



Universidad de Jaén

Escuela Politécnica Superior de Jaén

**Evaluación de técnicas de aprendizaje profundo
para la detección de productos fitosanitarios
en la industria del olivar**

Autor/a: María Cristina Fernández Ruiz

Máster: Ingeniería Industrial

**Tutores: Silvia Satorres Martínez
Diego Manuel Martínez Gila**

Departamento de Ingeniería Electrónica y Automática

Fecha: 02/12/2024



CREEA



Universidad de Jaén
Escuela Politécnica Superior de Jaén
Departamento de Ingeniería Electrónica y Automática

Doña Silvia María Satorres Martínez y Don Diego Manuel Martínez Gila, tutores del Proyecto Fin de Carrera titulado: Evaluación de técnicas de aprendizaje profundo para la detección de productos fitosanitarios en la industria del olivar, que presenta María Cristina Fernández Ruiz, autorizan su presentación para defensa y evaluación en la Escuela Politécnica Superior de Jaén.

Jaén, 2 de diciembre de 2024

La alumna:

María Cristina Fernández Ruiz

Los tutores:

Resumen

El objetivo del TFM es estudiar la capacidad de los métodos de aprendizaje profundo para la clasificación de tipos y niveles de concentración de fitosanitarios a partir de imágenes hiperespectrales de aceitunas y hojas.

Para ello, primero se obtuvieron las áreas de interés para aceitunas y hojas mediante segmentación semántica y umbralización, respectivamente. A continuación, se implementó una red neuronal convolucional 1D (CNN-1D) con la que se clasificaron las imágenes obteniendo exactitudes superiores al 90%, sobre todo en la clasificación por niveles, en algunos casos tras optimizar la red.

Los resultados se compararon con los obtenidos con métodos de aprendizaje automático, siendo los resultados también ajustados, aunque ligeramente inferiores en general a los de la red CNN-1D. Adicionalmente, se han aplicado métodos de selección de características, obteniendo bandas con significación física que permiten disminuir la cantidad de datos.

Se comprueba así la utilidad de estas técnicas, como método no invasivo para detectar fitosanitarios en las aceitunas y hojas.

Abstract

The objective of the Master's Thesis (TFM) is to study the ability of deep learning methods to classify types and concentration levels of pesticides based on hyperspectral images of olives and leaves.

To achieve this, areas of interest for olives and leaves were first obtained using semantic segmentation and thresholding, respectively. Subsequently, a one-dimensional convolutional neural network (1D-CNN) was implemented to classify the images, achieving accuracies above 90%, particularly in level classification, in some cases after optimizing the network.

The results were compared with those obtained using machine learning methods, which were also consistent, though generally slightly inferior to those of the 1D-CNN. Additionally, feature selection methods were applied, identifying bands with physical significance that allowed for a reduction in the amount of data.

These findings demonstrate the utility of these techniques as a non-invasive method for detecting pesticides in olives and leaves.

Índice

1. INTRODUCCIÓN	1
1.1. Justificación	1
1.2. Objetivos	4
1.3. Antecedentes	5
1.3.1. Aprendizaje automático y aprendizaje profundo	5
1.3.1.1. Conceptos básicos	5
1.3.1.2. Aprendizaje automático y reconocimiento de patrones	6
1.3.1.3. Proceso de clasificación y tipos de clasificadores	8
1.3.1.4. Métodos de clasificación en aprendizaje automático	15
1.3.1.5. Redes neuronales y aprendizaje profundo	20
1.3.1.6. Selección de características y aprendizaje automático	29
1.3.2. Aprendizaje automático y profundo en el análisis de imágenes	31
1.3.2.1. Visión por computador	31
1.3.2.2. Imagen digital	33
1.3.2.3. Preprocesado de imágenes	34
1.3.2.4. Análisis de imágenes	36
1.3.2.5. Reconocimiento de objetos e interpretación de imágenes	41
1.3.2.6. Aprendizaje automático en el análisis de imágenes	44
1.3.2.7. Aprendizaje profundo en el análisis de imágenes	45
1.3.2.8. Imágenes hiperespectrales	55
1.3.3. Aprendizaje automático y profundo para el análisis de imágenes en la agricultura	60
1.3.3.1. Escala de paisaje	60
1.3.3.2. Escala de árbol o plantas	63
1.3.3.3. Escala de órganos	67
1.3.3.4. Otras aplicaciones en el sector oleícola	70
2. MATERIALES Y MÉTODOS	73
2.1.1. Hardware y software	73
2.1.2. Preparación de las muestras y adquisición de imágenes	74
2.2. Metodología	80
2.2.1. Interfaz de visualización y consulta	81
2.2.2. Preprocesado	83
2.2.2.1. Obtención del área de interés de las imágenes	83
2.2.2.2. Corrección B/N y normalización	95

2.2.3.	Obtención de las curvas espectrales.....	96
2.2.3.1.	Consulta y extracción de las curvas espectrales de cada pixel	97
2.2.3.2.	Obtención de las curvas espectrales medias de un área de interés	97
2.2.3.3.	Obtención de las curvas espectrales medias de una ventana definida por el usuario	98
2.2.3.4.	Obtención de las curvas espectrales en cuadrículas creadas automáticamente	100
2.2.4.	Selección de características.....	101
2.2.4.1.	Análisis de componentes principales.....	103
2.2.4.2.	Análisis de la varianza.....	104
2.2.4.3.	Eliminación recursiva de características.....	106
2.2.4.4.	Aprendizaje automático: Bosques aleatorios.....	107
2.2.4.5.	Basado en activaciones internas de la red 1D.....	107
2.2.5.	Clasificación de las muestras	109
2.2.5.1.	Técnicas de aprendizaje automático	109
2.2.5.2.	Técnicas de aprendizaje profundo	115
2.2.6.	Segmentación y clasificación de las imágenes.....	126
3.	RESULTADOS Y DISCUSIÓN.....	129
3.1.	Curvas espectrales de las aceitunas	129
3.2.	Clasificación de las muestras de aceitunas	133
3.2.1.	Técnicas de aprendizaje automático	133
3.2.1.1.	Curvas espectrales medias	133
3.2.1.2.	Curvas espectrales extraídas de las ventanas manualmente	147
3.2.1.3.	Curvas espectrales extraídas de las cuadrículas automáticamente.....	161
3.2.2.	Técnicas de aprendizaje profundo.....	175
3.2.2.1.	Curvas espectrales medias	175
3.2.2.2.	Curvas espectrales extraídas de las ventanas manualmente	187
3.2.2.3.	Curvas espectrales extraídas de las cuadrículas automáticamente.....	199
3.3.	Selección de características de las aceitunas	210
3.4.	Clasificación de las muestras de hojas.....	225
3.5.	Selección de características de las hojas	236
4.	CONCLUSIONES.....	238
	Bibliografía	245

Índice de ilustraciones

Ilustración 1-1. Inteligencia artificial, aprendizaje automático y aprendizaje profundo.....	6
Ilustración 1-2. a) Aprendizaje supervisado; b) Aprendizaje no supervisado, tomada de (Torres, 2020)	8
Ilustración 1-3. División del conjunto de datos típica para Machine Learning.	10
Ilustración 1-4. Muestra del underfitting, overfitting y ajuste correcto (tomada de https://docplayer.es/63777861-Sobreajuste-overfitting.html).	11
Ilustración 1-5. Matriz de confusión en la clasificación binaria.	12
Ilustración 1-6. Matriz de confusión en la clasificación con más de dos clases.....	14
Ilustración 1-7. Esquema de trabajo de los bosques aleatorios (tomada de https://insideaiml.com/blog/Random-Forest-Algorithm-1029).....	18
Ilustración 1-8. Support Vector Machines (SVM, tomada de https://www.ml2projects.com/post/svm_kernelab).	19
Ilustración 1-9. Importancia de las dimensiones en SVM (tomada de https://www.cienciadedatos.net/documentos/34_maquinas_de_vector_soporte_support_vector_machines).	20
Ilustración 1-10. Funcionamiento de una neurona simple o perceptrón (tomada de Torres, 2020).....	21
Ilustración 1-11. Fases del entrenamiento de las neuronas (tomada de Torres, 2020).....	23
Ilustración 1-12. Descenso del gradiente (tomada de Torres, 2020).....	24
Ilustración 1-13. Perceptrón Multicapa (MLP, tomada de Torres, 2020).	26
Ilustración 1-14. Esquema de una red neuronal recurrente (tomada de (Bosch, et al., 2019)).	29
Ilustración 1-15. Aumento del brillo y expansión del contraste (tomada de Jiménez, 2020)..	35
Ilustración 1-16. Filtros de paso bajo (tomada de (Espinosa-Aranda, et al., 2016)).	35
Ilustración 1-17. Filtros de extracción de bordes (tomada de (Espinosa-Aranda, et al., 2016)).	36
Ilustración 1-18. Filtros de Roberts, Sobel, Prewitt y laplaciano (elaboración propia).	37
Ilustración 1-19. Detectores de esquinas y puntos característicos: a: Harris; b: SIFT (tomada de (Chen, et al., 2012)); b: SIFT (tomada de (Lowe, 2004)).	38
Ilustración 1-20. Umbralización mediante k-means (tomada de (Jiménez, 2020; Kanso & Ghebleh, 2017)).	39
Ilustración 1-21. Segmentación y clasificación de imágenes en color (tomada de https://jp.mathworks.com/discovery/image-segmentation.html).	39
Ilustración 1-22. Algoritmo de k-means (tomada de https://medium.com/@rohithramesh1991/unsupervised-text-clustering-using-natural-language-processing-nlp-1a8bc18b048d).....	43
Ilustración 1-23. Aprendizaje automático: Árboles de decisión (tomada de https://www.edrawsoft.com/es/decision-tree); b: Support Vector Machines (SVM, tomada de https://numerentur.org/svm/).....	44
Ilustración 1-24. Funcionamiento de la capa de convolución (tomada de Torres, 2020).	46
Ilustración 1-25. Mapa de características (tomada de Torres, 2020).	47
Ilustración 1-26. Función de activación (tomada de https://pythonwife.com/intro-to-deep-learning-neural-networks-for-the-purpose-of-nlp/).....	47
Ilustración 1-27. Funcionamiento del Max-pooling (tomada de Torres, 2020).....	48

Ilustración 1-28. Aplicación de una CNN a una imagen (tomada de https://medium.com/swlh/convolutional-neural-networks-for-multiclass-image-classification-a-beginners-guide-to-6dbc09fabbd).....	49
Ilustración 1-29. Resultado CNN para Clasificación b) Resultado para detección de objetos c) Resultado con segmentación semántica (tomada de (Sirmaçek, et al., 2019); https://deepai.org/publication/sequential-image-processing-methods-for-improving-semantic-video-segmentation-algorithms).	50
Ilustración 1-30. a) Segmentación semántica; b) Segmentación de instancias (tomada de (Lin, et al., 2014); https://link.springer.com/content/pdf/10.1007%2F978-3-319-10602-1_48.pdf).....	50
Ilustración 1-31. Bloque residual de una red ResNet (tomada de (He, et al., 2015)).....	53
Ilustración 1-32. Estructura básica de un autoencoder (tomada de (Bosch, et al., 2019)).....	54
Ilustración 1-33. Imagen hiperespectral (hipercubo e imagen RGB), tomada de (Bhargava, et al., 2024).	55
Ilustración 2-1. Cámara multispectral empleada.	77
Ilustración 2-2. Diagrama general de flujo seguido en este trabajo.	80
Ilustración 2-3. Regiones del espectro electromagnético entre el visible e infrarrojo cercano (tomada de (Ito, et al., 2023): https://ceur-ws.org/Vol-3459/paper3.pdf	81
Ilustración 2-4. Interfaz de visualización y consulta de imágenes.	82
Ilustración 2-5. Diferentes combinaciones de las imágenes en color real y falso color.	82
Ilustración 2-6. Curvas espectrales extraídas de los píxeles.....	83
Ilustración 2-7. Red Deeplab (tomada de (Chen, et al., 2017)).	85
Ilustración 2-8. DeeplabV3+ (tomada de (Wang, et al., 2022c)).....	86
Ilustración 2-9. Estructura de la red DeepLabV3+ con base ResNet 18.	87
Ilustración 2-10. Interfaz Image Labeler de Matlab para etiquetamiento de píxeles.	89
Ilustración 2-11. Progreso de entrenamiento en la red DeepLabV3+.....	90
Ilustración 2-12. Imágenes clasificadas de aceitunas mediante la red DeepLabV3+ Resnet 18, se muestra en azul lo considerado como el área de interés de las aceitunas (zonas sanas sin defectos).	91
Ilustración 2-13. Matrices de confusión para los datos de entrenamiento, validación y test en la red DeepLabV3+ empleada para la obtención de las áreas de interés de las imágenes...93	
Ilustración 2-14. Obtención de las áreas de interés en las hojas mediante umbralización del NDVI.	94
Ilustración 2-15. Interfaz de la aplicación para la extracción de los valores medios de reflectancia en ventanas definidas manualmente.	99
Ilustración 2-16. Creación de las cuadrículas definidas automáticamente para la extracción de los valores medios de reflectancia en aceitunas y hojas.....	100
Ilustración 2-17. Ventana del análisis de componentes principales (PCA) de Matlab.	104
Ilustración 2-18. Componentes principales y varianza explicada.	104
Ilustración 2-19. Ventana de Selección de Características (Feature Selection) en Matlab..	105
Ilustración 2-20. Características seleccionadas para el modelo.....	106
Ilustración 2-21. Classification Learner en el menú principal de Matlab.	111
Ilustración 2-22. Ventana principal del Classification Learner de Matlab.	111
Ilustración 2-23. Clasificadores disponibles en la versión empleada de Matlab.	112
Ilustración 2-24. Esquema de las capas de la red convolucional (CNN) 1D implementada.	118
Ilustración 2-25. Curvas de exactitud de entrenamiento y validación.....	120

Ilustración 2-26. Matriz de confusión de los datos de tipo test tras la aplicación de la red convolucional.	120
Ilustración 2-27. Red obtenida tras la optimización.....	122
Ilustración 2-28. Iteraciones realizadas para obtener la red óptima.	123
Ilustración 2-29. Hiperparámetros de la red optimizada.....	123
Ilustración 2-30. Gráficos obtenidos con información sobre la optimización.	123
Ilustración 2-31. Características de la red optimizada.....	124
Ilustración 2-32. Interfaz de visualización por ventanas de la concentración de pesticidas fitosanitarios de la imagen y porcentaje de defectos de la aceituna.	128
Ilustración 3-1. Curvas espectrales medias de las aceitunas tratadas con los distintos tipos de producto.	129
Ilustración 3-2. Curvas espectrales medias de las aceitunas tratadas con deltametrina.	130
Ilustración 3-3. Curvas espectrales medias de las aceitunas tratadas con diflufenican.....	130
Ilustración 3-4. Curvas espectrales medias de las aceitunas tratadas con lambda-cihalotrina.	131
Ilustración 3-5. Curvas espectrales medias de las aceitunas tratadas con oxifluorfen.	131
Ilustración 3-6. Curvas espectrales medias de las aceitunas tratadas con tebuconazol.	132
Ilustración 3-7. Matrices de confusión de los datos de test en el aprendizaje automático por tipos de producto en los espectros medios de las aceitunas: a: Análisis discriminante; b: Árbol de decisión.....	135
Ilustración 3-8. Matrices de confusión de los datos de test en el aprendizaje automático para los niveles de deltametrina en los espectros medios de las aceitunas: a: SVM; b: Bosques aleatorios.....	137
Ilustración 3-9. Matrices de confusión de los datos de test en el aprendizaje automático para los niveles de diflufenican en los espectros medios de las aceitunas: a: Análisis discriminante; b: Árboles de decisión.	139
Ilustración 3-10. Matrices de confusión de los datos de test en el aprendizaje automático para los niveles de lambda-cihalotrina en los espectros medios de las aceitunas: a: SVM; b: k-NN.....	141
Ilustración 3-11. Matrices de confusión de los datos de test en el aprendizaje automático para los niveles de oxifluorfen en los espectros medios de las aceitunas: a: Análisis discriminante; b: k-NN.	143
Ilustración 3-12. Matrices de confusión de los datos de test en el aprendizaje automático para los niveles de tebuconazol en los espectros medios de las aceitunas: a: SVM; b: Bosques aleatorios.....	145
Ilustración 3-13. Matrices de confusión de los datos de test en el aprendizaje automático por tipos de producto en los espectros medios de las ventanas definidas manualmente: a: Análisis discriminante; b: Árboles de decisión.	149
Ilustración 3-14. Matrices de confusión de los datos de test en el aprendizaje automático para los niveles de deltametrina en los espectros medios de las ventanas definidas manualmente: a: Red neuronal; b: Árboles de decisión.....	151
Ilustración 3-15. Matrices de confusión de los datos de test en el aprendizaje automático para los niveles de diflufenican en los espectros medios de las ventanas definidas manualmente: a: Análisis discriminante; b: k-NN.....	153
Ilustración 3-16. Matrices de confusión de los datos de test en el aprendizaje automático para los niveles de lambda-cihalotrina en los espectros medios de las ventanas definidas manualmente: a: Red neuronal; b: k-NN.	155

Ilustración 3-17. Matrices de confusión de los datos de test en el aprendizaje automático para los niveles de oxifluorfen en los espectros medios de las ventanas definidas manualmente: a: Análisis discriminante; b: Árboles de decisión.	157
Ilustración 3-18. Matrices de confusión de los datos de test en el aprendizaje automático para los niveles de tebuconazol en los espectros medios de las ventanas definidas manualmente: a: Análisis discriminante; b: Árboles de decisión.	159
Ilustración 3-19. Matrices de confusión de los datos de test en el aprendizaje automático por tipos de producto en los espectros medios de las cuadrículas extraídas automáticamente: a: Análisis discriminante; b: Árboles de decisión.	163
Ilustración 3-20. Matrices de confusión de los datos de test en el aprendizaje automático para los niveles de deltametrina en los espectros medios de las cuadrículas extraídas automáticamente: a: Red neuronal; b: Árboles de decisión.	165
Ilustración 3-21. Matrices de confusión de los datos de test en el aprendizaje automático para los niveles de diflufenican en los espectros medios de las cuadrículas extraídas automáticamente: a: Red neuronal; b: Árboles de decisión.	167
Ilustración 3-22. Matrices de confusión de los datos de test en el aprendizaje automático para los niveles de lambda-cihalotrina en los espectros medios de las cuadrículas extraídas automáticamente: a: Red neuronal; b: k-NN.	169
Ilustración 3-23. Matrices de confusión de los datos de test en el aprendizaje automático para los niveles de oxifluorfen en los espectros medios de las cuadrículas extraídas automáticamente: a: Red neuronal; b: Árboles de decisión.	171
Ilustración 3-24. Matrices de confusión de los datos de test en el aprendizaje automático para los niveles de tebuconazol en los espectros medios de las cuadrículas extraídas automáticamente: a: Red neuronal; b: Árboles de decisión.	173
Ilustración 3-25. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para los tipos de producto en los espectros medios de las aceitunas.	177
Ilustración 3-26. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para los niveles de deltametrina en los espectros medios de las aceitunas.	179
Ilustración 3-27. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para los niveles de diflufenican en los espectros medios de las aceitunas.	180
Ilustración 3-28. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para los niveles de lambda-cihalotrina en los espectros medios de las aceitunas.	182
Ilustración 3-29. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para los niveles de oxifluorfen en los espectros medios de las aceitunas.	183
Ilustración 3-30. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para los niveles de tebuconazol en los espectros medios de las aceitunas.	186
Ilustración 3-31. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para el tipo de producto en los espectros medios de las ventanas definidas manualmente.	189

Ilustración 3-32. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para el nivel de deltametrina en los espectros medios de las ventanas definidas manualmente.	191
Ilustración 3-33. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para el nivel de diflufenican en los espectros medios de las ventanas definidas manualmente.	192
Ilustración 3-34. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para el nivel de lambda-cihalotrina en los espectros medios de las ventanas definidas manualmente.	194
Ilustración 3-35. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para el nivel de oxifluorfen en los espectros medios de las ventanas definidas manualmente.	195
Ilustración 3-36. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para el nivel de tebuconazol en los espectros medios de las ventanas definidas manualmente.	198
Ilustración 3-37. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para el tipo de producto en los espectros medios de las cuadrículas extraídas automáticamente.....	200
Ilustración 3-38. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para el nivel de deltametrina en los espectros medios de las cuadrículas extraídas automáticamente.	202
Ilustración 3-39. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para el nivel de diflufenican en los espectros medios de las cuadrículas extraídas automáticamente.	203
Ilustración 3-40. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para el nivel de lambda-cihalotrina en los espectros medios de las cuadrículas extraídas automáticamente.	205
Ilustración 3-41. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para el nivel de oxifluorfen en los espectros medios de las cuadrículas extraídas automáticamente.	206
Ilustración 3-42. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para el nivel de tebuconazol en los espectros medios de las cuadrículas extraídas automáticamente.	207
Ilustración 3-43. Selección de características mediante diversos métodos: mapa de importancia (red 1D); eliminación recursiva; bosques aleatorios; análisis de la varianza (ANOVA) para los tipos de producto.....	212
Ilustración 3-44. Regiones del espectro marcadas como más importantes en los métodos de selección de características para diferenciar por tipos de producto.	213
Ilustración 3-45. Selección de características mediante diversos métodos: mapa de importancia (red 1D); eliminación recursiva; bosques aleatorios; análisis de la varianza (ANOVA) para los niveles de deltametrina.	214
Ilustración 3-46. Regiones del espectro marcadas como más importantes en los métodos de selección de características para diferenciar por niveles de deltametrina.....	215
Ilustración 3-47. Selección de características mediante diversos métodos: mapa de importancia (red 1D); eliminación recursiva; bosques aleatorios; análisis de la varianza (ANOVA) para los niveles de diflufenican.....	216

Ilustración 3-48. Regiones del espectro marcadas como más importantes en los métodos de selección de características para diferenciar por niveles de diflufenican.	217
Ilustración 3-49. Selección de características mediante diversos métodos: mapa de importancia (red 1D); eliminación recursiva; bosques aleatorios; análisis de la varianza (ANOVA) para los niveles de lambda-cihalotrina.	218
Ilustración 3-50. Regiones del espectro marcadas como más importantes en los métodos de selección de características para diferenciar por niveles de lambda-cihalotrina.	219
Ilustración 3-51. Selección de características mediante diversos métodos: mapa de importancia (red 1D); eliminación recursiva; bosques aleatorios; análisis de la varianza (ANOVA) para los niveles de oxifluorfen.	220
Ilustración 3-52. Regiones del espectro marcadas como más importantes en los métodos de selección de características para diferenciar por niveles de oxifluorfen.	221
Ilustración 3-53. Selección de características mediante diversos métodos: mapa de importancia (red 1D); eliminación recursiva; bosques aleatorios; análisis de la varianza (ANOVA) para los niveles de tebuconazol.	222
Ilustración 3-54. Regiones del espectro marcadas como más importantes en los métodos de selección de características para diferenciar por niveles de tebuconazol.	223
Ilustración 3-55. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para los tipos de producto en los espectros medios de las ventanas definidas automáticamente en las hojas.	228
Ilustración 3-56. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para el nivel de deltametrina en los espectros medios de las ventanas definidas automáticamente en las hojas.	230
Ilustración 3-57. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para el nivel de diflufenican en los espectros medios de las ventanas definidas automáticamente en las hojas.	231
Ilustración 3-58. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para el nivel de lambda-cihalotrina en los espectros medios de las ventanas definidas automáticamente para las hojas.	232
Ilustración 3-59 Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para el nivel de oxifluorfen en los espectros medios de las ventanas definidas automáticamente para las hojas.	234
Ilustración 3-60. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para el nivel de tebuconazol en los espectros medios de las ventanas definidas automáticamente en las hojas.	235
Ilustración 3-61. Bandas seleccionadas para cada una de las redes entrenadas.	237

Índice de tablas

Tabla 2-1. Concentraciones de los productos fitosanitarios aplicados a las aceitunas en este trabajo.	75
Tabla 2-2. Concentraciones de los productos fitosanitarios en las aceitunas resultantes de los análisis.....	76
Tabla 2-3. Características de la cámara multispectral.	77
Tabla 2-4. Imágenes empleadas en el trabajo.	79
Tabla 2-5. Hiperparámetros seleccionados en la red empleada.	88
Tabla 2-6. Datos del entrenamiento de la red DeepLabV3+.	90
Tabla 2-7. Exactitud o tasa de acierto para los tres conjuntos de datos.	91
Tabla 2-8. Arquitectura de la red e hiperparámetros de entrenamiento para las distintas redes de clasificación de aceitunas.	125
Tabla 2-9 Arquitectura de la red e hiperparámetros de entrenamiento para las distintas redes de clasificación de hojas.....	125
Tabla 3-1. Resultados de la exactitud (accuracy) en validación y test del aprendizaje automático por tipos de producto en los espectros medios de las aceitunas.	134
Tabla 3-2. Resultados de la exactitud (accuracy) en validación y test del aprendizaje automático, para los niveles de deltametrina en los espectros medios de las aceitunas.9 .	136
Tabla 3-3. Resultados de la exactitud (accuracy) en validación y test del aprendizaje automático, para los niveles de diflufenican en los espectros medios de las aceitunas.	138
Tabla 3-4. Resultados de la exactitud (accuracy) en validación y test del aprendizaje automático, para los niveles de lambda-cihalotrina en los espectros medios de las aceitunas.	140
Tabla 3-5. Resultados de la exactitud (accuracy) en validación y test del aprendizaje automático, para los niveles de oxifluorfen en los espectros medios de las aceitunas.	142
Tabla 3-6. Resultados de la exactitud (accuracy) en validación y test del aprendizaje automático, para los niveles de tebuconazol en los espectros medios de las aceitunas.	144
Tabla 3-7. Resultados de la exactitud (accuracy) en validación y test del aprendizaje automático por tipos de producto en los espectros medios de las ventanas definidas manualmente.	148
Tabla 3-8. Resultados de la exactitud (accuracy) en validación y test del aprendizaje automático, para los niveles de deltametrina en las ventanas definidas manualmente.....	150
Tabla 3-9. Resultados de la exactitud (accuracy) en validación y test del aprendizaje automático, para los niveles de diflufenican en los espectros medios de las ventanas definidas manualmente.	152
Tabla 3-10. Resultados de la exactitud (accuracy) en validación y test del aprendizaje automático, para los niveles de lambda-cihalotrina en los espectros medios de las ventanas definidas manualmente.	154
Tabla 3-11. Resultados de la exactitud (accuracy) en validación y test del aprendizaje automático, para los niveles de oxifluorfen en los espectros medios de las ventanas definidas manualmente.	156
Tabla 3-12. Resultados de la exactitud (accuracy) en validación y test del aprendizaje automático, para los niveles de tebuconazol en los espectros medios de las ventanas definidas manualmente.	158
Tabla 3-13. Resultados de la exactitud (accuracy) en validación y test del aprendizaje automático por tipos de producto en los espectros medios de las cuadrículas extraídas automáticamente.....	162

Tabla 3-14. Resultados de la exactitud (accuracy) en validación y test del aprendizaje automático, para los niveles de deltametrina de las cuadrículas extraídas automáticamente.	164
Tabla 3-15. Resultados de la exactitud (accuracy) en validación y test del aprendizaje automático, para los niveles de diflufenican en los espectros medios de las cuadrículas extraídas automáticamente.	166
Tabla 3-16. Resultados de la exactitud (accuracy) en validación y test del aprendizaje automático, para los niveles de lambda-cihalotrina en los espectros medios de las cuadrículas extraídas automáticamente.9	168
Tabla 3-17. Resultados de la exactitud (accuracy) en validación y test del aprendizaje automático, para los niveles de oxifluorfen en los espectros medios de las cuadrículas extraídas automáticamente.	170
Tabla 3-18. Resultados de la exactitud (accuracy) en validación y test del aprendizaje automático, para los niveles de tebuconazol en los espectros medios de las cuadrículas extraídas automáticamente.	172
Tabla 3-19. Resultados de la exactitud (accuracy) en entrenamiento, validación y test de la red convolucional para los tipos y niveles de producto en los espectros medios de las aceitunas.....	175
Tabla 3-20. Resultados de la exactitud (accuracy) en entrenamiento, validación y test de la red convolucional para los tipos y niveles de producto en los espectros medios de	187
Tabla 3-21. Resultados de la exactitud (accuracy) en entrenamiento, validación y test de la red convolucional para los tipos y niveles de producto en los espectros medios de las cuadrículas extraídas automáticamente.	199
Tabla 3-22. Resultado selección de características en los datos de entrenamiento, validación y test para tres métodos: evaluación de activaciones en red 1-D (mapa de importancia), selección recursiva y random forest.	210
Tabla 3-23. Resultado para la exactitud (accuracy) de clasificación de las hojas de los datos de entrenamiento, validación y test mediante las redes entrenados	226
Tabla 3-24. Resultado para la exactitud (accuracy) de clasificación de las hojas de los datos de entrenamiento, validación y test mediante las redes entrenadas tras la selección de características	236

Índice de ecuaciones

(Ecuación 1-1)	8
(Ecuación 1-2).....	8
(Ecuación 1-3).....	13
(Ecuación 1-4).....	13
(Ecuación 1-5).....	13
(Ecuación 1-6).....	13
(Ecuación 1-7).....	16
(Ecuación 1-8).....	21
(Ecuación 1-9).....	23
(Ecuación 1-10).....	27
(Ecuación 1-11).....	27
(Ecuación 1-12).....	27
(Ecuación 2-1).....	93
(Ecuación 2-2).....	95
(Ecuación 2-3).....	96

1. INTRODUCCIÓN

1.1. Justificación

De acuerdo con la FAO (Organización de las Naciones Unidas para la Alimentación y la Agricultura), el sector agrícola utiliza aproximadamente tres millones de toneladas de pesticidas cada año. Sin estos productos, las pérdidas en la producción alcanzarían el 78% en frutas, el 54% en verduras y el 32% en cereales (Zhang, et al., 2011). El uso de estos productos se extiende más allá del propio cultivo, previniendo el deterioro de las frutas después de la cosecha (Min, et al., 2024).

En el cultivo del olivar, el uso de pesticidas está igualmente justificado, sobre todo ante las previsiones que estiman que en las regiones del sur de la península ibérica se produzca una reducción del rendimiento de los cultivos del olivar de secano de entre un 17% y un 20%, por la aparición de nuevas plagas que afecten a los cultivos y el cambio de las condiciones climáticas (Fraga, et al., 2020; Lantero, et al., 2023). Hay que tener en cuenta que España es líder mundial tanto en producción de aceite de oliva como en superficie cultivada, suponiendo la española el 70% de la producción de la Unión Europea y el 45% de la mundial (Velasco, 2009; Fazari, et al., 2018; MAPA, 2024). Más específicamente, en Andalucía se localiza un 80% del total y en Jaén un 37%, por lo que la citada reducción en la producción puede afectar no solo al sector del aceite de oliva, sino a la economía general y la sociedad de la región.

El uso de pesticidas y en general productos fitosanitarios en este sector varía según factores como la ubicación geográfica, las condiciones climáticas, las plagas específicas y las prácticas agrícolas de cada región (Therios, 2009). Sin embargo, aunque los pesticidas son efectivos para aumentar el rendimiento de los cultivos, también presentan serias desventajas para los ecosistemas y la vida en general. Según una revisión realizada por (de Souza, et al., 2020), se ha comprobado que los pesticidas, incluso cuando se aplican directamente al suelo o a los árboles, pueden llegar a contaminar aguas superficiales y subterráneas mediante procesos como la escorrentía, la percolación, la evaporación y la precipitación. Este impacto generalizado afecta a la microflora, la microfauna y las plantas, integrando estos compuestos químicos en la cadena alimentaria y generando riesgos tanto para el

equilibrio del ecosistema (Tudi, et al., 2021) como para la salud humana (Elahi, et al., 2019; El-Nahhal & El-Nahhal, 2021).

Debido a estos riesgos ambientales y de la salud, se han establecido regulaciones estrictas para limitar los residuos de pesticidas en los alimentos, incluidas las aceitunas recolectadas, lo que es crucial para proteger a los consumidores y preservar el medio ambiente. Así, en este contexto, la Unión Europea y la Comisión del *Codex Alimentarius* de la FAO han definido límites máximos de residuos (LMR) para cada fitosanitario presente en el aceite de oliva, garantizando así la seguridad del consumidor (Luchetti, 2002; García-Reyes, et al., 2007). Estos LMR representan la concentración máxima permitida legalmente de residuos de pesticidas en los alimentos. En España, el Ministerio de Agricultura, Pesca y Alimentación (MAPA) ha introducido el concepto de periodo de seguridad, que se refiere al tiempo mínimo que debe transcurrir entre la aplicación del producto fitosanitario en el olivar y la cosecha del fruto. Este periodo asegura el cumplimiento de los LMR en el aceite de oliva. Sin embargo, hasta ahora no existe actualmente una herramienta práctica que permita verificar rápidamente, en el momento de la entrega en las cooperativas, si se ha respetado dicho periodo de seguridad (Martínez-Gila, et al., 2024).

Las técnicas de detección de plaguicidas se pueden dividir en dos tipos, invasivas y no invasivas. Las primeras requieren extraer la muestra del proceso (técnicas *off-line*) y transportarla a un laboratorio analítico donde se prepara y procesa mediante instrumentación específica. Ejemplos de estas técnicas son la cromatografía de gases (Ferrer, et al., 2005) y la cromatografía de líquidos (Hu, et al., 2014), si bien estas últimas podrían adaptarse para su uso directamente en la línea de producción (*in-line*). Entre las técnicas no invasivas (Sindhu & Manickavasagan, 2023), se puede citar la espectroscopia *Raman* (Zhang, et al., 2022b), la espectroscopia fotoacústica (Liu, et al., 2015), la nariz electrónica (Martínez Gila, et al., 2021; Martínez Gila, et al., 2023; Navarro Soto, et al., 2024) o la espectroscopia de fluorescencia (Wang, et al., 2021); todas ellas técnicas contrastadas, pero difíciles de implementar en procesos industriales en línea por diversos motivos (Martínez-Gila, et al., 2024).

Dadas estas limitaciones, se ha realizado un estudio por parte de investigadores del Grupo de Robótica y Automática y Visión por Computador (GRAV) de la

Universidad de Jaén, en el que propone la utilización de una técnica basada en la obtención y análisis de imágenes hiperespectrales en el espectro visible e infrarrojo cercano (*near infrared*, NIR). Las imágenes de las aceitunas, previamente tratadas con diversos productos fitosanitarios de tipo pesticidas (herbicidas, como el diflufenican y oxilfluorfen; insecticidas, como la deltametrina y lambda-cihalotrina; y fungicidas, como el tebuconazol), han sido adquiridas en el laboratorio por miembros del grupo, en condiciones controladas. Una vez adquiridas, las imágenes han sido etiquetadas a partir de los datos proporcionados por el análisis químico de las aceitunas y a partir de ahí clasificadas mediante diversas técnicas de aprendizaje automático (Martínez-Gila, et al., 2024). Se trata de una alternativa rápida, sin contacto y rentable, adecuada para el monitoreo en línea y en tiempo real de residuos de pesticidas. Este enfoque no solo permite la detección directa de residuos de pesticidas en aceitunas recién cosechadas, sino que también aborda la necesidad de una solución escalable e industrialmente viable que las técnicas actuales no logran satisfacer.

Una vez contrastada la validez de este enfoque, en este trabajo se propone aplicar técnicas de aprendizaje profundo sobre las imágenes hiperespectrales del mismo conjunto anterior de aceitunas, junto con un conjunto adicional de imágenes de hojas de olivo, tratadas y adquiridas de igual forma que las aceitunas. Las imágenes analizadas se han proporcionado por los tutores a cargo de estas investigaciones. En el trabajo se han desarrollado técnicas de aprendizaje profundo, como redes neuronales convolucionales 1D (*convolutional neural networks*, CNN) que, según la amplia bibliografía consultada, son las más adecuadas para este tipo de estudios centrados en la hiperespectralidad de las imágenes. Una vez desarrolladas, las técnicas se han aplicado para detectar y diferenciar distintos tipos de pesticidas y, sobre todo, su concentración, de cara al control de calidad de acuerdo a las normas europeas sobre productos fitosanitarios antes mencionadas. Adicionalmente se han analizado y seleccionado las bandas o longitudes onda más efectivas para la detección de los tipos y niveles de pesticidas.

1.2. Objetivos

Este trabajo se centra en la aplicación de técnicas de inteligencia artificial para la detección de productos fitosanitarios en aceitunas y hojas de olivo a partir de imágenes hiperespectrales tomadas en laboratorio. Así, se han ensayado técnicas de aprendizaje automático y profundo, en especial redes neuronales convolucionales, sobre las imágenes de aceitunas y hojas a las que se han aplicado distintos productos.

De esta forma, el objetivo principal del trabajo es desarrollar y evaluar distintas técnicas de aprendizaje automático y profundo para la detección y estimación de la concentración de productos sanitarios en aceitunas y hojas de olivo, a partir de imágenes hiperespectrales de estos órganos tomadas en laboratorio.

Como objetivos específicos se pueden indicar:

- Desarrollar y entrenar distintas técnicas de aprendizaje automático y profundo supervisado, y las aplicaciones informáticas correspondientes, para la detección de productos sanitarios en aceitunas y hojas de olivo a partir de imágenes hiperespectrales.
- Comparar y evaluar los resultados de las distintas técnicas desarrolladas y aplicadas a partir de muestras etiquetadas de validación, determinando aquellas que muestren una mejor exactitud para la detección de fitosanitarios en las aceitunas y hojas de olivo.
- Determinar las bandas más relevantes para la detección de fitosanitarios en las aceitunas e hojas de olivo a partir de las imágenes hiperespectrales, y así aplicar de técnicas de reducción de la dimensionalidad y selección de características para la mejorar la eficiencia de las técnicas.

1.3. Antecedentes

1.3.1. Aprendizaje automático y aprendizaje profundo

1.3.1.1. Conceptos básicos

Inteligencia artificial, aprendizaje automático y aprendizaje profundo son tres términos muy relacionados entre sí, pero igualmente diferenciados. El concepto de inteligencia artificial es el más extenso y complejo, de tal forma que puede ser abordado desde distintas perspectivas que tienen que ver con el pensamiento, la racionalidad y las actuaciones prácticas (ingenieriles) que de ello se derivan (Russell & Norvig, 2010). Sin embargo, una aproximación más simple y general conduciría a decir que por inteligencia artificial se entiende aquella inteligencia que muestran las máquinas, en contraste con la inteligencia natural de los humanos. En este sentido, desde una visión ingenieril, la inteligencia artificial se podría definir como el esfuerzo para automatizar tareas intelectuales normalmente realizadas por humanos (Torres, 2020).

Así, el área de inteligencia artificial (IA) es un campo muy amplio que abarca muchas áreas del conocimiento, como las matemáticas, el lenguaje, la ingeniería, etc., que está permitiendo el desarrollo de aspectos tan dispares como el reconocimiento de voz, el procesado de lenguaje natural o la visión por computador que a su vez repercuten en diversas aplicaciones de la ingeniería y la industria (la robótica, la fabricación, la automoción, el control, la aeronáutica, etc.), la medicina, y un largo etcétera, que incluye la agricultura y su industria derivada.

Por su parte, el aprendizaje automático (muchas veces nombrado por su traducción y siglas en inglés: *Machine Learning*, ML), se puede definir igualmente de una forma práctica como el subcampo de la inteligencia artificial que proporciona a los ordenadores la capacidad de aprender sin ser explícitamente programados, es decir, sin que necesiten que el programador indique las reglas que debe seguir para lograr su tarea, sino que lo hace automáticamente (Torres, 2020). En términos prácticos, el aprendizaje automático consiste en desarrollar un algoritmo de predicción para la resolución de un problema particular, mediante un proceso de aprendizaje a partir de los datos que permite encontrar patrones o tendencias en ellos.

Un tipo especial de algoritmos dentro del aprendizaje automático son las redes neuronales artificiales (*artificial neural networks*, ANN), así denominadas porque su estructura trata de emular a las neuronas humanas y su capacidad para la obtención de resultados. Dentro de las redes neuronales, hay modelos que están compuestos de múltiples capas de procesamiento, con múltiples niveles de abstracción, en cuyo caso se habla de aprendizaje automático (conocido también por sus siglas en inglés, *Deep Learning*, DL). A partir de los datos de partida y a través de las múltiples capas neuronas, se llevan a cabo una serie de transformaciones lineales y no lineales para generar una salida que se aproxima iterativamente a la esperada (etiqueta o *label*). En el aprendizaje automático se persigue crear modelos que sean capaces de representar conceptos complejos y/o abstractos a partir de otros más sencillos.

En conclusión, se puede decir que el aprendizaje automático es un subconjunto parte del aprendizaje automático (Bosch, et al., 2019), que a su vez es una parte de la inteligencia artificial (Ilustración 1-1).

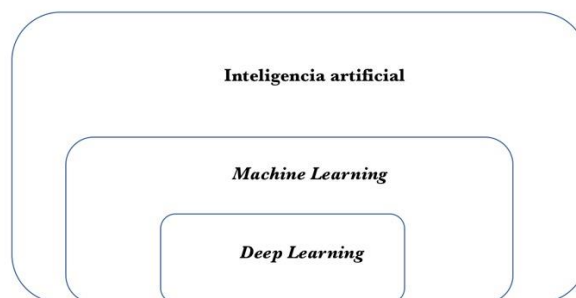


Ilustración 1-1. Inteligencia artificial, aprendizaje automático y aprendizaje profundo.

1.3.1.2. Aprendizaje automático y reconocimiento de patrones

Como se ha indicado anteriormente, **el reconocimiento de patrones es la finalidad básica del aprendizaje automático**. Este se basa en asignar etiquetas a los objetos indicando a qué clase pertenecen, a través de la representación de un conjunto de propiedades o atributos, mediante un vector n-dimensional (García-Ordás, et al., 2016).

Para ello, se emplean métodos supervisados y no supervisados (García-Ordás, et al., 2016; Bosch, et al., 2019; Torres, 2020), entre los que se considera un tercer

grupo, el aprendizaje por refuerzo, que goza de características de los dos primeros. En el aprendizaje supervisado se dispone de un conjunto de datos junto con sus etiquetas indicando a qué clase pertenece cada dato o ejemplo, también denominado patrón. El objetivo que se persigue es conseguir un modelo que permita etiquetar datos nuevos que no se hayan empleado en el propio ajuste del clasificador. Este proceso se conoce como entrenamiento del clasificador.

Por su parte, en las técnicas de aprendizaje no supervisado se dispone de los datos sin las etiquetas de clase. Tratan de modelar los datos en sí mismos, normalmente formando grupos (*clustering*) de patrones con atributos similares, buscando características, regularidades y correlaciones entre los datos, de forma que se puedan separar las categorías similares entre sí. Se buscan características que hacen a estos datos similares dentro del grupo y diferentes del resto.

En la comprensión del concepto de aprendizaje automático, además de los tipos de aprendizaje, es interesante considerar las tareas fundamentales que se realizan a partir de él y que son la base de tareas y aplicaciones más complejas. Entre estas tareas básicas se pueden citar la clasificación, la regresión y el agrupamiento (Bosch, et al., 2019). La clasificación es uno de los procesos cognitivos más importantes, tanto en la vida cotidiana como en los negocios; consiste en asignar instancias de un dominio dado, tanto de valor discreto como continuo, a un conjunto de clases, siendo estas los valores de un atributo discreto seleccionado. Un proceso inductivo muy similar es el de agrupamiento o *clustering*, en el que no dispone de una etiqueta de clase a predecir; consiste en dividir un conjunto de instancias de un dominio dado, discreto o continuo, en un conjunto de grupos (*clusters*) basándose en la similitud entre las instancias. Por su parte, la regresión es igualmente una tarea de aprendizaje inductivo, que permite predecir valores numéricos en lugar de etiquetas de clase discretas; consiste por tanto en asignar valores numéricos a instancias de un dominio dado, tanto de valor discreto como de valor continuo.

Podría decirse que la clasificación y la regresión se resuelven con métodos de aprendizaje supervisado, mientras que el agrupamiento lo hace mediante métodos no supervisados. Así, en la Ilustración 1-2a se muestra un ejemplo de clasificación mediante aprendizaje supervisado, en la que se conoce las clases a la que pertenecen

los datos (cruces rojas y círculos azules). Por otra parte, en la Ilustración 1-2b se muestra un ejemplo de aprendizaje no supervisado, que respondería a un proceso de agrupamiento, donde no se conoce la clase a la que pertenecen los datos disponibles y se buscan técnicas que los agrupen en función de sus características.

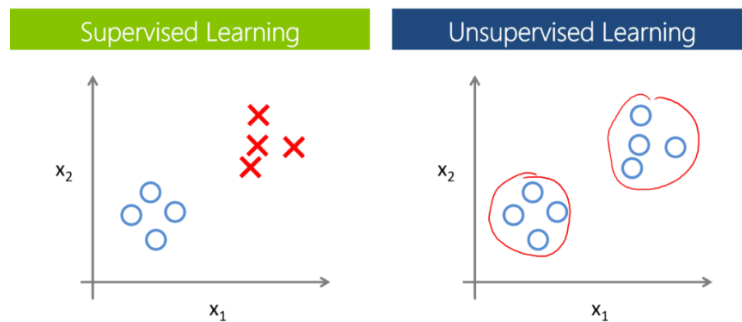


Ilustración 1-2. a) Aprendizaje supervisado; b) Aprendizaje no supervisado, tomada de (Torres, 2020)

1.3.1.3. Proceso de clasificación y tipos de clasificadores

Centrándonos en la **clasificación**, que es el proceso que se aplica fundamentalmente en este trabajo, hay distintos enfoques que se basan bien en una función discriminante que divida el espacio de las características (clasificadores estadísticos) o bien empleando reglas de pertenencia (clasificadores sintácticos) (Platero, 2008). En el primer caso, las funciones de decisión son las encargadas de particionar el espacio de características en distintas clases o grupos representativos. El ejemplo más simple de estas funciones de decisión es el que se expresa mediante una relación lineal entre características y etiquetas (Torres, 2020):

$$y = w x + b \quad (\text{Ecuación 1-1})$$

Siendo:

y: la etiqueta del ejemplo de entrada.

x: la característica.

w: la pendiente de la recta que se denomina peso y es uno de los parámetros que se aprenderán durante el entrenamiento.

b: el punto de intersección de la recta en el eje, llamado sesgo. Este parámetro también debe ser aprendido por el modelo.

Normalmente, no se tendrá solo una característica, sino varias, por lo que de manera general la ecuación puede expresarse como:

$$y = \sum_i w_i x_i + b \quad (\text{Ecuación 1-2})$$

Este tipo de funciones y otras más complejas de tipo potencial, cuadrático, etc., están en la base de clasificadores tan diferentes como el análisis discriminante, las máquinas de vectores de soporte o las propias redes neuronales en todos sus tipos, que generalmente pueden aplicarse también en problemas de regresión. Por el contrario, dentro de los clasificadores sintácticos se pueden citar los árboles de decisión o los bosques aleatorios. Además de estos, hay otros métodos basados de una u otra forma en la proximidad o la probabilidad de pertenecer a una clase, como como los clasificadores de mínima distancia, máxima probabilidad o los K-vecinos más próximos. Todos ellos se integran junto a los basados en una función discriminante en los métodos paramétricos (Platero, 2008).

Desde un punto de vista práctico, todos los métodos de aprendizaje automático supervisado se caracterizan por presentar unas fases bien diferenciadas en los procesos de clasificación – y, en su caso de regresión –, que comienzan con entrenamiento, para lo que se emplea una muestra de la población. Con el entrenamiento se construye el modelo, para lo que se emplean unos parámetros iniciales que se van ajustando de una forma iterativa, con diversas técnicas que dependen de la técnica empleada. Una vez entrenado y ajustados los parámetros del modelo, se procede a la validación del mismo, mediante una muestra independiente de la de entrenamiento.

Por lo tanto, con carácter previo se necesita realizar una **partición de los datos** en diferentes conjuntos, que aún con diferencias entre los distintos investigadores (Ripley, 1996; Russell & Norvig, 2010; Torres, 2020), pueden clasificarse en: datos de entrenamiento (*training*), datos de validación y datos de prueba (*test*), como se muestra en la Ilustración 1-3:

- Conjunto de datos de entrenamiento: Son los datos que entrenan los modelos. La calidad del modelo va a ser directamente proporcional a la calidad de estos datos.
- Conjunto de datos de validación: Son realmente una parte de los datos de entrenamiento, no son un conjunto en sí. Con ellos se ajustarán los hiperparámetros del modelo y se podrá seleccionar el mejor de los modelos entrenados.

- Conjunto de datos de test: Son datos reservados e independientes, que no han pasado por el entrenamiento y servirán para comprobar si la predicción del modelo entrenado es correcta, en definitiva, si el clasificador funciona correctamente.



Ilustración 1-3. División del conjunto de datos típica para Machine Learning.

De hecho, en muchas ocasiones, se simplifica este esquema y se habla únicamente de datos de entrenamiento y datos de validación o test (James, et al., 2017; Bosch, et al., 2019). La proporción entre unos y otros es variable, aunque un esquema bastante común es de 60/20/20, respectivamente. En caso de que solo se consideren dos conjuntos, la proporción varía entre 60/40, 70/30 o 80/20.

La existencia de conjuntos de datos independientes para el entrenamiento, el ajuste y la validación previene contra uno de los principales problemas del aprendizaje automático que es el sobreajuste, también conocido como sobre-entrenamiento o en inglés *overfitting*. Este fenómeno aparece como efecto de sobreentrenar el modelo de forma que este llega a ser tan complejo que solo se ajustará a los casos particulares de entrenamiento, siendo incapaz de clasificar correctamente nuevos datos de entrada (Ilustración 1-4). Así, cuando se entrena con una gran cantidad de datos se pueden introducir muestras atípicas o con ruido, o muestras que pueden no ser del todo representativas.

Para prevenir el sobreajuste es necesario:

- Tener una cantidad de muestras suficiente para cada uno de los conjuntos de entrenamiento, test y validación.
- Clases variadas y equilibradas: Los datos de entrenamiento deben de estar balanceados, es decir, aproximadamente el mismo número de datos por clase.

- Considerar una cantidad de dimensiones apropiada: Esta cantidad viene dada por el número de atributos o características consideradas para el modelo y, si el número de características es demasiado alto, se puede incurrir en sobreajuste.
- Conjunto de datos de test: Es necesario contar con un conjunto de datos proveniente de la división del espacio de datos que se mantenga oculto al entrenamiento.

También puede aparecer el problema contrario, el subajuste o *underfitting*. En él se tienen tan pocas muestras de entrenamiento que el modelo no realizará una buena predicción, por falta de ajuste de los datos (Ilustración 1-4). Por ejemplo, si tan solo se obtiene un buen acierto para los datos de test para una clase y para el resto no, será síntoma de sub-ajuste en las clases mal clasificadas. En este caso, no hay suficientes datos para realizar un ajuste del modelo.

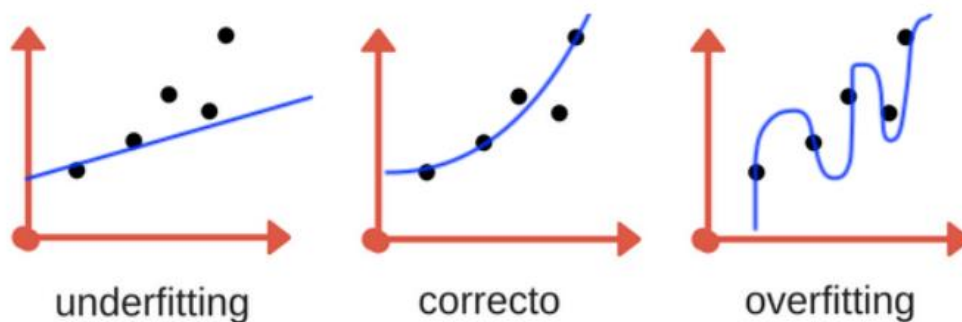


Ilustración 1-4. Muestra del *underfitting*, *overfitting* y ajuste correcto (tomada de <https://docplayer.es/63777861-Sobreajuste-overfitting.html>).

Hay distintas formas de aplicar la **validación**, que básicamente se pueden agrupar en las siguientes (Torres, 2020):

- Validación de retención (*hold-out*): Es el esquema más simple que consiste en dividir el conjunto de datos en dos: datos de entrenamiento y datos de validación, de acuerdo a los porcentajes antes indicados (por ejemplo, 80/20 o 70/30).
- Validación cruzada (*cross validation*) (Devijver & Kittler, 1982; James, et al., 2017; Bosch, et al., 2019): Se realiza dividiendo los datos de entrenamiento en un número k de grupos (5 o 10 generalmente). Así, uno de estos grupos se utiliza como datos de validación mientras que se entrena con el resto (4

en el caso de 5 grupos, es decir una proporción de 80/20; o 9 en el caso de 10 grupos, con una proporción de 90/10), repitiendo el proceso k -veces (*k-folds*). La exactitud final será una media de la exactitud en cada iteración. En este esquema se entrenará con todos los datos del muestreo de entrenamiento, como ocurre en el caso anterior, pero también se validará con todos ellos, por lo que dará una buena aproximación de la exactitud en la predicción del modelo. La desventaja es que requiere varias iteraciones por lo que es necesario más tiempo de ejecución y más potencia computacional, y así está recomendado para conjunto de datos pequeños.

Sea con un esquema u otro, la validación de los modelos de aprendizaje para clasificación se realiza mediante la matriz de confusión y todos los parámetros que se derivan de ella (Congalton, 1991; Ting, 2011). Por su parte, en los modelos de regresión las métricas son diferentes, y se definen el error absoluto medio, el error cuadrático medio o su raíz, etc. (Bosch, et al., 2019).

La matriz de confusión tiene diversas expresiones y definiciones, dependiendo de diversos criterios, como el hecho de emplear clasificaciones binarias (presencia/ausencia de un determinado elemento) o bien clasificaciones con un mayor número de clases (3 o más). En el caso de utilizar una clasificación binaria, la expresión de la matriz de confusión es bastante sencilla y permite definir claramente las celdas resultantes, así como los parámetros o índices que se derivan de ella (Ilustración 1-5).

		Predicho	
		P'	N'
Real	P	VP	FN
	N	FP	VN

Ilustración 1-5. Matriz de confusión en la clasificación binaria.

Siendo:

- P y N: Instancias positivas y negativas en la realidad;
- P' y N': Instancias predichas como positivas y negativas en la realidad.
- Verdaderos positivos (VP): Instancias que siendo positivas se clasifican correctamente como positivas.

- Falsos negativos (FN): Instancias que, siendo positivas, se clasifican erróneamente como negativas. También se conocen como errores de omisión.
- Falsos positivos (FP): Instancias que, siendo negativas, se clasifican erróneamente como positivas. También se conocen como errores de comisión.
- Verdaderos negativos (VN): Instancias que, siendo negativas, se clasifican correctamente como negativas.

A partir de ellas, se definen diversos índices, como:

- Exactitud (*accuracy*) o porcentaje de acierto: Razón de las instancias clasificadas correctamente, respecto al total.

$$Exac = \frac{VP+VN}{P+N} \quad (\text{Ecuación 1-3})$$

- Sensibilidad o recuerdo (*sensitivity o recall*): Razón de verdaderos positivos respecto al total de positivos reales.

$$Sen = \frac{VP}{P} \quad (\text{Ecuación 1-4})$$

- Especificidad (*specificity*): Razón de verdaderos negativos respecto a los negativos totales:

$$Esp = \frac{VN}{N} \quad (\text{Ecuación 1-5})$$

- Precisión (*precision*): Relación de verdaderos positivos respecto al total de positivos predichos:

$$Prec = \frac{VP}{VP+FP} \quad (\text{Ecuación 1-6})$$

Todos ellos se pueden expresar en tanto por 1 o más frecuentemente en porcentaje.

A partir de este esquema, se pueden calcular algunos otros gráficos y parámetros como la curva ROC, que se construye representando las parejas de 1 – especificidad (en el eje X) y la sensibilidad (en el eje Y), que resulta de la clasificación

binaria en función de distintos umbrales (por ejemplo, en pasos de 0,1). A partir de la curva se pueden determinar parámetros como el área bajo la curva (AUC) ROC.

Sin embargo, en el caso en el que la clasificación de lugar a más de dos clases, la matriz de confusión y, sobre todo, los parámetros derivados de ella, cambian en cierta medida. Así la forma de la matriz sería la que se muestra en la Ilustración 1-6 (Bosch, et al., 2019):

		Predicho		
		A'	B'	C'
Real	A	AA'	AB'	AC'
	B	BA'	BB'	BC'
	C	CA'	CB'	CC'

Sens.	E.Om.
AA'/A	1-SA
BB'/B	1-SB
CC'/C	1-SC

Prec.	AA'/A'	BB'/B'	CC'/C'
E. Co.	1-PA	1-PB	1-PC

Ilustración 1-6. Matriz de confusión en la clasificación con más de dos clases.

En este caso, la exactitud o porcentaje de acierto global viene definido por la suma del número de instancias en la diagonal principal, es decir, las instancias clasificadas correctamente, respecto al total de instancias. Por su parte el error global viene dado por el número de instancias fuera de la diagonal principal, es decir, las clasificadas incorrectamente; o 1 (100%) menos el porcentaje de acierto global.

Además, se pueden definir la sensibilidad y la precisión de cada clase, así como los errores de omisión o comisión correspondientes (Ilustración 1-6). La sensibilidad y precisión globales, se definen como la media ponderada de las sensibilidades y precisiones de cada clase, en función del número de instancias reales o predichas de cada una de ellas, respectivamente.

1.3.1.4. Métodos de clasificación en aprendizaje automático

En cuanto a los **métodos de aprendizaje automático**, a continuación, se resumen los más empleados, abarcando desde métodos estadísticos multivariantes hasta las redes neuronales más complejas.

Análisis discriminante: El análisis discriminante es una herramienta clásica del análisis multivariante (Johnson & Wichern, 2007; James, et al., 2017), que, no obstante, se suele encontrar en las librerías de programación de aprendizaje automático. Esta técnica se basa en maximizar la separación entre grupos, como el análisis de la varianza (ANOVA), lo que permite predecir a qué grupo pertenece un determinado caso basado en sus características. Parte de la premisa de que los grupos pueden ser representados por distribuciones normales multivariantes. Dentro de él, es muy común utilizar el análisis discriminante lineal (*linear discriminant analysis*, LDA), que asume que las covarianzas son iguales entre los grupos y utiliza funciones lineales para la clasificación.

El procedimiento comienza por el cálculo de las medias y matrices de covarianza para cada grupo, a lo que sigue la construcción de la función discriminante, mediante el desarrollo de una función que maximiza la razón de varianza entre los grupos respecto a la varianza dentro de los grupos. Finalmente, para la clasificación, se aplican los coeficientes de la función discriminante a las observaciones nuevas para predecir su grupo. Tiene aplicaciones en campos como la medicina (diagnóstico de enfermedades), el marketing (segmentación de clientes) y las finanzas (análisis de crédito), así como en la clasificación de imágenes.

Regresión logística (logistic regresión, LR): Al igual que la regresión lineal multivariante y el análisis discriminante, es una técnica estadística que, como las anteriores, se incluyen dentro de los paquetes informáticos de aprendizaje automático. La regresión logística se utiliza para modelar la relación entre una variable dependiente categórica y una o más variables independientes, que pueden ser continuas o categóricas (James, et al., 2017). Es especialmente útil cuando la variable dependiente tiene solo dos categorías (binaria), como "sí" o "no".

A diferencia de la regresión lineal, que predice valores continuos, la regresión logística utiliza la función logística (o sigmoide) para transformar la salida en una probabilidad entre 0 y 1. La fórmula es:

$$P(Y = 1/X) = \frac{1}{1 + e^{-(B_0 + B_1X_1 + B_2X_2 + \dots + B_nX_n)}}$$

(Ecuación 1-7)

Para la estimación de parámetros se utiliza el método de máxima verosimilitud, en lugar de los mínimos cuadrados usados en la regresión lineal. La regresión logística es fundamental en situaciones donde el resultado es categórico, ofreciendo un enfoque que se adapta mejor a la naturaleza de los datos que la regresión lineal. Se emplea en diversas aplicaciones, desde la medicina hasta el marketing, la modelización de procesos naturales basados en parámetros continuos y categóricos, y la clasificación de imágenes.

Árboles de decisión (*decision trees*, DT): Son un modelo de predicción usado en ámbitos muy diversos (Breiman, et al., 1986; Hastie, et al., 2009). Dado un conjunto de datos se fabrican diagramas de construcciones lógicas, muy similares a los sistemas de predicción basados en reglas, que sirven para representar y categorizar una serie de condiciones que ocurren de forma sucesiva, para la resolución de un problema. Se emplean en clasificación, pero también en regresión en lo que se conoce como la técnica de árboles de clasificación y regresión (*classification and regression trees*, CART).

Los árboles de decisión están formados por nodos, vectores de número, flechas y etiquetas. Su lectura se realizará de arriba hacia abajo. Existen diferentes tipos de nodos:

- Primer nodo o nodo raíz: En él se produce la primera división en función de la variable más importante.
- Nodos internos o intermedios: Se encuentran tras la primera división y volverán a dividir el conjunto de datos en función de las distintas variables.
- Nodos terminales u hojas: Se ubican en la parte inferior del esquema y su función es indicar la clasificación definitiva.

Otros conceptos interesantes son la profundidad del árbol, que viene determinada por el número máximo de nodos de una rama; o los criterios de división en cada nodo:

- Índice de diversidad de Gini (*Gini's diversity index*)
- Regla de dos (*Twoing rule*)
- Reducción máxima de la desviación (*maximum deviance reduction*)

Los árboles de decisión son sencillos de interpretar, con muy bajos tiempos en entrenamiento y predicción y con requisitos bajos de memoria. Pero su desventaja es que tienen una baja precisión predictiva, y sufren sobre-ajuste fácilmente, conforme aumentan el número de ramificaciones.

Bosques aleatorios: Generalmente conocidos por su término en inglés, *random forest* (Breiman, 2001; Hastie, et al., 2009), son un método de aprendizaje automático utilizado principalmente para problemas de clasificación y regresión. Se basan en la idea de crear un conjunto de árboles de decisión (un "bosque") y combinar sus predicciones para mejorar la precisión y robustez del modelo. Cada árbol se entrena sobre una muestra aleatoria de los datos (*bootstrap*), lo que permite que cada árbol aprenda patrones diferentes de los datos. En cada nodo de un árbol, se selecciona un subconjunto aleatorio de características para determinar la mejor división. Esto aumenta la diversidad entre los árboles y reduce el riesgo de sobreajuste.

En problemas de clasificación, cada árbol emite un voto para la clase de la observación. La clase final predicha es aquella que recibe la mayoría de los votos de todos los árboles del bosque (Ilustración 1-7). Esta técnica es menos propensa al sobreajuste en comparación con un solo árbol de decisión, debido a la combinación de múltiples modelos y la aleatoriedad en su construcción, lo que la hace muy efectiva para conjuntos de datos grandes y complejos. Además, calcula la importancia de cada variable en la clasificación, identificando las características o variables más relevantes para el modelo, lo que permite emplearla como técnica de clasificación y selección de variables. Los bosques aleatorios son ampliamente utilizados en diversas áreas, incluyendo, la medicina, finanzas, modelización de procesos naturales y análisis de imágenes.

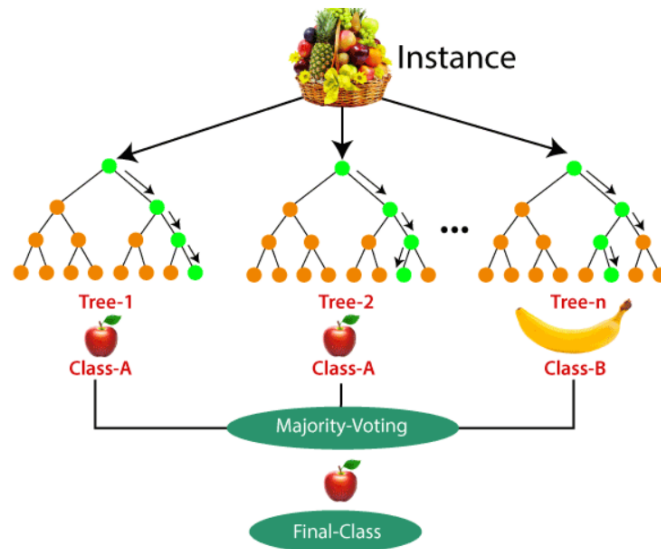


Ilustración 1-7. Esquema de trabajo de los bosques aleatorios (tomada de <https://insideaiml.com/blog/Random-Forest-Algorithm-1029>)

K-vecinos más próximos: Más conocido por el término en inglés, *k-Nearest Neighbors* (k-NN) (Hastie, et al., 2009), es uno de los algoritmos de clasificación más conocidos dentro del aprendizaje supervisado que sirve para estimar la función de densidad $F(x/C_j)$ de las predictoras x por cada Clase C_j . El algoritmo se llama así porque clasifica cada nuevo ejemplo calculando la distancia de ese ejemplo con todos los del conjunto de entrenamiento. La clase predicha para este nuevo ejemplo viene dada por la clase a la que pertenezcan los ejemplos más cercanos. El valor de k determina en cuantos vecinos se tendrán en cuenta para perder la clase. Así con un valor de $k=1$, la clase predicha para cada ejemplo será la clase a la que pertenezca el ejemplo.

Lo primero que hace el algoritmo k-NN es calcular la distancia de cada ejemplo nuevo a clasificar con todos los del conjunto de entrenamiento. Los ejemplos tienen una serie de atributos, donde cada atributo tiene un rango de valores que puede ser muy distinto del rango de valores de otro atributo, con lo cual se hace necesario normalizar todos los atributos para que su rango de valores esté entre 0 y 1, de forma que todos los atributos afectan por igual al cálculo de la distancia. Una vez todos los atributos tienen el mismo rango de valores se puede medir la distancia de cada nuevo ejemplo debidamente. Para calcular la distancia se pueden usar diferentes métodos como la distancia euclidiana, que es la usada generalmente, o la distancia Manhattan, Canberra, máxima, etc.

Máquinas de vectores de soporte, también suelen ser conocidas por su nombre y siglas en inglés (*Support Vector Machines*, SVM), fue desarrollado en la década de los 90, dentro de campo de la ciencia computacional (Vapnik, 1995; Hastie, et al., 2009). Si bien originariamente se desarrolló como un método de clasificación binaria, su aplicación se ha extendido a problemas de clasificación múltiple y regresión. SVM ha resultado ser uno de los mejores clasificadores para un amplio abanico de situaciones, por lo que se considera uno de los referentes dentro del ámbito de aprendizaje estadístico y automático (*machine learning*).

Intuitivamente, SVM es un modelo de aprendizaje supervisado que, partiendo de un conjunto de entrenamiento, permite etiquetar en diferentes clases y representar dichas muestras en puntos en el espacio que separen las diferentes clases mediante un espacio lo más amplio posible. Así, cuando las nuevas muestras de los casos de test se pongan en correspondencia con dicho modelo puedan ser clasificadas correctamente en función de su proximidad. Es el concepto de separación óptima donde reside la característica fundamental de SVM, que busca encontrar el hiperplano que maximice el margen m entre las clases de este espacio como se observa en la Ilustración 1-8.

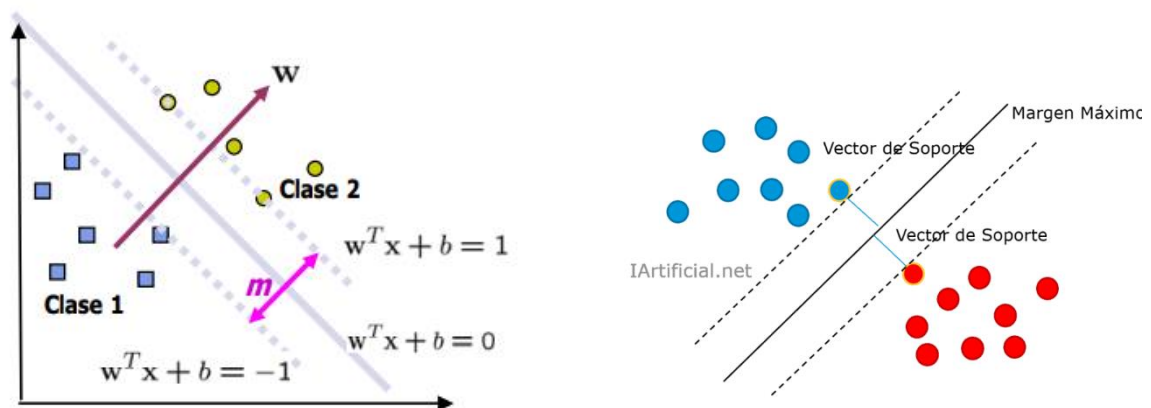


Ilustración 1-8. Support Vector Machines (SVM, tomada de https://www.ml2projects.com/post/svm_kernlab).

El nombre SVM (Máquinas de Vectores de Soporte) viene de máquina debido a que es un método de *machine learning* y los vectores de soporte que definen el margen máximo de separación del hiperplano que separa las clases. El hiperplano del espacio n -dimensional tendrá $n-1$ dimensiones.

El clasificador SVM descrito anteriormente consigue buenos resultados cuando el límite de separación entre clases es aproximadamente lineal. Si no lo es, su capacidad decae drásticamente. En la mayoría de ocasiones lo normal es que los problemas no sean tan sencillos, por lo que no se podrá encontrar un hiperplano que separe las clases, debido a que las curvas de separación no sean lineales, que existan más de dos clases o que los datos no sean completamente separables. Para resolver este problema una solución es expandir las dimensiones del espacio original. El hecho de que los grupos no sean linealmente separables en el espacio original no significa que no lo sean en un espacio de mayores dimensiones como se muestra en la Ilustración 1-9.

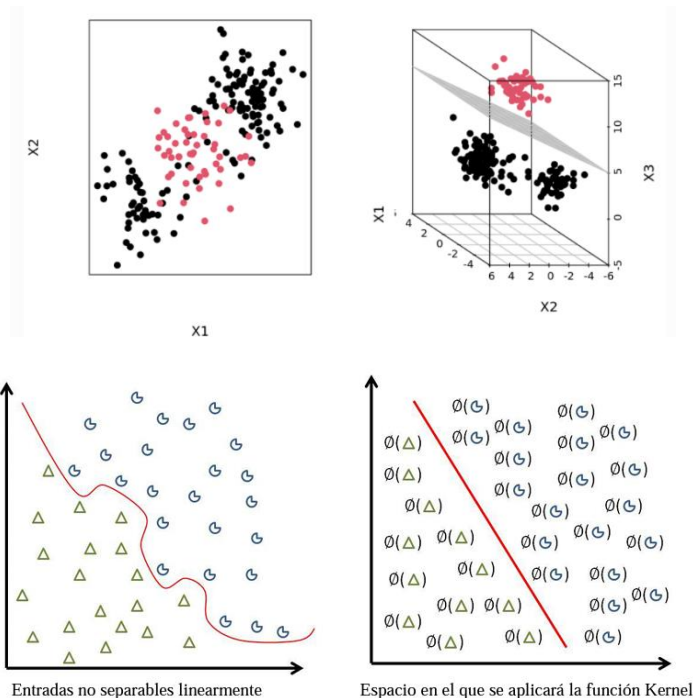


Ilustración 1-9. Importancia de las dimensiones en SVM (tomada de https://www.cienciadedatos.net/documentos/34_maquinas_de_vector_soporte_support_vector_machines).

1.3.1.5. Redes neuronales y aprendizaje profundo

Son un conjunto especial dentro de las técnicas de aprendizaje son las redes neuronales artificiales (Artificial Neural Networks, ANN) (Torres, 2020; Bosch, et al., 2019). Se definen como unas estructuras algorítmicas que permiten crear modelos formados por múltiples capas de procesamiento con múltiples niveles de abstracción que realizarán una serie de transformaciones lineales y no lineales, en las que, a partir de unos datos de entrada, se genere una salida próxima a la esperada. El hecho de que se utilicen múltiples capas y niveles de abstracción es por lo que estas estructuras

son conocidas como aprendizaje profundo. No obstante, la base de muchas de estas redes está en la unidad funcional que se conoce como perceptrón o neurona simple (Ilustración 1-10).

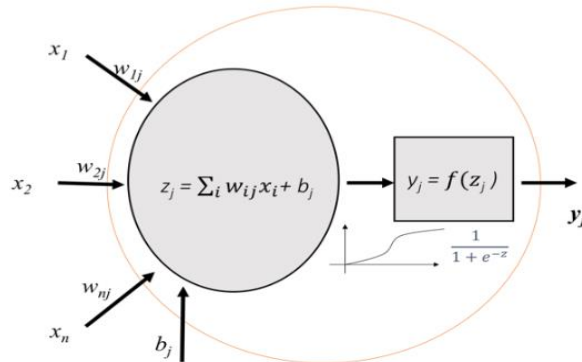


Ilustración 1-10. Funcionamiento de una neurona simple o perceptrón (tomada de Torres, 2020).

El perceptrón funciona como un modelo de regresión lineal simple. Tras la observación de muchos datos, calcula los parámetros de pesos y sesgos (w y b). El cálculo de los parámetros consiste en un proceso de aprendizaje iterativo para todos los datos de entrenamiento, en los que se comparan la etiqueta real y la obtenida a través del modelo. Tras cada iteración se irán ajustando los pesos para que el error se minimice. A continuación, la neurona aplica el vector de pesos calculados sobre los valores del elemento “ x ” de entrada y tras sumar el sesgo b , se calcula el valor de z en lo que se conoce como función de propagación. El resultado se envía a una función de activación que producirá un resultado de 0 o 1. La función de la neurona artificial queda definida con las siguientes expresiones:

$$z = b + \sum_i w_i x_i$$

$$y = f_{act}(z)$$

(Ecuación 1-8)

Las funciones de activación se usan para propagar hacia adelante la salida de una neurona, de manera, que, si hay varias capas de neuronas conectadas, la salida de esta la recibirá la siguiente hasta la salida. El objeto principal de las funciones de activación (excepto la lineal) es agregar no linealidad a la salida de las neuronas. Hay distintas funciones de activación, entre las que destacan (Torres, 2020; Bosch, et al., 2019):

- Linear: Es una función identidad en la que la señal no cambia.
- Sigmoide: Permite reducir valores extremos o atípicos sin eliminarlos. Para valores altos y positivos de z , la y tomará un valor de 1; para valores altos y negativos de z , resultará $y = 0$.
- Tanh-: Representa la relación entre el seno y el coseno hiperbólico, produciendo un rango normalizado entre -1 y 1, que es la entrada que le va bien a algunas redes neuronales.
- Softmax: Para cada dato de entrada, se obtendrá la probabilidad de pertenencia a cada una de las posibles clases. Funciona calculando las evidencias de que un determinado dato pertenezca a una clase, a partir de los datos de entrenamiento. Una vez se han calculado las evidencias se deben de convertir estas en probabilidades cuya suma de todos sus componentes suma 1. Para ello softmax usa el valor exponencial de las evidencias calculadas y luego las normaliza de modo que sumen 1, formando así una distribución de probabilidad.
- ReLU (*Rectified Linear Unit*): Es una función que solo se activa si la entrada está por encima de un cierto umbral. Normalmente, el umbral es el 0, siendo la salida 0 para una entrada menor que 0 (negativa). Cuando la entrada se eleva por encima del umbral la salida será una relación lineal.

El aprendizaje para el cálculo de los parámetros se realiza mediante un proceso iterativo de ida (*forwardpropagation*) y vuelta (*backpropagation*) para todos los datos de entrenamiento, en los que se compara la etiqueta real y la obtenida a través del modelo. Tras cada iteración se irán ajustando los pesos para que el error se minimice. Así, el entrenamiento se puede dividir en tres fases (Torres, 2020) (Ilustración 1-11):

- *Forwardpropagation*: Una vez se introducen los datos de entrenamiento estos atravesarán toda la red. Cada neurona recibirá unos valores de entrada y realizará una transformación dando un resultado a la salida que se transmite a la siguiente. Cuando los datos han cruzado todas las capas y las neuronas han realizado sus cálculos, se llega a la capa final que dará un resultado de predicción de la clase a la que pertenece cada ejemplo de entrada.

- A continuación, la función de pérdida (LOSS), que se definirá más adelante, comparará la predicción con el resultado correcto, ya que los datos introducidos están etiquetados. Lo ideal será que el coste sea 0, es decir, no haya diferencia entre el resultado real y la predicción. Por eso, a medida que se entrena el modelo se irán ajustando los pesos de las interconexiones de las neuronas. Para ello se utilizará la función descenso del gradiente (*gradient descent*).
- Una vez se tiene el resultado de la función LOSS, el resultado se retropropaga (*backpropagation*) a todas las neuronas de la capa oculta que contribuyen directamente a la salida. Las neuronas de la capa oculta reciben una fracción de la señal total de LOSS, basándose en la contribución que haya aportado cada neurona a la salida original.

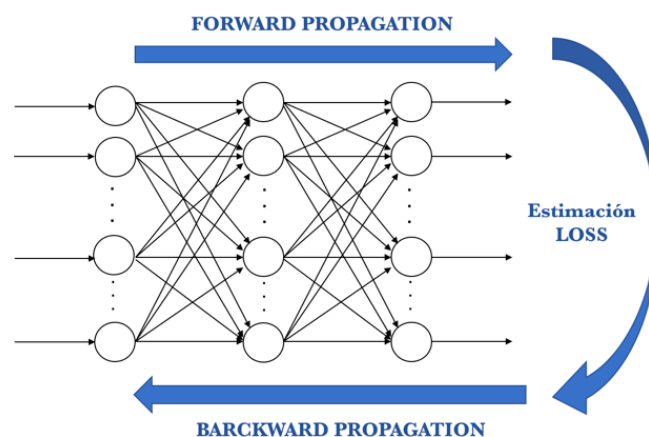


Ilustración 1-11. Fases del entrenamiento de las neuronas (tomada de Torres, 2020).

La función de pérdida, coste, o error (*loss function*) indica lo cerca que se encuentra la predicción de la solución real. La función de pérdida se calcula realizando una media de los costes de los N datos de entrenamiento, según la expresión:

$$C(w, b) = \frac{1}{N} \sum_{i=1}^N C_i(w, b)$$

(Ecuación 1-9)

Donde:

w y b representan los pesos y sesgos respectivamente.

C_i representa la función coste para la instancia i del conjunto de entrenamiento.

Cuanto más próximo sea el valor de la función de pérdida a 0, mejor será la estimación. El valor de la función de pérdida se puede calcular a partir del error absoluto medio, el error cuadrático medio o su raíz (*Root Mean Square Error*, RMSE), la entropía cruzada categórica, etc.

No obstante, es frecuente y conveniente la aplicación de algoritmos de optimización como el descenso del gradiente, que se trata de un método iterativo que sirve para encontrar los mínimos locales de la función LOSS. El descenso de gradiente usa la primera derivada (gradiente) de la función LOSS cuando realiza la actualización de los parámetros. Una vez todas las capas tienen su valor del gradiente de la función LOSS que les corresponde, se actualizan los parámetros en sentido contrario a la que indica el gradiente. En la Ilustración 1-12 se muestra el proceso de actualización de los pesos (Torres, 2020). Se comienza por un punto inicial y tras calcular el gradiente se modifica el valor del peso en sentido contrario al gradiente (gradiente negativo). La magnitud de este cambio de peso está especificada por el parámetro de tasa de aprendizaje o *learning rate*. El algoritmo repite este proceso hasta que se alcanza un mínimo, donde la función LOSS no puede decrecer más.

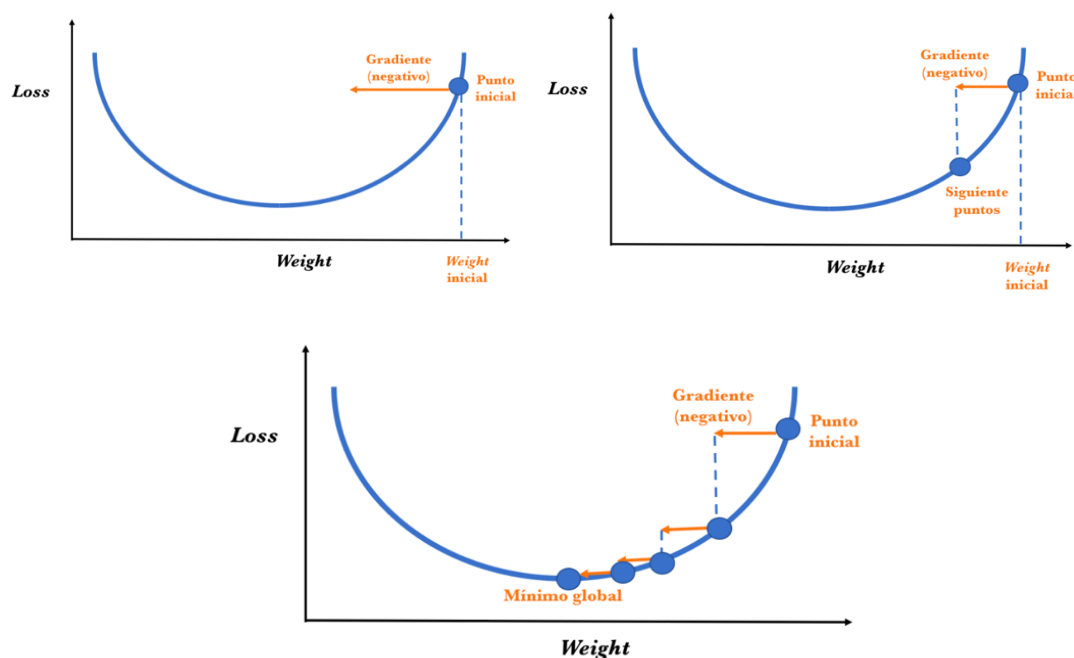


Ilustración 1-12. Descenso del gradiente (tomada de Torres, 2020).

La tasa o ratio de aprendizaje (*learning rate*), que formará aparte de los hiperparámetros que ahora se describirán, define cuanto se avanza en cada paso, es

decir, cuanto afecta el gradiente a la actualización de los parámetros en cada iteración. Un *learning rate* muy pequeño provocará una convergencia extremadamente lenta y, por el contrario, si el valor es muy alto el algoritmo no llegará a converger.

Tras obtener el conjunto de datos habrá que definir los hiperparámetros de la red neuronal. Los hiperparámetros son aquellos valores modificables de los que dependerán el entrenamiento del modelo y que no son obtenidos a partir del conjunto de datos. Modificarán los pesos y sesgos de las diferentes capas. Los hiperparámetros más importantes son (Torres, 2020):

- Selección del tamaño de *batch*: Conjunto de datos que se procesan en cada iteración de cada época. La mejor elección del tamaño del *batch* será por tanto, la máxima que se pueda soportar computacionalmente (aunque el proceso se ralentice).
- Tasa de aprendizaje o *learning rate*. Indica la rapidez de ajuste del modelo al problema, como se ha descrito previamente.
- Iteración y época. La iteración es una actualización de los pesos del modelo, que se define como el número de lotes de datos que el algoritmo ha visto. Las épocas es el número de veces que la red procesa el conjunto de datos completo en el proceso de entrenamiento. Al aumentar el número de épocas, los pesos asociados a las distintas capas se actualizarán en más ocasiones, por lo que será una solución para resolver el problema de *underfitting*, aunque si el aumento es muy grande se puede llegar a sobreajuste u *overfitting*.

Como se viene apuntando, las redes neuronales están formadas por varias, y en ocasiones, numerosas neuronas, formando estructuras más o menos complejas. Una de ellas es la conocida como Perceptrón Multicapa (*Multi-Layer Perceptron*, MLP), que es una estructura formada por una capa de entrada de varios perceptrones, uno o más capas de compuestas por perceptrones, llamadas capas ocultas (*hidden layers*) y una capa final con varios perceptrones llamada la capa de salida. Visualmente se puede presentar con el siguiente esquema (Ilustración 1-13) (Torres, 2020):

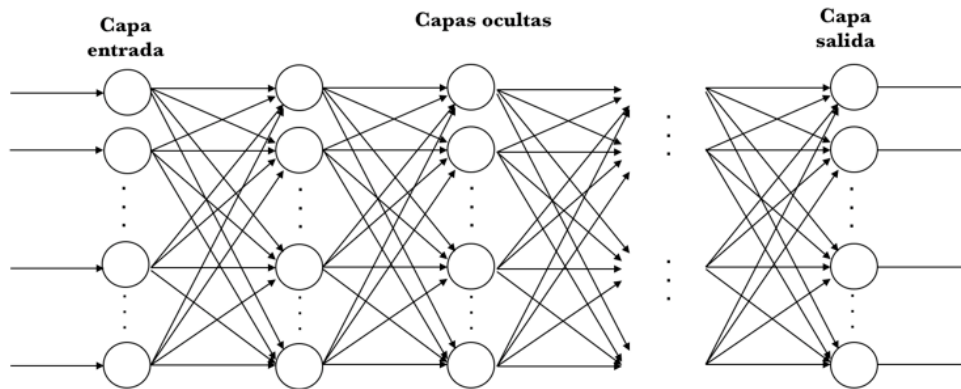


Ilustración 1-13. Perceptrón Multicapa (MLP, tomada de Torres, 2020).

Todas las salidas de cada una de las neuronas de una capa estarán conectadas con todas las entradas de las neuronas de la siguiente. Este tipo de red se conoce como densamente conectada o *fully connected*. Se aplica a la salida de las neuronas de la capa de salida una de las funciones de activación anteriormente mencionadas. Si las clases son excluyentes, normalmente se utiliza softmax.

Otros tipos de estructuras o arquitecturas de red más complejas dan lugar a los tipos de redes convolucionales, recurrentes, radiales, etc. Así, las **redes neuronales convolucionales** (*convolutional neural networks*, CNN) son muy usadas en tareas de visión por computador y son las que han permitido grandes avances en el campo de la inteligencia artificial. La CNN es un tipo de red neuronal artificial con aprendizaje supervisado que procesa sus capas imitando al córtex visual del ojo humano para identificar distintas características en las entradas que permiten identificar objetos, es decir “ver”. Las redes neuronales convolucionales se encuentran dentro del área del aprendizaje profundo o *deep learning*. Fueron desarrolladas primeramente a finales de los 90, aunque recientemente se han popularizado por sus buenos resultados en el análisis y reconocimiento de la imagen, provocando un gran avance en el campo de la visión por computador (Torres, 2020). Las redes convolucionales se pueden aplicar en distintas áreas y aplicaciones tales como probabilidad, estadística, visión por computador, procesamiento del lenguaje natural, imagen y procesamiento de la señal, ingeniería y ecuaciones diferenciales.

Las CNN son muy similares a las MLP en lo básico: están formadas por neuronas que tienen parámetros en forma de sesgos y pesos que se pueden aprender. Sin embargo, presentan unos elementos característicos que las hacen muy útiles en el

reconocimiento de objetos en las imágenes u otros tipos de patrones. La operación básica en las redes convolucionales es la convolución. La convolución es una operación matemática sobre dos funciones (f y g) que produce una tercera función (s) que expresa como la forma de una es modificada por la otra. Usualmente, la operación de convolución se denota con un asterisco (*) (Bosch, et al., 2019):

$$s(t) = (f * g)(t) \int f(\tau) g(t - \tau) d\tau$$

(Ecuación 1-10)

En la terminología de las redes convolucionales, el primer argumento (la función f) de la convolución se refiere a la entrada (*input*) y el segundo argumento (la función g) es el filtro o *kernel*. La salida de la función (s) se refiere al mapa de características o *feature map*. Normalmente, los datos usados por las redes convolucionales son datos discretos, por lo que se aplica la convolución discreta:

$$s(t) = (f * g)(t) \sum_{\tau=-\infty}^{\infty} f(\tau) g(t - \tau) d\tau$$

(Ecuación 1-11)

En aplicaciones de aprendizaje automático la entrada es un vector multidimensional de datos y el *kernel* es un vector multidimensional de parámetros. En visión por computador, trabajamos en imágenes y las convoluciones se aplican sobre más de un eje. Por ejemplo, si usamos una imagen (I) de dos dimensiones como entrada, también usaremos un *kernel* (K) de dos dimensiones y, dado que la convolución cumple con la propiedad conmutativa, podemos usar la siguiente fórmula:

$$S(i, j) = I * K(i, j) \sum_m \sum_n I(i - m, j - n) K(m, n)$$

(Ecuación 1-12)

La operación de convolución se aplica mediante unas capas especializadas de la red, denominadas capas convolucionales. La diferencia principal entre una capa densamente conectada (como las que existen en las MLP) y una capa convolucional, es que la capa densa aprende patrones globales en su espacio global de entrada, mientras que las capas convolucionales aprenden patrones locales en pequeñas matrices (ventanas en las imágenes) de dos dimensiones (Torres, 2020).

Una característica importante es que las capas convolucionales pueden aprender con una determinada jerarquía. Por ejemplo, en el caso de las imágenes, la primera capa convolucional aprende elementos básicos como aristas, mientras una segunda capa puede aprender patrones compuestos de elementos básicos aprendidos en la capa anterior (Torres, 2020). Esto continúa sucesivamente hasta poder aprender patrones muy complejos que permiten reconocer formas complejas como un rostro o la silueta de un animal. Otra propiedad muy interesante es que, una vez aprendida una característica en una capa convolucional en un punto concreto de una imagen, posteriormente la podrá reconocer en cualquier parte de la misma. En una red densamente conectada tendría que aprender un patrón nuevamente cada vez.

Por su parte, las **redes neuronales recurrentes** (*recurrent neural networks*, RNN) son un tipo de arquitectura de red neuronal diseñada para trabajar con datos secuenciales, donde la información en una secuencia depende de elementos anteriores. Son especialmente efectivas en tareas donde el contexto temporal es importante, como el procesamiento del lenguaje natural y la predicción de series temporales.

En la mayor parte de las redes neuronales que se han descrito la información fluye siempre en una única dirección, desde las neuronas de entrada hacia las neuronas de salida, lo que se conoce como redes *feed-forward*. Sin embargo, en una red neuronal puede haber conexiones que vayan desde una capa posterior hacia una capa anterior. Así, en la Ilustración 1-14 se puede ver un ejemplo en el que la capa de neuronas ocultas está conectada a los datos de entrada y, a su vez, a la salida de las neuronas de la propia capa. De esta forma, cuando la red procesa un registro concreto está utilizando, a la vez, información del propio registro e información que generó la red con el registro anterior. Este tipo de conexiones pueden ser muy útiles para tratar secuencias de datos, como es el caso de las series temporales o los textos.

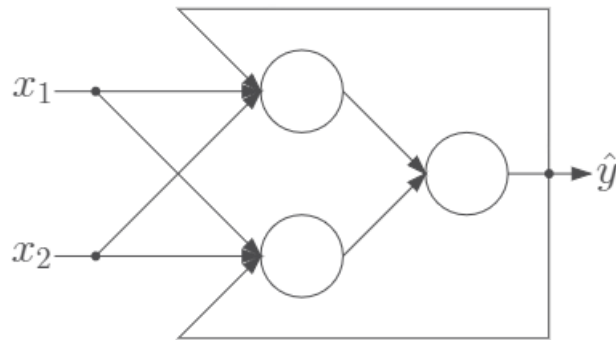


Ilustración 1-14. Esquema de una red neuronal recurrente (tomada de (Bosch, et al., 2019)).

La estructura básica de una red recurrente se compone de:

- Neuronas Recurrentes: Las RNN tienen conexiones que permiten que la información fluya no solo hacia adelante, sino también hacia atrás, creando ciclos en la red. Esto les permite mantener un "estado" que recuerda información pasada.
- Celdas de Memoria: Las RNN suelen utilizar celdas de memoria (como LSTM o GRU) que ayudan a mitigar el problema del descenso del gradiente, permitiendo que la red aprenda dependencias a largo plazo (Hochreiter & Schmidhuber, 1997).
- Capas: Típicamente, una RNN consta de capas de entrada, capas recurrentes y capas de salida, donde las capas recurrentes pueden ser apiladas para crear redes más profundas.

1.3.1.6. Selección de características y aprendizaje automático.

Por último, se va a hacer una breve mención a un procedimiento de especial importancia cuando se trabaja con conjuntos de datos de un elevado número de atributos o dimensiones, como las imágenes hiperespectrales, como es la selección de características. Este procedimiento se lleva a cabo mediante técnicas diferentes, algunas de ellas a su vez relacionadas con el aprendizaje automático.

La selección de características es un proceso que implica elegir un subconjunto de variables relevantes de un conjunto de datos para mejorar el rendimiento de modelos de aprendizaje automático (Guyon & Elisseeff, 2003). Este proceso busca

reducir la dimensionalidad, minimizar el riesgo de sobreajuste y mejorar la interpretabilidad del modelo.

Hay diversos métodos de selección de características:

- Métodos de filtro: Se basan en realizar una evaluación de características de manera independiente del modelo, empleando métricas estadísticas como la correlación y la información mutua. Como ejemplos se pueden citar la selección basada en p-valores o los coeficientes de correlación, como el factor de inflación de la varianza (*variance inflation factor*, VIF).
- Métodos de envoltura (**wrapper**): Evalúan el rendimiento del modelo usando diferentes subconjuntos de características, lo que implica un entrenamiento repetido del modelo. Como ejemplo se puede citar la eliminación recursiva de características (*Recursive Feature Elimination*, RFE).
- Métodos de incrustación (*embedded*): Integran la selección de características en el proceso de entrenamiento del modelo. Como ejemplos se pueden citar el LASSO, que utiliza regularización para penalizar características irrelevantes; o los métodos de clasificación de variables, que permiten evaluar además la importancia de las mismas, como los árboles de decisión o los bosques aleatorios, que en sí mismos se pueden considerar métodos de aprendizaje automático.

A las técnicas de selección de características hay sumar las de reducción de la dimensionalidad, dentro de las cuales destaca el análisis de componentes principales (en inglés *principal component analysis*, PCA) (Jolliffe, 2002). El PCA es un método de análisis que transforma un conjunto de datos con múltiples variables correlacionadas en un nuevo conjunto de variables no correlacionadas llamadas componentes principales. Estos son combinaciones lineales de las variables originales, y se ordenan de tal manera que el primer componente principal captura la mayor variabilidad en los datos y así sucesivamente hasta un número determinado de componentes. Otro método que está comenzando a emplearse de forma generalizada son las redes convolucionales, fundamentalmente a través de sus capas de *pooling*.

1.3.2. Aprendizaje automático y profundo en el análisis de imágenes

1.3.2.1. Visión por computador

El ser humano percibe la estructura tridimensional del espacio en torno a ellos con una cierta facilidad, al menos aparentemente. Se observan objetos, su número, sus colores y sus formas, lo que puede llevar incluso a adivinar emociones, como en el caso de la observación de personas (Szeliski, 2010). El funcionamiento de la percepción visual ha sido estudiado por los psicólogos perceptivos, aunque todavía siguen quedando bastantes aspectos por descifrar (Marr, 1982). En cualquier caso, la información visual tiene una gran importancia en la vida de las personas, particularmente en la toma de decisiones.

La visión artificial o visión por computador pretende emular a la percepción visual humana, de tal manera que, a partir de la realidad registrada por unos sensores y una serie de reglas establecidas en experiencias previas, se pueda interpretar la realidad y tomar decisiones al respecto. En los últimos años se han producido avances significativos que, no obstante, no han llegado a simular completamente al complejísimo proceso de la visión humana (Escalera, 2001; Szeliski, 2010). De esta manera, las investigaciones en visión artificial han ido desarrollando desde los años 1960 técnicas matemáticas y computacionales para extraer, describir e interpretar los objetos de la realidad a partir de las imágenes, tanto en 2 dimensiones, como en 3, e incluso incorporando el tiempo y el movimiento. Se trata de un ámbito de investigación multidisciplinar, en el que se combinan aspectos matemáticos, informáticos, de ingeniería electrónica y mecánica, e incluso fisiológicos, psicológicos, etc. Actualmente, se dispone de técnicas bien asentadas para:

- Detectar, diferenciar, describir e interpretar objetos en las imágenes utilizando diversas propiedades como el color, forma, tamaño, textura, etc.
- Disponer de modelos 3D a partir de imágenes parcialmente superpuestas (estereoscopía).
- Rastrear objetos en movimiento como una persona que se mueve en un contexto complejo.

En base a esto, de forma general, la visión artificial se puede definir como la ciencia que estudia la interpretación de imágenes mediante computadores digitales, y de forma más específica se hace referencia al proceso de extracción de información del mundo físico a partir de imágenes utilizando para ello un computador (Jiménez, 2020). En este sentido, la visión por computador es un concepto más amplio que el análisis de imágenes, que se puede entender como una subdisciplina dentro de aquel. Por lo tanto, mientras que el análisis de imágenes proporciona técnicas para descomponer y procesar imágenes, la visión por computador utiliza estos procesos para construir una comprensión más profunda y semántica del entorno visual, lo que permite a las máquinas interactuar con el mundo de manera más inteligente, especialmente en la toma de decisiones.

En cuanto a las aplicaciones, la visión por computador se aplica en una gran y creciente variedad de ámbitos (Escalera, 2001; Szeliski, 2010; Lowe, 2015):

- Detección automática y el seguimiento de objetivos militares, el análisis del terreno y las armas inteligentes.
- Control de calidad, mediante la inspección completa y la uniformidad en los criterios de inspección de piezas, productos, máquinas, etc.
- Robótica, para el reconocimiento de los objetos en el entorno de los robots, ya que los estos exigen una gran precisión en muchas de sus tareas.
- En la agricultura, el medio ambiente y la ingeniería civil, el análisis de imágenes aéreas o satelitales permite determinar los diversos tipos de terreno, vegetación, cultivos, bosques, zonas urbanas, comunicaciones, etc.
- En biomedicina hay muchas aplicaciones como la determinación del número de glóbulos rojos, la detección de tumores y otras patologías, a partir de imágenes de distinta procedencia (rayos X, resonancias, ultrasonidos, etc.).
- La identificación y diagnóstico, como en los analizadores de huellas dactilares o el iris, para permitir el acceso a instalaciones o sistemas.
- En el ámbito de la seguridad se utiliza en el reconocimiento facial o en la detección de intrusos en instalaciones o propiedades.
- En tráfico se identifican matrículas, vehículos, adelantamientos, infracciones, etc., así como los avances en los sistemas de conducción automática.

- Otros aspectos muy conocidos como la venta minorista utilizan sistemas como la lectura código de barras y reconocimiento de caracteres (OCR).
- Finalmente, aunque no menos importante, el mundo del ocio y especialmente de la animación, con aplicaciones como el *matchmove* para la fusión de imágenes generadas por computadora (CGI), y otras muy populares debido a su implementación en dispositivos móviles (*stitching, morphing, etc.*).

1.3.2.2. Imagen digital

En cualquier caso, todo parte del análisis de imágenes. Una imagen puede definirse como una distribución bidimensional de radiación electromagnética proveniente de una escena, reflejada por los objetos y capturada por un sensor (Escalera, 2001; Jiménez, 2020). El proceso de digitalización convierte esta imagen en una matriz de dimensiones $N \times M$, generalmente cuadrada, compuesta por píxeles, cada uno con un valor digital que refleja la luminosidad del punto correspondiente en la escena. El histograma de una imagen muestra la frecuencia de los niveles digitales registrados, permitiendo calcular parámetros estadísticos como la media, mínimo, máximo y desviación típica. Estos parámetros ayudan a caracterizar visualmente la imagen en términos de propiedades como el brillo y el contraste. El color de la imagen es resultado de la respuesta de los objetos a diferentes longitudes de onda de luz. Para generar imágenes en color, se requiere información de al menos tres bandas del espectro visible (rojo, verde y azul). El sistema RGB es el más utilizado para especificar colores, asignando colores a combinaciones de niveles digitales en las bandas.

No obstante, además de las imágenes, se están desarrollando métodos de representación 3D del mundo real, mediante sistemas como LiDAR y las técnicas fotogramétricas que generan modelos 3D a partir de imágenes estereoscópicas. Los videos, como secuencias de imágenes 2D, pueden acercarse a la representación 3D dependiendo del movimiento del objeto o de la cámara.

El procesamiento de imágenes en visión por computador comienza con la **adquisición de imágenes por sensores**, seguido de operaciones de realce y mejora. Luego se procede a la extracción de elementos, segmentación, clasificación y finalmente al reconocimiento de objetos y la interpretación de la escena.

La adquisición de las imágenes se lleva a cabo mediante un sistema de captura, que se basa en un sensor que registra radiación electromagnética. Los sensores van desde sencillas cámaras digitales B/N o RGB hasta sensores hiperespectrales. El proceso de adquisición de imágenes ha de tener en cuenta diferentes aspectos como la iluminación adecuada, el propio sistema óptico del sensor o el sistema de conversión analógico/digital (Jiménez, 2020). El sistema de iluminación es crucial para obtención de imágenes de calidad, para lo que se emplean distintas fuentes de luz como lámparas incandescentes, halógenas, fluorescentes, LED o láseres. La óptica de la cámara, que se encarga de captar los rayos de luz y proyectarlos en el sensor de imagen, está compuesta por un conjunto de lentes en un cilindro metálico llamado objetivo, mientras que mediante un diafragma se regula la cantidad de luz que llega al sensor (Sánchez-Salmerón & Ricolfe-Viala, 2016). Las cámaras digitales utilizan sensores CCD (*Charge Coupled Device*) o CMOS (*Complementary Metal Oxide Semiconductor*) para convertir la luz en señales eléctricas.

1.3.2.3. Preprocesado de imágenes

Una vez se han adquirido las imágenes, se suelen someter a **operaciones de preprocesado** con vistas a mejorar sus propiedades, lo que incidirá en la calidad del análisis (extracción, segmentación, clasificación y reconocimiento) y la toma de decisiones consiguiente. Este proceso incluye operaciones para optimizar el contraste, el brillo y la geometría de la imagen, y se dividen en dos categorías: las que afectan a la geometría y las que lo hacen. Entre los operadores que no afectan a la geometría hay que citar los puntuales, locales y globales (Herrera, et al., 2016).

Los operadores puntuales operan con los píxeles individuales y son utilizados para mejoras del contraste y el brillo de la imagen, o bien para calibrar o normalizar la imagen. Suelen realizarse mediante operaciones algebraicas y estadísticas, como las expansiones de contraste, que se basan en ecuaciones lineales o no lineales para ajustar los niveles digitales y ocupar todo el rango de resolución de la imagen (Jiménez, 2020) (Ilustración 1-15). La calibración a una escala de blanco/negro también responde a este tipo de operadores.

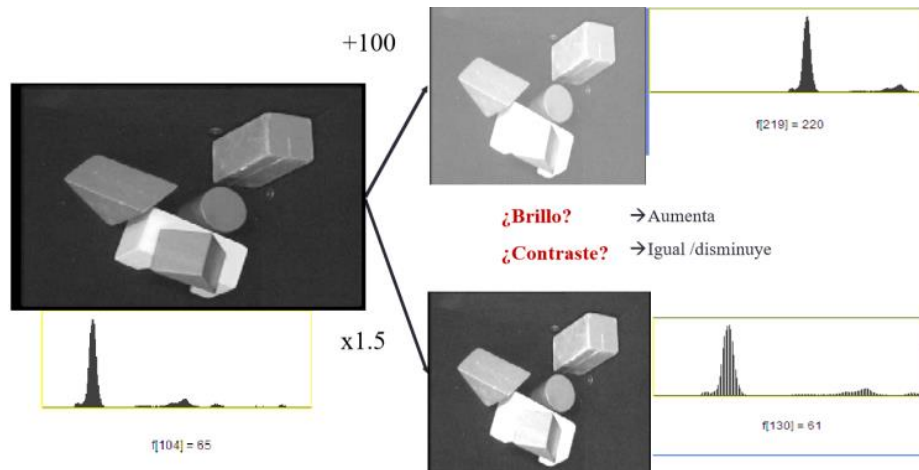


Ilustración 1-15. Aumento del brillo y expansión del contraste (tomada de Jiménez, 2020).

Los operadores locales operan con píxeles y sus vecinos (matrices de 3x3 o mayores) y básicamente corresponden a los filtros y operadores de gradiente. Los filtros son esenciales en el análisis de imágenes, utilizados para eliminar ruido o ya directamente para la extracción de objetos; de hecho, son uno de los elementos fundamentales de las redes convolucionales. Hay numerosos tipos y criterios para clasificarlos, pero de una forma sencilla se puede hablar de (Espinosa-Aranda, et al., 2016): filtros de paso bajo, que eliminan ruido aplicando medidas de tendencia central (media, distribución gaussiana) como se muestra en la Ilustración 1-16; filtros de paso alto, que aumentan las diferencias entre píxeles, detectando bordes y contornos (Sobel y laplaciano) como en la Ilustración 1-17; direccionales, que son un caso particular de los anteriores, pero detectando líneas y bordes en determinadas direcciones (Prewitt) como en la Ilustración 1-17; texturales, describen la textura de la imagen utilizando medidas de dispersión. La morfología matemática se utiliza también para el pre-procesado y segmentación, aplicando operaciones que simplifican las imágenes preservando su forma (erosión, dilatación). También existen la posibilidad de aplicar filtros en el dominio de la frecuencia, para lo cual es necesario hacer una transformación de la imagen, que se suele hacer mediante la transformada de Fourier.



Ilustración 1-16. Filtros de paso bajo (tomada de (Espinosa-Aranda, et al., 2016)).

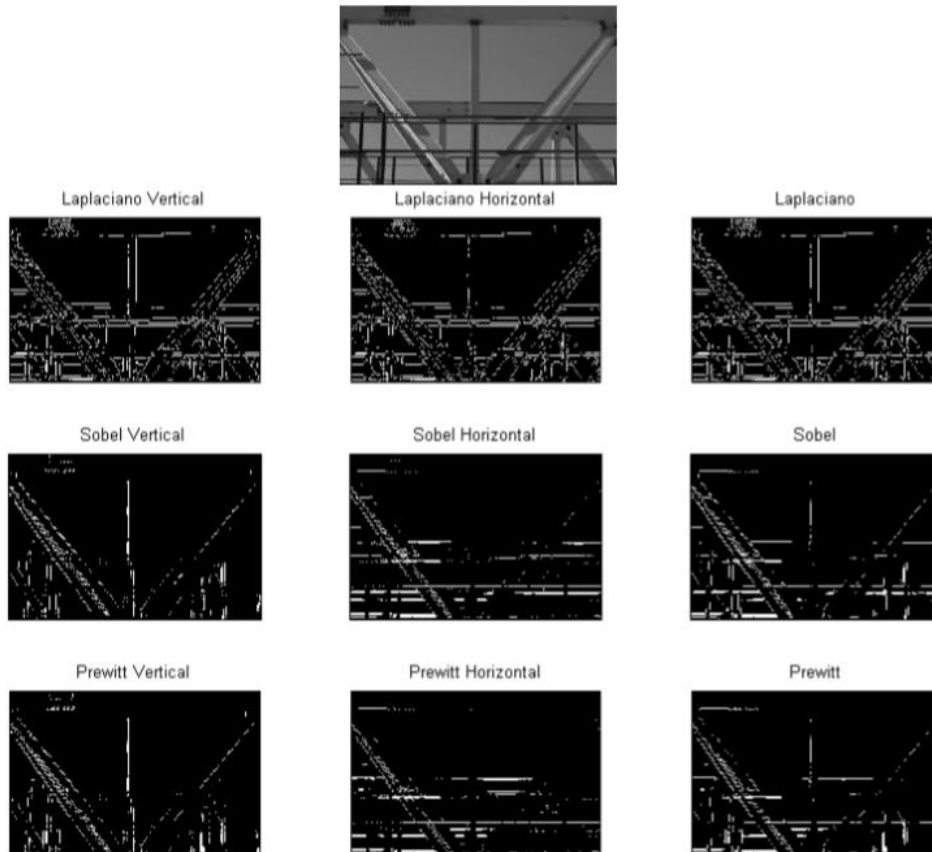


Ilustración 1-17. Filtros de extracción de bordes (tomada de (Espinosa-Aranda, et al., 2016)).

Finalmente, los operadores globales son transformaciones que afectan a toda la imagen, como cambios de dominio o del espacio de color, por ejemplo, desde RGB a otros sistemas como HSV o Lab, facilitando la visualización de elementos no claros. Las transformaciones al dominio de la frecuencia, antes comentadas en los filtros de la frecuencia, también pueden considerarse dentro de este grupo.

Por su parte, las transformaciones geométricas modifican la posición de los píxeles sin cambiar sus valores, permitiendo operaciones como la rotación, escalado y deformación (*warping*). Este tipo de operaciones va a ser importante en el reconocimiento de objetos ya que un mismo objeto puede aparecer desplazado, rotado o con diferente tamaño o forma dentro de una imagen.

1.3.2.4. Análisis de imágenes

Una vez se han preparado las imágenes, se pasa al propio proceso de análisis que suele comenzar por la **detección y extracción de elementos de bajo nivel** dentro de los objetos como los bordes y esquinas. La detección de bordes es

fundamental en el análisis de imágenes, ya que identifica los límites entre los objetos y el fondo. Para esto, se utilizan operadores de derivada que permiten localizar cambios en los niveles digitales, aplicando la primera derivada (gradiente) para encontrar grandes variaciones (picos de la intensidad); y la segunda derivada (laplaciana) para detectar pasos de negativo a positivo (Espinosa-Aranda, et al., 2016). Posteriormente, se debe establecer un umbral mínimo para distinguir los bordes del ruido.

Para utilizar operadores de gradiente, se calcula la diferencia entre píxeles vecinos, aunque esta forma simple de operar es bastante sensible al ruido. Por ello, se emplean filtros más sofisticados como los de Roberts, Sobel y Prewitt (Ilustración 1-18), que suavizan la imagen y mejoran la detección de bordes. El operador laplaciano es muy efectivo, pero también sensible al ruido, y se usa comúnmente en combinación con el filtro laplaciano de la gaussiana (LoG) en el método de Marr-Hildreth (Marr & Hildreth, 1980; Marr, 1982). Igualmente, el detector de bordes de Canny (Canny, 1986) destaca por minimizar la distancia entre bordes detectados y reales, utilizando un filtro gaussiano seguido de la obtención del gradiente y la eliminación de máximos no relevantes.

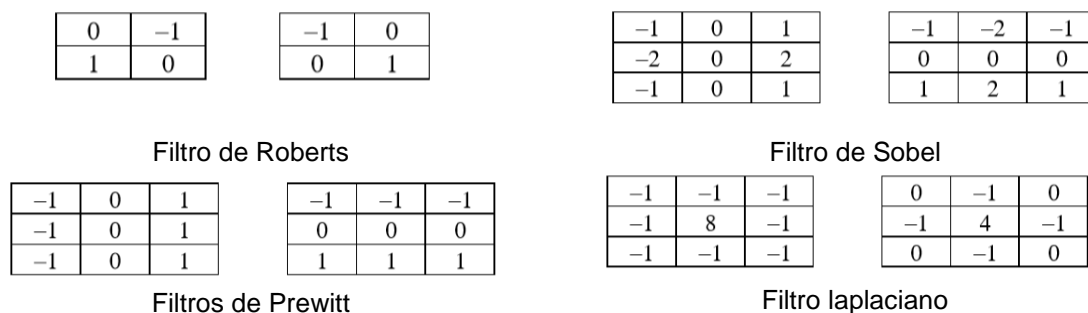


Ilustración 1-18. Filtros de Roberts, Sobel, Prewitt y laplaciano (elaboración propia).

Por su parte, las esquinas se detectan mediante gradientes y son más complejos y difíciles de detectar que los bordes, ya que pueden pasar más desapercibidos al tener menor continuidad y por el difuminado. Se emplean métodos específicos, como Kitchen-Rosenfeld o Beaudet, basados en la determinación de primeras y segundas derivadas; u otros como el operador del Harris que parte de una matriz *Hessiana* (Ilustración 1-19).



Ilustración 1-19. Detectores de esquinas y puntos característicos: a) Harris; b) SIFT (tomada de (Chen, et al., 2012)); b: SIFT (tomada de (Lowe, 2004)).

Finalmente, la extracción de puntos característicos (*featured points*) amplía la detección de bordes y esquinas (Alegre & Fernández-Robles, 2016), mediante el añadido de descriptores a los puntos detectados, que luego son extraídos. Se utilizan métodos robustos frente a variaciones en la imagen, como SIFT (*Scale Invariant Feature Transform*), basado en filtros gaussianos (Ilustración 1-19); y SURF (*Speed up Robust Features*), basado en la matriz *Hessiana*.

La **segmentación de la imagen** es el proceso de dividir una imagen en regiones homogéneas, a partir de atributos como el color o la intensidad. Existen dos enfoques principales: el primero es la división de la imagen en regiones que luego se etiquetan y que permiten la definición de fronteras; el segundo parte de la detección de bordes, completándolos y definiendo fronteras. En ambos casos, una vez que se han definido las fronteras entre objetos se pasaría a la descripción de los mismos.

Dentro de las técnicas del primer tipo, hay que destacar la umbralización, que separa la imagen en regiones según el valor de intensidad de los píxeles. Un umbral (*threshold*, T) se identifica como un nivel digital o valor de intensidad a partir del cual un grupo determinado de píxeles serán considerados como pertenecientes a un subconjunto determinado, mientras que el resto se asignan a un segundo subconjunto (Guijarro, et al., 2016). La umbralización se determina habitualmente a partir del histograma de la imagen y puede segmentar la imagen en dos clases (figura y fondo) o más clases. Hay a su vez dos grandes tipos de técnicas de umbralización que son la umbralización global, basada en el histograma de la imagen completa o la umbralización local, basada en los histogramas de sectores de la imagen.

Para el primer caso, hay métodos como las rupturas naturales (*natural breaks*) que buscan mínimos en el histograma para establecer los umbrales y segmentar la imagen. Muy importantes, son los de agrupamiento o *clustering*, como el análisis de la varianza o el k-medias (*k-means*), que agrupan píxeles iterativamente. En el algoritmo de k-medias se parte de dos (o más) grupos de píxeles, cada uno de los cuales tiene un histograma con sus estadísticos correspondientes; según esto, todos los píxeles cuyo valor esté más cerca de la media de cada objeto se clasifican dentro de esa clase (Ilustración 1-20). Otros métodos ampliamente utilizados son el método de Otsu (Otsu, 1979), que optimiza el umbral minimizando la varianza dentro de las clases y maximizando la diferencia entre ellas; el método de Kapur (Kapur, et al., 1985), basado en la entropía; o la propia umbralización local, antes comentada.

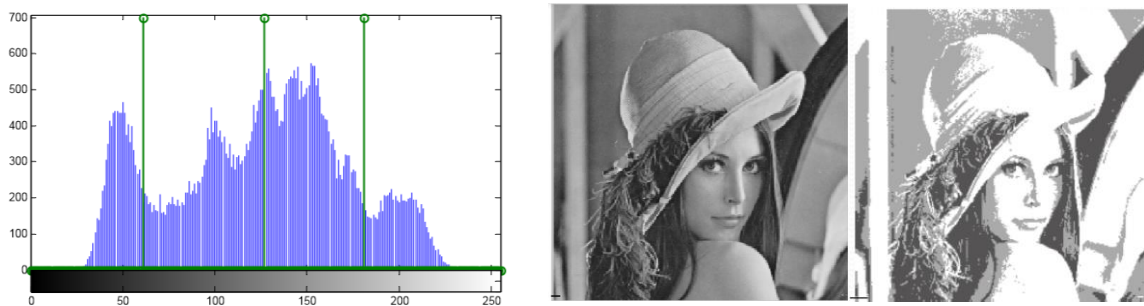


Ilustración 1-20. Umbralización mediante k-means (tomada de (Jiménez, 2020; Kanso & Ghebleh, 2017)).

La segmentación por color es útil para imágenes a color, y generalmente se basa en técnicas como k-medias y Otsu, extendidas a imágenes multispectrales (Ilustración 1-21). Actualmente, los métodos de segmentación basados en aprendizaje automático y profundo están en auge, combinando segmentación y clasificación de imágenes, como se analizará en apartados posteriores.



Ilustración 1-21. Segmentación y clasificación de imágenes en color (tomada de <https://jp.mathworks.com/discovery/image-segmentation.html>).

El segundo conjunto de métodos se enfoca en la extracción de bordes utilizando operadores como Prewitt, Sobel, Roberts, y algoritmos como Marr-Hildreth o Canny, ya comentados en el apartado anterior. Estos requieren verificar y completar los bordes detectados, mediante técnicas de seguimiento de contornos, como los algoritmos de conectividad, el etiquetado de componentes conexas y otros. Por otra parte, la transformada de Hough es otra técnica usada para detectar curvas y bordes, mientras que las serpientes o *snakes* (Kass, et al., 1988) y las tijeras inteligentes o *intelligent scissors* (Mortensen & Barrett, 1998) ayudan a delimitar objetos en tiempo real. Todos estos métodos permiten delinear los límites entre objetos que luego habría que rellenar y describir.

Además, los métodos de crecimiento de regiones agrupan píxeles similares según alguna propiedad como la intensidad o el color de los píxeles, partiendo de semillas (técnicas de abajo a arriba); mientras que los de división de regiones parten de una única región y la dividen según un criterio de homogeneidad que, si no es cumplido, produce una nueva división.

En general, las tareas de segmentación no suelen dar por sí mismas un resultado preciso en la delimitación de los objetos, por lo que, antes de extraer las características se requiere generalmente una etapa de post-procesado. Un conjunto de técnicas que dan buen resultado en este sentido son las de morfología matemática, que incluyen la erosión, dilatación, cierre, apertura, etc. Una técnica avanzada en esta línea, que produce buenos resultados en el algoritmo de *watershed* (Beucher, 1992).

La **extracción de características** en imágenes segmentadas se realiza mediante descripciones matemáticas de elementos y regiones, considerando atributos como intensidad, color, tamaño, forma y posición. Estas descripciones pueden enfocarse en el contorno (frontera del objeto) o en su interior; y deben ser únicas, completas y resistentes a transformaciones geométricas como rotación o escalado. También pueden realizarse sobre puntos característicos u otros elementos extraídos de la imagen.

Para los contornos, se utilizan métodos como los códigos de cadena, que codifican la dirección de avance en una imagen binaria; y la aproximación poligonal,

que ajusta segmentos lineales al contorno (Jiménez, 2020). En la descripción de regiones, se extraen características geométricas y topológicas de todos los píxeles del objeto, como los momentos geométricos. Dentro de estos se calculan momentos de orden 0 (área) y 1 (centroide), pero especialmente interesantes son los de orden 2 que se muestran invariantes ante transformaciones. También se consideran descriptores topológicos, como el número de componentes conexas o el número de Euler, que permiten una descripción global de las propiedades de las regiones.

En cuanto a las características no geométricas, se extraen datos sobre los niveles de intensidad de los objetos, mediante medidas como la media, moda y mediana, útiles para describir objetos en imágenes en color RGB, multispectrales o incluso hiperspectrales (en este caso se pueden obtener los espectros medios de las regiones o de toda la imagen). También se analiza la textura, que describe la suavidad o rugosidad de una superficie, usando técnicas estadísticas como la varianza o la desviación típica a partir del histograma. En otros casos se aplican métodos específicos como la entropía o avanzados como los Patrones Locales Binarios (LBP) (García-Olalla & Alegre, 2016).

Para una descripción global de la imagen, aunque también para las regiones, se emplean técnicas como el Análisis de Componentes Principales (PCA), los descriptores de Fourier y los Histogramas del Gradiente (HOG), que permiten seleccionar la información y eliminar la no relevante. En el caso de puntos característicos, obtenidos a partir de los métodos SIFT y SURF, se miden gradientes locales alrededor de cada punto o sumas de derivadas de primer orden, aunque presentan limitaciones, como el tamaño del descriptor. Para resolver esto, se pueden agrupar descriptores mediante técnicas como el *Bag of Words* (BoW).

1.3.2.5. Reconocimiento de objetos e interpretación de imágenes

El **reconocimiento de objetos**, como rostros o máquinas, y la **interpretación de imágenes** es una tarea sencilla y automática para los seres humanos, pero programar este proceso para que sea realizado por una máquina es sumamente complicado. A lo largo de los años, especialmente en los últimos, se han logrado importantes avances en este campo gracias al desarrollo de técnicas de aprendizaje automático y minería de datos. El reconocimiento de patrones, que es el núcleo de

este problema, consiste en asignar etiquetas a objetos para indicar a qué clase pertenecen. Este proceso utiliza un conjunto de medidas, llamadas atributos o características, que se representan en un vector de características.

El reconocimiento de patrones se basa en dos tipos principales de métodos: supervisado y no supervisado (García-Ordás, et al., 2016; Torres, 2020). En los métodos supervisados, los datos están etiquetados, es decir, se sabe a qué clase pertenecen los objetos. El objetivo es entrenar un clasificador que pueda etiquetar automáticamente nuevos datos, es decir, clasificar. En el caso de que las etiquetas no sean discretas sino continuas, este proceso se denomina regresión. Por otro lado, en el reconocimiento no supervisado, no se tienen etiquetas previas y el objetivo es descubrir grupos o *clusters* de datos que compartan características similares, lo que se conoce como *clustering* (Bosch, et al., 2019).

Para que el proceso de reconocimiento de patrones sea efectivo, es fundamental que el conjunto de datos con el que se entrena el clasificador sea representativo de las condiciones reales de operación. Es esencial seleccionar las características más relevantes y definir un procedimiento adecuado para obtener las medidas necesarias. Para ello hay distintos métodos de selección de características ya mencionados, tanto de filtro (análisis de correlación), como de envoltura (eliminación recursiva) y embebidos (LASSO, árboles de decisión o bosques aleatorios). En muchos otros casos, lo que se aplican son método de reducción de la dimensionalidad como el análisis de componentes principales (PCA) o las propias redes convolucionales para simplificar o agrupar las características más importantes.

Dentro de los métodos no supervisados, uno de los más utilizados es el algoritmo k-medias, que organiza los datos en grupos o *clusters* en los que los miembros de cada grupo son similares entre sí y diferentes del resto de grupos (Lloyd, 1982) (Ilustración 1-22).

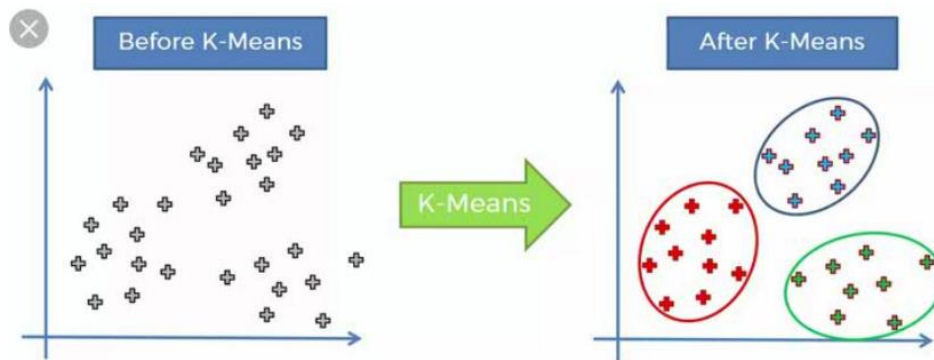


Ilustración 1-22. Algoritmo de k-means (tomada de <https://medium.com/@rohithramesh1991/unsupervised-text-clustering-using-natural-language-processing-nlp-1a8bc18b048d>).

. En el ámbito de los métodos supervisados, hay métodos muy empleados como los clasificadores estadísticos (Platero, 2008), y dentro de ellos los algoritmos de proximidad como el de mínima distancia, que asigna los objetos a la clase más cercana en términos de características o el k-vecinos más próximos (K-nearest neighbours, k-NN), que clasifica los objetos según sus vecinos más cercanos en el espacio de características. Por otra parte, los clasificadores probabilísticos, basados en la teoría bayesiana, también juegan un papel importante en el reconocimiento de patrones. El clasificador de máxima probabilidad, por ejemplo, asigna los objetos a las clases a las que tienen mayor probabilidad de pertenecer y suele proporcionar mejores resultados que enfoques más simples, como el de mínima distancia.

Otros clasificadores se basan en ajustar líneas o superficies que separan las clases de objetos en el espacio de características. Los más simples son los que utilizan regresión lineal, que ajusta una línea recta para separar las clases para lo que se emplean métodos como el análisis discriminante (James, et al., 2017). Para problemas más complejos, donde las clases no pueden separarse mediante una línea recta, se utilizan métodos más avanzados como las máquinas de vectores de soporte (SVM), que emplean hiperplanos para separar las clases en un espacio multidimensional (Vapnik, 1995; Bishop, 2006). Otra alternativa es la regresión logística que utiliza conjuntamente clases continuas o categóricas (James, et al., 2017; Bishop, 2006) para separar entre clases de una variable igualmente categórica o binaria. En este punto, se entra ya en las técnicas de aprendizaje automático, que ya se trataron en el apartado anterior, y que proporcionan resultados óptimos en problemas de clasificación y reconocimiento complejos, como el caso de las imágenes.

1.3.2.6. Aprendizaje automático en el análisis de imágenes

El aprendizaje automático o *machine learning* (Bosch, et al., 2019; Torres, 2020) ha impulsado el desarrollo de técnicas que permiten a las computadoras aprender a partir de datos, con una extensa aplicación en análisis de imágenes y reconocimiento de objetos (Bishop, 2006). Estas técnicas incluyen métodos de clasificación de tipo sintáctico, basados en algoritmos sencillos como los árboles de decisión, que permiten clasificar elementos en clases mediante reglas simples (Breiman, et al., 1986; Bishop, 2006; Hastie, et al., 2009) (Ilustración 1-23 a). Una extensión de este método es el los bosques aleatorios o *random forest* (Breiman, 2001; Belgiu & Drăgu, 2016), que utiliza muchos árboles de decisión desarrollados a partir de submuestras de los datos, asignando los objetos a las clases basadas en la mayoría de las decisiones tomadas por los árboles. Las máquinas de soporte vectorial (SVM) también pertenecen a este grupo (Ilustración 1-23 b).

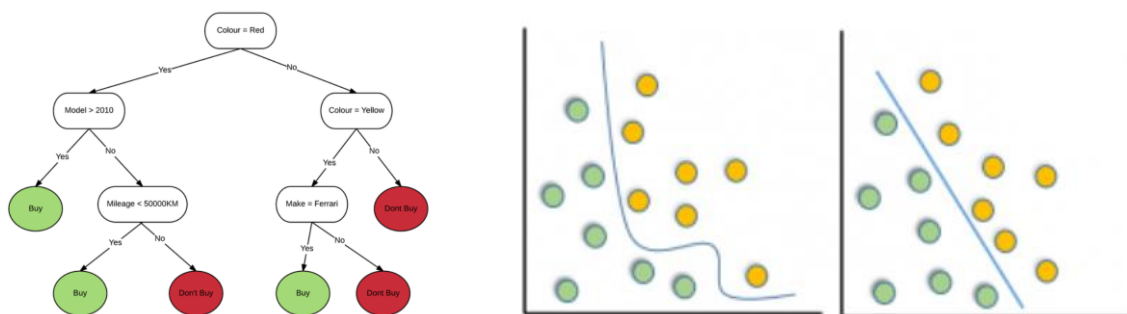


Ilustración 1-23. Aprendizaje automático: Árboles de decisión (tomada de <https://www.edrawsoft.com/es/decision-tree>); b: Support Vector Machines (SVM, tomada de <https://numerentur.org/svm/>).

Dentro del aprendizaje automático, hay un grupo que se emplea muy profusamente en el análisis de imágenes y el reconocimiento de objetos, como son las redes neuronales (Bosch, et al., 2019; Torres, 2020; Bishop, 2006). Como se ha descrito anteriormente, las redes neuronales son estructuras de procesamiento que intentan emular las funciones del cerebro humano mediante un conjunto de variables numéricas (entradas) que se transforman en otro conjunto de variables (salidas). Estas redes están organizadas en capas: una capa de entrada, una capa de salida y varias capas ocultas intermedias. A través de ella, las redes aplican funciones de propagación, activación y transferencia que modifican la respuesta final. Existen diferentes tipos de redes neuronales, entre ellas las redes convolucionales (CNN), que son ampliamente utilizadas en la clasificación de imágenes debido a su capacidad

para detectar patrones visuales complejos y la adecuación de la propia operación de convolución a la estructura 2D de las imágenes (Torres, 2020).

Un aspecto importante es el tipo de clasificación o reconocimiento que se lleva a cabo cuando se trabaja con imágenes, en concreto a si se trata una clasificación de la imagen en el sentido de detectar qué tipo de objeto u objetos hay en una imagen (Bosch, et al., 2019); o bien realizar una segmentación semántica de la imagen, asignando cada uno de los píxeles de la imagen a una determinada clase (Sirmaçek, et al., 2019).

Una parte crucial del proceso de clasificación es la validación o verificación de los resultados obtenidos. Esto se realiza utilizando una matriz de confusión, que compara los resultados del clasificador con la realidad, es decir, con los datos reales etiquetados. En una clasificación binaria, por ejemplo, se pueden contar los verdaderos positivos, falsos positivos, falsos negativos y verdaderos negativos. El número total de elementos bien clasificados se obtiene sumando los verdaderos positivos y los verdaderos negativos, lo que se expresa en un porcentaje que indica la exactitud del clasificador.

1.3.2.7. Aprendizaje profundo en el análisis de imágenes

Como se ha comentado anteriormente, las redes convolucionales son la técnica de aprendizaje profundo con una mayor aplicación en el análisis de imágenes y el reconocimiento de objetos, debido a la posibilidad de trabajar sobre dos o tres dimensiones y el reescalado, que permite reconocer objetos de distinto tamaño. Aunque hay diversas posibilidades de aplicarlas, lo más sencillo es considerar las dos dimensiones de la altura y anchura de la imagen; y como tercer eje, lo que se conoce como el número de canales o profundidad que dependerá del tipo de imagen que se estudie. Así, para imágenes RGB la dimensión del eje profundidad será 3, mientras que para imágenes en escala de grises será 1.

El funcionamiento de las capas convolucionales se puede explicar con el ejemplo de una imagen de 28x28 píxeles, en escala de grises (Ilustración 1-24).

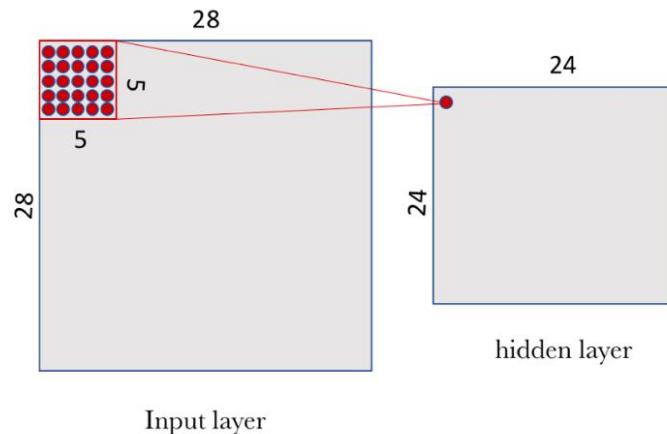


Ilustración 1-24. Funcionamiento de la capa de convolución (tomada de Torres, 2020).

El ejemplo corresponde a 784 neuronas de entrada. Cada neurona de la capa oculta será conectada a una pequeña región de 5x5 neuronas de entrada (25 neuronas de las 784). Esta ventana de 5x5 irá recorriendo toda la capa de entrada que contiene la imagen, de izquierda a derecha y de arriba hacia abajo. Por cada posición de la ventana se genera una neurona de la nueva capa de neuronas ocultas. Se observa como una capa de entrada de 28x28 con una ventana de 5x5 define un espacio de 24x24 neuronas en la primera capa oculta. En este caso, el paso de la ventana es de 1 píxel, pero este se podrá configurar según las necesidades de la red (es lo que se conoce como paso o *stride*). También, se puede rellenar de ceros alrededor de la imagen para mejora el barrido que se realiza con la ventana (*padding*), obteniendo así una capa oculta del mismo tamaño que la de entrada.

Para conectar cada neurona de la capa oculta con las 25 neuronas correspondientes de la capa de entrada se usa un sesgo b y una matriz de pesos W de tamaño 5x5, llamado filtro o *kernel*. El valor de cada punto de la capa oculta corresponderá al producto escalar entre el *kernel* y las 25 neuronas de entrada, lo que corresponde a la operación de convolución. Normalmente, no solo se aplicará un solo filtro, sino un conjunto de ellos generando así diferentes capas ocultas resultantes. Por ejemplo, si se aplican 32 filtros, se generará una salida de tamaño 24x24x32, es decir, una matriz 3D que contiene las 32 salidas de 24x24 píxeles resultado de computar los 32 filtros sobre la entrada. El conjunto de estas 32 matrices de salida se conoce como mapa de características (Ilustración 1-25).

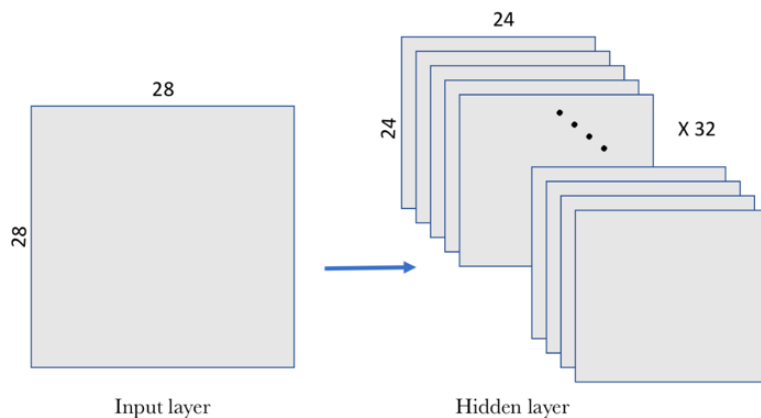


Ilustración 1-25. Mapa de características (tomada de Torres, 2020).

Cada filtro estará definido por una matriz de pesos w y un sesgo b , lo que permitirá detectar una característica concreta. Por lo que, para poder realizar el reconocimiento de imágenes habrá que, efectivamente, usar varios filtros a la vez. Así, cada capa de convolución debe capturar algún patrón de la imagen y pasarla a la siguiente capa de convolución. Se establece que los valores negativos no son importantes, siendo los valores positivos los que deberán pasar a la siguiente capa, por lo que se utilizará como función de activación una del tipo ReLU, antes definida, cuyo funcionamiento se muestra en la Ilustración 1-26.

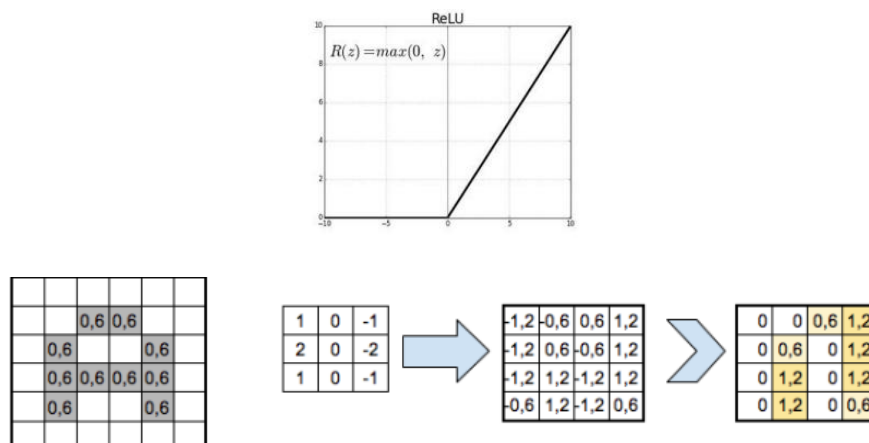


Ilustración 1-26. Función de activación (tomada de <https://pythonwife.com/intro-to-deep-learning-neural-networks-for-the-purpose-of-nlp/>).

Tras la primera convolución puede ser necesario realizar más, pero para ello será necesario reducir la cantidad de neuronas. En el ejemplo, se obtiene una primera capa oculta de 25088 neuronas ($28 \times 28 \times 32$). Si se realizase una nueva convolución de esa capa, el número de neuronas de la siguiente capa de convolución aumentaría considerablemente, provocando problemas de procesamiento. Por ello será necesario

realizar un remuestreo para reducir el tamaño de la siguiente capa de neuronas, reduciendo el tamaño de las imágenes filtradas, pero en las que prevalezcan las características más importantes. Las capas de *pooling* suelen ser aplicadas inmediatamente después de las capas convolucionales, para simplificar la información recogida en la capa convolución. El funcionamiento de estas capas de *pooling* consiste en tomar una ventana $n \times n$ de la capa convolucional y condensar la información, para lo que hay diferentes alternativas. Destaca el método *max-pooling*, en la que el valor que se toma será el valor máximo contenido en la ventana, pero también puede aplicarse el *average-pooling*, en la que se calcula el valor promedio de la ventana.

En la Ilustración 1-27, se muestra un ejemplo de *max-pooling* con una ventana de 2×2 para el procesamiento de la capa convolucional anterior (24×24). Se observa que, la capa de *pooling* se reduce quedando en una imagen de 12×12 . Con la transformación *pooling* se mantendrá la relación espacial de los objetos. El *pooling* se aplicará a cada uno de los filtros de la capa convolucional, de manera que, partiendo de una matriz $24 \times 24 \times 32$, tras aplicar el *pooling*, se obtendrá una matriz de $12 \times 12 \times 32$.

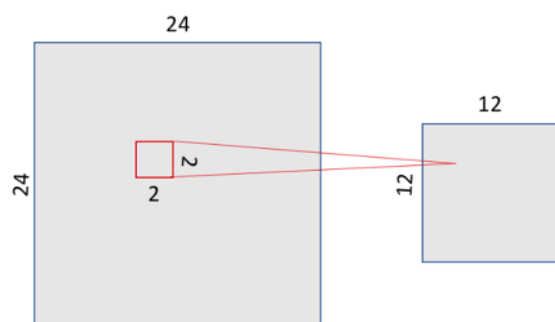


Ilustración 1-27. Funcionamiento del Max-pooling (tomada de Torres, 2020).

En la Ilustración 1-28 se muestra el proceso completo de aplicar una red CNN a una imagen. Tras la etapa de aprendizaje donde se irán sucediendo las diferentes capas de convolución, seguidas de las de *pooling*, habrá que pasar a la etapa de clasificación. Así, la última capa de pooling que será una matriz 3D $n \times n \times m$ se aplanará, pasando a ser una capa de neuronas tradicional, densamente conectada.

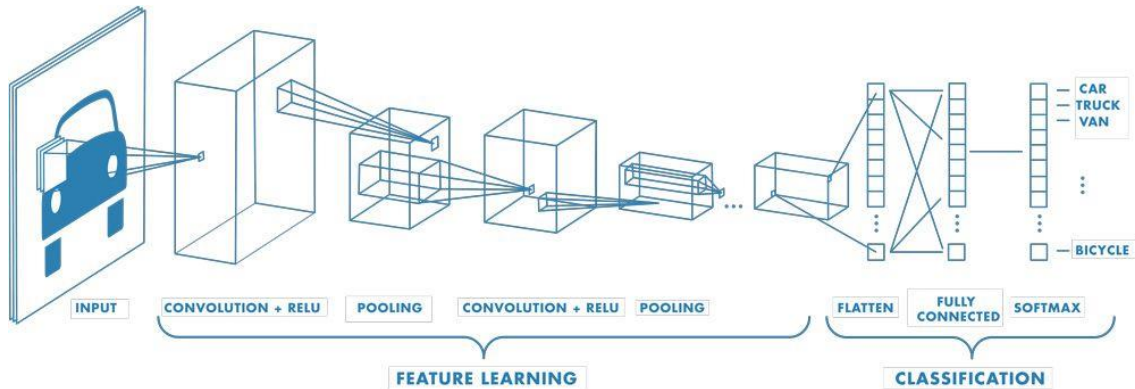


Ilustración 1-28. Aplicación de una CNN a una imagen (tomada de <https://medium.com/swlh/convolutional-neural-networks-for-multiclass-image-classification-a-beginners-guide-to-6dbc09fabbd>).

A esta nueva capa oculta “tradicional” se le aplicará una función de activación softmax (detallada anteriormente) que conecta con la capa de salida final, que tendrá una cantidad de neuronas correspondiente al número de clases que se hayan fijado.

Las redes neuronales aplicadas a imágenes pueden tener diferentes objetivos como se muestra en la Ilustración 1-29, entre los que destacan (Sirmaçek, et al., 2019; Bosch, et al., 2019):

- Clasificación de la imagen: Realizar una predicción sobre qué clase de objeto hay en la escena; por ejemplo, detectar si en una imagen como la del ejemplo hay un gato o un perro (Ilustración 1-29 a).
- Detección de objetos: Se trata de hacer predicciones en una escena acerca de dónde pueden encontrarse los objetos pertenecientes a una determinada clase. Es un problema más complejo que el de clasificación de una imagen completa; se crearán una serie de regiones a examinar y sobre estas se aplicará el mismo método que para la clasificación; posteriormente habrá que introducir una serie de técnicas para evitar que un mismo objeto se clasifique varias veces. Normalmente, se marcan los diferentes objetos con un rectángulo encuadrante (*bounding box*) donde se indica también la clase a la que pertenecen como en la Ilustración 1-29 b).
- Segmentación semántica. Se clasificará cada uno de los píxeles de la imagen, creando crearán unos contornos precisos de los objetos en la escena. No se determinará el número de objetos de una determinada clase que hay en la imagen, si no el número de píxeles de esta. En el ejemplo de

la Ilustración 1-29 c, se muestra cómo se segmentan los píxeles que pertenecen al gato, por un lado; y los que pertenecen al perro, por el otro.

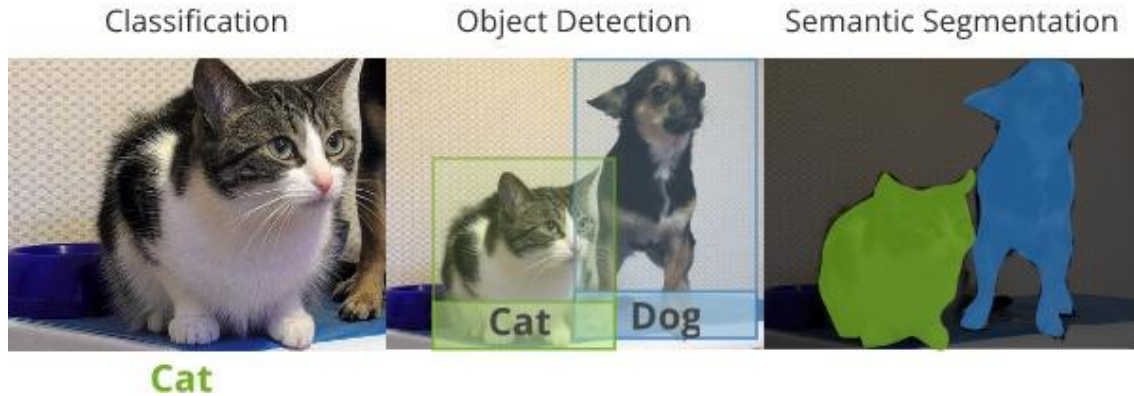


Ilustración 1-29. Resultado CNN para Clasificación b) Resultado para detección de objetos c) Resultado con segmentación semántica (tomada de (Sirmaçek, et al., 2019); <https://deepai.org/publication/sequential-image-processing-methods-for-improving-semantic-video-segmentation-algorithms>).

- Segmentación de instancias (Lin, et al., 2014). Es la combinación entre la detección de objetos y la segmentación semántica. Primero se realiza una localización de los diferentes objetos de la imagen y posteriormente se realiza una segmentación dentro del marco delimitado de este. La diferencia en el resultado de esta técnica y la segmentación semántica es que, en la segmentación de instancias, se etiquetarán cada uno de los objetos que pertenezcan a la misma clase: oveja 1, oveja 2..., como se muestra en la Ilustración 1-30.

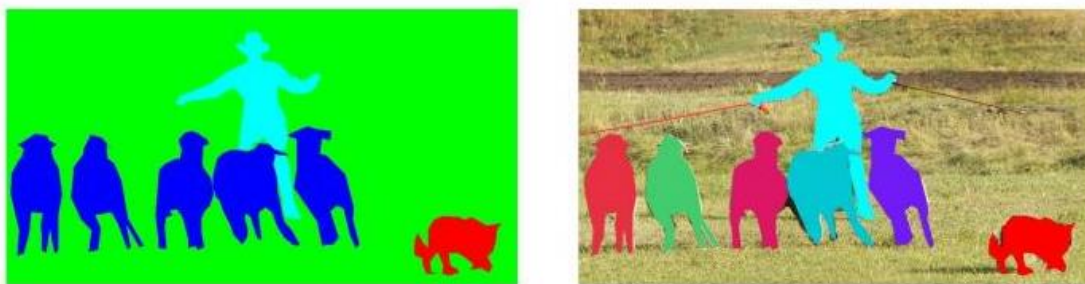


Ilustración 1-30. a) Segmentación semántica; b) Segmentación de instancias (tomada de (Lin, et al., 2014); https://link.springer.com/content/pdf/10.1007%2F978-3-319-10602-1_48.pdf).

- Otras aplicaciones como la transferencia de estilos, que permiten, por ejemplo, pintar una imagen con un estilo diferente, transferido de otra imagen.

Un concepto importante, relacionado con estos métodos avanzados de aprendizaje profundo es el aumento de datos (*data augmentation*) (Torres, 2020) (Bosch, et al., 2019). Los modelos resultantes de estos métodos a menudo tienen una gran cantidad de parámetros, lo que les permite generalizar bien cuando tienen una cantidad masiva de datos etiquetados. Sin embargo, en la práctica, estos grandes conjuntos de datos etiquetados no estarán siempre disponibles, lo que conduce a un riesgo elevado de sobreajuste, al entrenar con un número más reducido y sesgado de datos. El aumento de datos es una técnica de uso común para aumentar tanto el tamaño como la diversidad de los conjuntos de entrenamiento etiquetados realizando transformaciones en estos, pero aprovechando sus etiquetas. Actualmente es una técnica muy común para combatir el sobreajuste en CNN. Las transformaciones más comunes que se le realizan a las imágenes son:

- Volteo sobre el eje horizontal o vertical.
- Rotación medida en grados.
- Translación horizontal o vertical.
- Aplicación de un factor de escalado aleatorio.
- Modificaciones en contraste, saturación y brillo.

No obstante, incluso tras emplear la técnica de aumento de datos, el riesgo de sobreajuste continúa, puesto que, al aumentar los datos, parte de las características de las nuevas imágenes son repeticiones de las anteriores. Para solucionar este problema, se utiliza la técnica de aprendizaje por transferencia (*transfer learning*), una de las más importantes dentro del aprendizaje profundo (Torres, 2020; Bosch, et al., 2019). Su interés radica en que, en lugar de entrenar una red desde cero, lo que implicaría una gran cantidad de datos y tiempo de computación (días o semanas), se hace desde una red entrenada. El conjunto de datos con los que se entrena la red original es muy grande, de manera que el aprendizaje de la red permite al modelo preentrenado actuar como un modelo genérico del mundo virtual, y, por lo tanto, sus características pueden ser útiles para muchos problemas de visión por computador.

El proyecto Imagenet (Krizhevsky, et al., 2012) es una gran base de datos visuales diseñada para su uso en el reconocimiento de objetos. El proyecto ImageNet organiza un concurso de software anual, *ImageNet Large Scale Visual Recognition*

Challenge (ILSVRC), en el que los programas de software compiten para clasificar y detectar correctamente objetos y escenas. Las redes que participan en este concurso, pueden ser utilizadas posteriormente por usuarios, aprovechando el conocimiento adquirido al entrenar con enormes conjuntos de datos. Algunas de las redes preentrenadas más conocidas son:

- LeNet-5 (1998) es una red convolucional pionera de 7 niveles creada por (Lecun, et al., 1998). Clasifica dígitos y es usada para reconocer números escritos a mano en cheques digitalizados. La entrada es una imagen en escala de grises de 32x32 píxeles.
- AlexNet (Krizhevsky, et al., 2012) ganó la competencia ILSVRC en 2012 (Desafío de reconocimiento virtual a gran escala) reduciendo el error respecto a sus mejores competidores. La arquitectura es similar a la LeNet pero más profunda, con más filtros por capa y con capas convolucionales apiladas. Consiste en capas de convolución con ventanas de 11x11, 5x5 y 3x3, capas de *max-pooling*, capas de activación ReLU, SGD con momentum y aumento de datos.
- GoogLeNet/Inception (Szegedy, et al., 2015) fue la ganadora de ILSVRC de 2015. Creada por Google, consiguió reducir el error hasta un valor muy cercano al nivel de rendimiento de reconocimiento humano. Su diseño está inspirado en una CNN LeNet pero implementó un elemento novedoso llamado módulo de inepción, basado en convoluciones muy pequeñas para reducir drásticamente el número de parámetros. Tiene 22 capas de profundidad, pero redujo la cantidad de parámetros a 4 millones.
- VGGNet, desarrollado por (Simonyan & Zisserman, 2014), fue el subcampeón de la competencia ILSVRC en 2014. Consta de 16 capas, 13 convolucionales y las tres últimas densamente conectadas. Similar a la AlexNet pero solo con convoluciones de ventana 3x3 píxeles y una gran cantidad de filtros. Actualmente, es una de las más usadas para extracción de características de la imagen y en estructuras de segmentación semántica. Sin embargo, VGGNet consta de 138 millones de parámetros por lo que puede ser difícil de manejar por el usuario.

- ResNet (Residual Network) fue propuesta por el equipo de Microsoft Research (He, et al., 2015). Resuelve el problema de la posible degradación de la información al entrenar redes cada vez más profundas, mediante la introducción de bloques residuales. Como se muestra en la Ilustración 1-31, en estos bloques se producen conexiones directas entre algunas capas del modelo, llamadas “conexión de salto”.

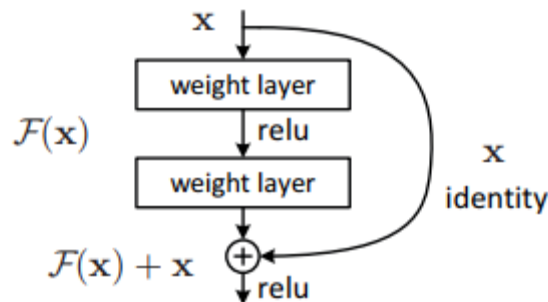


Ilustración 1-31. Bloque residual de una red ResNet (tomada de (He, et al., 2015)).

Finalmente hay que hacer referencia a un tipo especial de estructuras, los *autoencoders*, que tienen relevancia en el ámbito de la segmentación semántica cuando se combinan o integran en redes convolucionales. Los *autoencoders* son un tipo especial de redes neuronales totalmente conectadas (*fully-connected*) que funcionan intentando reproducir los datos de entrada en la salida de la red (Bosch, et al., 2019). La arquitectura de estas redes permite que puedan realizar algunas tareas muy interesantes que las diferencian de otras redes neuronales. En concreto, tres de sus principales aplicaciones son la reducción de la dimensionalidad, el preentrenamiento de redes neuronales, detectando los atributos más relevantes, y la generación de nuevos datos sintéticos que permitan aumentar el conjunto de datos de entrenamiento.

Su estructura básica presenta dos partes claramente diferenciadas:

- Codificador (*encoder*), que es el encargado de convertir las entradas a una representación interna, generalmente de menor dimensión que los datos de entrada. A veces también recibe el nombre de red de reconocimiento.

- Decodificador (*decoder*), que se encarga de transformar la representación interna a la salida de la red. También puede recibir el nombre de red generativa.

La Ilustración 1-32 muestra la estructura básica de un *autoencoder*. Aunque pueda parecer similar a la estructura de una red totalmente conectada, tiene algunas diferencias relevantes. En primer lugar, los *autoencoders* tienen el mismo número de neuronas en la capa de salida que en la capa de entrada, ya que este intentará reproducir en la salida la entrada que ha recibido. En segundo lugar, la capa oculta (o capas ocultas) deben tener un número de neuronas inferior a las capas de entrada y salida. Así, la representación interna debe preservar la información de entrada en un formato de menor dimensionalidad.

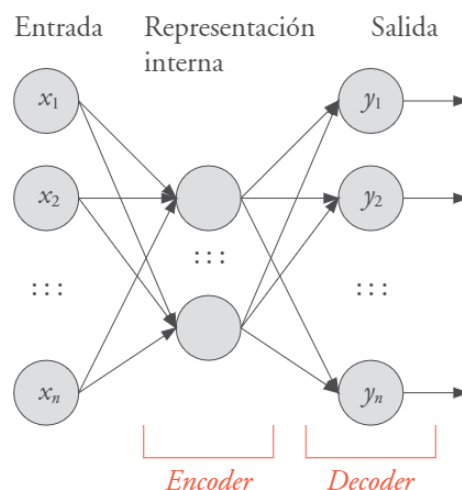


Ilustración 1-32. Estructura básica de un autoencoder (tomada de (Bosch, et al., 2019)).

La combinación de *autoencoders* y redes convolucionales permite aprovechar sus fortalezas en el procesamiento de datos, especialmente en la extracción de características, la reducción de la dimensionalidad, la compresión de imágenes y la generación de imágenes. Concretamente los denominados *autoencoders* convolucionales consisten en implementar *autoencoders* utilizando capas convolucionales, de tal forma que el *encoder* utiliza capas convolucionales para extraer características espaciales de la imagen, mientras que el *decoder* utiliza capas deconvolucionales (o transpuestas) para reconstruir la imagen (Masci, et al., 2011). Algunas de las redes convolucionales más empleadas hoy como DeepLab+, Segnet o UNet emplean estas estructuras.

1.3.2.8. Imágenes hiperespectrales

Las imágenes hiperespectrales (en inglés *hyperspectral imaging*, HSI), también conocida como imagen espectroscópica, se obtienen como cualquier imagen por la interacción de la radiación electromagnética con el material observado (Bhargava, et al., 2024). En la obtención de imágenes hiperespectrales se combina espectroscopía e imágenes para crear una base de datos tridimensional (3D), conocida como hipercubo, que integra información espectral y espacial (Selci, 2019). Este proceso recopila datos espectrales de cada píxel en un detector bidimensional (2D). A diferencia de las imágenes RGB, que solo tienen tres bandas espectrales, el hipercubo captura una gama casi continua de longitudes de onda con alta resolución espectral (Bhargava, et al., 2024) (Ilustración 1-33). Esto lo hace útil para detectar variaciones espectrales bajo diferentes condiciones. La HSI mide cómo la radiación es reflejada, transmitida o emitida por un objeto, en multitud de longitudes de onda proporcionando una visión detallada que supera a las imágenes convencionales en precisión y utilidad analítica.

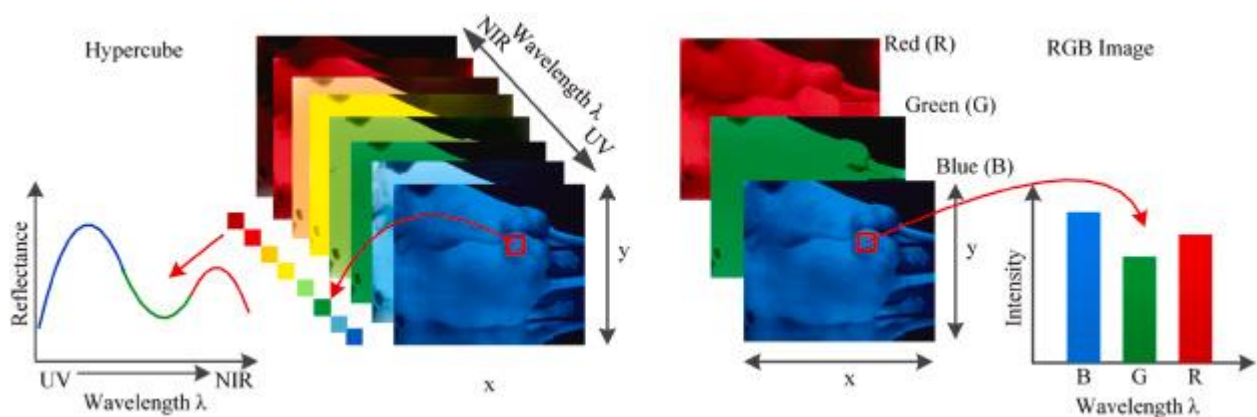


Ilustración 1-33. Imagen hiperespectral (hipercubo e imagen RGB), tomada de (Bhargava, et al., 2024).

Así, los sensores hiperespectrales, también conocidos como espectrómetros de imágenes, suelen capturar espectros de visible (RGB), infrarrojo cercano (NIR) e infrarrojo de longitud de onda corta o medio (SWIR o MIR) en el rango de la región de 0,4 a 2,5 μm , produciendo espectros espectrales con cientos de longitudes de onda distintas (Goetz, et al., 1985). Hay distintos modos de adquisición, que se resumen en: el escaneo de puntos (*whiskbroom*), el escaneo de líneas (*push broom*), la observación fija de imágenes (*imagers staring*) y las instantáneas de imágenes (*spanshots*) (Bhargava, et al., 2024).

Como cualquier tipo de imagen, las imágenes hiperespectrales se describen por su resolución espectral y espacial (Bhargava, et al., 2024). La resolución espacial determina el detalle más pequeño detectable en una imagen, definido por el tamaño de la entidad más pequeña que puede distinguirse como independiente. La resolución espectral, que es la distintiva de una imagen hiperespectral, mide el rango del espectro electromagnético y las bandas capturadas por un sensor. Un sensor con muchas bandas estrechas tiene alta resolución espectral, ideal para diferenciar firmas espectrales similares. Las imágenes multispectrales, con baja resolución espectral, no detectan detalles finos; por el contrario, los sensores hiperespectrales capturan múltiples bandas estrechas en segmentos del espectro visible e infrarrojo, ofreciendo un gran potencial para identificar materiales según sus firmas espectrales únicas, permitiendo obtener más información de la superficie (Jensen & Lulla, 1987). En las imágenes hiperespectrales, la resolución temporal se refiere al tiempo que un sensor tarda en volver a capturar datos de la misma ubicación, conocido como tiempo de retorno.

Uno de los tipos de imágenes hiperespectrales más conocidas son la que se obtienen en teledetección, tanto espacial como aerotransportada. En este caso, la superficie terrestre interactúa con la radiación solar reflejándola, absorbiéndola o transmitiéndola según su material (Jia, et al., 2021). Los sensores hiperespectrales miden esta energía reflejada en multitud de bandas espectrales, permitiendo identificar materiales como bosques, agua y suelo mediante curvas de respuesta espectral, que representan sus firmas características.

Las ventajas de utilizar las imágenes hiperespectrales en diferentes aplicaciones, como la agricultura y la estimación de la calidad de los alimentos, son las siguientes (Lu & Fei, 2014; Bhargava, et al., 2024):

- Es una tecnología no invasiva, sin contacto y no destructiva que garantiza la calidad y la seguridad de los alimentos.
- Los experimentos no utilizan productos químicos, por lo que son seguros para el medio ambiente.

- El tiempo necesario para procesar la evaluación de la calidad y el control/almacenamiento de los alimentos es bajo en comparación con los enfoques químicos y tradicionales.
- Proporcionan una mejor comprensión de los elementos químicos de los productos alimenticios.
- Proporcionan una selección de área adecuada para el análisis crítico de la imagen.
- Obtienen conjuntamente información espacial y espectral, para proporcionar datos más apropiados y precisos sobre las muestras.

A pesar de sus ventajas, la obtención de imágenes hiperespectrales también tiene algunas desventajas (Lu & Fei, 2014; Bhargava, et al., 2024):

- Son sistemas muy costosos en comparación con otras técnicas de captura de imágenes.
- Necesidad de computadoras de alta velocidad para el procesamiento de datos y unidades de gran capacidad para el almacenamiento de datos.
- La señal puede verse afectada por el entorno ambiental, como la dispersión, la iluminación, etc., lo que produce una mala relación señal-ruido.
- La detección e identificación de diferentes elementos dentro de la imagen equivalente utilizando datos espectrales, es en su mayoría difícil.

En cuanto a las aplicaciones, hay un primer grupo relacionado con la teledetección hiperespectral (Jia, et al., 2021), que permite detectar y analizar con precisión atributos de objetos terrestres. Gracias a las características espectrales únicas de cada material, las imágenes hiperespectrales permiten una clasificación más precisa del terreno y los objetos situados sobre él. Así, se ha empleado en:

- La agricultura de precisión y la monitorización de cultivos (Strachan, et al., 2002; Teke, et al., 2013). Este tipo de aplicaciones se desarrollan extensamente en el apartado siguiente (Wang, et al., 2021; Guerri, et al., 2023; Barbedo, 2023; Shuai, et al., 2024).
- Especialmente importante es la detección del estrés de las plantas, ya que la productividad de los cultivos se reduce significativamente con éste. El

estrés en las plantas puede ser el resultado de las enfermedades y la escasez del agua y nutrientes (Lowe, et al., 2017; Wan, et al., 2022).

- Los recursos terrestres geológicos de tipo geológico-minero (van der Meer, et al., 2012; Sabine, et al., 2014), la cobertura vegetal natural como en bosques y prados (Zagolski, et al., 1996) (Jänicke, et al., 2020); y todo lo relacionado con los recursos hídricos. Así, la gestión eficiente, el análisis y el monitoreo de la calidad del agua requieren mucha consideración por parte de los investigadores (Li, et al., 2018), junto a la detección y monitorización de inundaciones (Brakenridge, et al., 2007).
- Los aspectos relativos a la contaminación y protección ambiental, como la detección de gas (Kuflik & Rotman, 2012) o derrames de petróleo (Foudan, et al., 2001), entre otros.

En otro tipo de escalas se ha aplicado con gran profusión a aspectos como:

- Seguridad y calidad de los alimentos: La exigencia de una producción de bajo coste y una gran eficiencia impone diversos retos en la seguridad y calidad de los alimentos y sus productos. Para evaluarlas se examinan distintos atributos de los alimentos, como los biológicos, químicos y físicos, tanto en productos ganaderos (Tan, 2004), como agrícolas (Wang, et al., 2021; Guerri, et al., 2023; Barbedo, 2023; Shuai, et al., 2024). Estos últimos serán tratados en el apartado siguiente.
- Análisis médicos: Durante las últimas décadas, la mejora de las técnicas hiperespectrales ha permitido obtener resultados significativos por parte de los investigadores médicos. Las técnicas más utilizadas son las relacionadas con las imágenes como las resonancias magnéticas y la tomografía (Paty, et al., 1988). Además, las características ópticas del tejido permiten obtener resultados de diagnóstico valiosos (Kumar, et al., 2013), así como la realización de pruebas cutáneas posibilitan examinar la salud de la piel humana (Du, et al., 2007). También se utiliza en cirugías guiadas por imágenes (Bhargava, et al., 2024).

Por su parte, la clasificación de imágenes hiperespectrales (Lv & Wang, 2020), como uno de los aspectos claves en su utilización en todos estos campos, ha

evolucionado significativamente, dividiéndose en tres enfoques principales: características espectrales, espaciales y espectral-espaciales (Jia, et al., 2021; Shuai, et al., 2024). Inicialmente, los métodos se centraban en la determinación de características espectrales puras (*endmembers*) y la utilización de clasificadores de aprendizaje automático como análisis discriminante, regresión logística SVM o redes neuronales, pero estos enfrentaban problemas debido a la redundancia de datos y la no linealidad entre espectros y objetos terrestres (Jia, et al., 2021). Esto llevó al desarrollo de técnicas de reducción de dimensiones, como análisis de componentes principales o selección de características, aunque con resultados limitados. No obstante, la variabilidad espectral causada por iluminación, ambiente y condiciones atmosféricas incrementa la clasificación errónea, destacando la necesidad de incorporar información espacial, que aporta detalles sobre la forma y contexto de los objetos, mejorando la precisión y robustez (Jia, et al., 2021).

Sin embargo, en la última década, el aprendizaje profundo ha ganado relevancia gracias a su capacidad para aprender patrones automáticamente, superando las limitaciones de los enfoques tradicionales de aprendizaje automático (Paoletti, et al., 2019; Jia, et al., 2021). Su éxito en áreas como procesamiento del lenguaje, reconocimiento de voz y conducción autónoma lo ha llevado a ser aplicado en la clasificación de imágenes hiperespectrales. Los métodos basados en características espectral-espaciales se dividen en dos enfoques: arquitecturas de dos flujos, que separan la extracción espectral y espacial, y de flujo único, que las combinan (Jia, et al., 2021). Estas técnicas utilizan diversos módulos de aprendizaje profundo, como capas convolucionales y recurrentes. Sin embargo, entrenar modelos profundos requiere grandes conjuntos de datos etiquetados, cuya obtención es costosa y limitada en la práctica. Para abordar este desafío, se han desarrollado paradigmas que permiten entrenar modelos con pocas muestras etiquetadas. Aunque prometedores, estos métodos aún enfrentan importantes retos.

1.3.3. Aprendizaje automático y profundo para el análisis de imágenes en la agricultura

1.3.3.1. Escala de paisaje

A escala de paisaje se emplean tanto imágenes satelitales como imágenes aéreas capturadas con plataformas convencionales (avionetas o helicópteros) como con sistemas/vehículos aéreos remotamente pilotados (VARP o RPAS por sus siglas inglesas) o no tripulados (VANT o UAV/UAS). Así, tanto la teledetección espacial como la fotogrametría aérea han sido empleadas en multitud de aplicaciones como una fuente primaria de información ya que permiten una visión sinóptica y muy completa de la superficie terrestre (Jensen, 2009; Lillesand, et al., 2015; Wang, et al., 2022a). En los últimos años, el desarrollo de los VARP/VANT ha permitido la extensión de estas aplicaciones a la escala de detalle (Terwilliger & Robbins, 2020).

En cuanto a los sensores, se emplean desde sensores con 3-4 bandas espectrales (RGB-NIR) a sensores multiespectrales, hiperespectrales, térmicos, radar (SAR) y LiDAR. Entre ellas, las imágenes hiperespectrales destacan por su capacidad para adquirir información espectral de una forma más completa y continua, respecto a los RGB y multiespectrales (Khan, et al., 2018; Lv & Wang, 2020) lo que las hace muy proclives a ser analizadas mediante técnicas de aprendizaje automático y profundo (Paoletti, et al., 2019; Jia, et al., 2021; Moharram & Sundaram, 2023; Shuai, et al., 2024).

Las aplicaciones del análisis de imágenes en el sector oleícola, a la escala de paisaje, están lógicamente relacionadas con la agricultura, tanto a nivel de cartografía (estática) como de monitorización (análisis multitemporal), en lo que se conoce como agricultura de precisión (Zarco-Tejada, et al., 2014). Dependiendo de la plataforma, satélites o plataformas aéreas, las aplicaciones y las técnicas empleadas son distintas (Wang, et al., 2021; Ang & Seng, 2021 ; Guerri, et al., 2023; Shuai, et al., 2024).

De esta manera, las imágenes de satélite son el punto de partida para la cartografía, clasificación y monitorización de cultivos, así como para la predicción del rendimiento y la gestión de tierras agrícolas, que son aspectos cruciales para la planificación y asignación de recursos (Wang, et al., 2022a). Los programas espaciales más conocidos como Landsat y Sentinel ofrecen resoluciones espaciales

de decenas de metros (10-30 m) y múltiples bandas espectrales que permiten realizar análisis multiespectral y multitemporal (10-13 bandas). Entre las técnicas más comunes está el uso de índices de vegetación, como el NDVI (Rouse, et al., 1974; Bannari, et al., 1995), que utiliza las bandas de infrarrojo cercano y rojo para estimar la biomasa y el estado de los cultivos. También se emplea el índice de humedad MDNWI, que incluye bandas de infrarrojo medio (Xu, 2006). Estos índices son útiles para la agricultura de precisión, ayudando a tomar decisiones sobre riego y cosecha.

Además de los índices, se aplican técnicas como el análisis de componentes principales (ACP o PCA) y el uso de clasificadores, desde los clásicos hasta los basados en aprendizaje automático y profundo para evaluar la cobertura del suelo. Los resultados se utilizan en políticas agrarias, control de erosión, seguimiento de plagas y enfermedades, entre otros (Zarco-Tejada, et al., 2014; Weiss, et al., 2020; Zhang, et al., 2020a; Lowe, et al., 2017; Wan, et al., 2022). Particularmente interesante es el uso de imágenes hiperespectrales en distintas aplicaciones relacionadas con los cultivos, como el rendimiento, el vigor o la presencia de enfermedades (Kuswidiyanto, 2022; Barbedo, 2023; Guerri, et al., 2023; Shuai, et al., 2024). Entre los sensores hiperespectrales más empleados hay que citar Hyperion, con 30 m de resolución espacial y 220 bandas espectrales (Li, et al., 2017; Paoletti, et al., 2019).

Por su parte, las imágenes capturadas por satélites de alta resolución (Pleides, Worldview, etc.) o por sensores aerotransportados ofrecen una resolución espacial métrica o submétrica, aunque su resolución espectral y temporal es más limitada en comparación con satélites de resolución más baja. A pesar de esto, estas imágenes permiten realizar análisis de precisión a nivel de sectores dentro de las parcelas agrícolas (Zhang & Wang, 2019). Las técnicas aplicadas incluyen análisis multiespectral, textural y de reconocimiento de objetos, como el conteo automático de árboles y arbustos, lo que mejora la precisión en el seguimiento de cultivos y en las políticas agrarias. Debido a la menor resolución temporal, se suelen combinar con imágenes satelitales convencionales para un monitoreo más eficiente. En este caso, hay también sistemas aerotransportados que capturan imágenes hiperespectrales, a través de sensores como AVIRIS o ROSIS, lo que permite abordar análisis más complejos para la clasificación y el seguimiento de cultivos (Mou, et al., 2017; Zhong, et al., 2017; Li, et al., 2017; Zhu, et al., 2018; Xu, et al., 2018; He, et al., 2018; Zhong,

et al., 2018; Haut, et al., 2019; Cao & Guo, 2020; Ladi, et al., 2022). En algún caso, se ha trabajado con imágenes de muy alta resolución, similares a las que se obtienen con sensores a bordo de VARP (Wambugu, et al., 2021).

Precisamente, el uso de VARP mejora aún más la resolución espacial, alcanzando la escala centimétrica. Estos sistemas, que permiten la instalación de diferentes cámaras, incluidas las multiespectrales, hiperespectrales y térmicas, ofrecen unas mayores prestaciones en la resolución espectral, aunque a veces en detrimento de la espacial. A pesar de sus limitaciones en cuanto al área de cobertura, los VARP proporcionan una alta resolución temporal, siendo ideales para el seguimiento preciso de cultivos, la estimación de biomasa y el conteo de plantas a nivel de parcela o sector. Están ya en la transición hacia la escala de árbol o planta, en las que se pueden utilizar estos dispositivos, planificando vuelos a poca altura (Neupane & Baysal-Gurel, 2021; Shi, et al., 2022; Nguyen, et al., 2023). En este caso, la instalación de sensores hiperspectrales en estas plataformas permite la cartografía detallada de cultivos, además de otras aplicaciones relacionadas con la caracterización de las plantas, el rendimiento y la detección de enfermedades (Yang, et al., 2021; Shi, et al., 2022; Karila, et al., 2022 ; Nguyen, et al., 2023).

En este punto hay que indicar que las técnicas tradicionales de análisis de imágenes satelitales y aéreas enfrentan limitaciones al mapear cultivos complejos y áreas con variaciones topográficas o mezcladas con zonas urbanas. De esta manera, la combinación de los sensores hiperespectrales y las técnicas de aprendizaje profundo se han revelado como claves para mejorar la identificación de cultivos y obtener información detallada sobre la distribución espacial y la estructura de plantación (Wan, et al., 2021; Ladi, et al., 2022; Moharram & Sundaram, 2023). Aun así, el uso de aprendizaje profundo en el mapeo de cultivos se ve obstaculizado por la dificultad de obtener conjuntos de datos adecuados de satélites, lo que complica la coherencia y precisión de los resultados (Shuai, et al., 2024).

Además, existen desafíos técnicos, como la extracción de características efectivas de datos hiperespectrales y la selección de bandas espectrales complementarias, que dificultan una clasificación e identificación más precisa de los cultivos. Aunque hay muchas investigaciones en áreas como el estrés de plagas y el

seguimiento de la salud de los cultivos, faltan estudios detallados sobre otras tareas de detección agrícola a esta escala (Shuai, et al., 2024). Las limitaciones en los entornos estructurados, los sensores y la generalización de los modelos, junto con la falta de conjuntos de datos robustos, dificultan el avance en este campo a gran escala (Aggarwal, et al., 2023). Para superar estas barreras, es crucial integrar datos multidimensionales de múltiples fuentes y a varias escalas, lo que facilitará una toma de decisiones más precisa y en tiempo real en la gestión agrícola.

1.3.3.2. Escala de árbol o plantas

La adquisición de imágenes a esta escala se lleva a cabo generalmente mediante fotografía terrestre fija o incluso de objeto cercano, aunque también se puede efectuar mediante sistemas móviles, tanto aéreos (VARP/VARS) como terrestres, los conocidos como sistemas de mapeo terrestre móvil (*mobile mapping system*, MMS). Los más frecuentes son los primeros, con multitud de aplicaciones, sobre todo en la monitorización de los cultivos a nivel de plantas, o incluso frutos (Mahlein, 2016; Lowe, et al., 2017; Wan, et al., 2022), y en la detección de plagas y enfermedades, aunque los vuelos VANT de detalle también se aplican (Shi, et al., 2022; Nguyen, et al., 2023) y las imágenes obtenidas con MMS empiezan igualmente a ser empleados (Elhashash, et al., 2022).

La principal aplicación del análisis de imágenes en esta escala es la monitorización del crecimiento de los cultivos, y más concretamente, de las plantas y los frutos que producen, lo que es clave para regular su desarrollo, detectar deficiencias nutricionales y predecir rendimientos (García-Berná, et al., 2020; Wang, et al., 2022b). Así, las técnicas de la visión por computador y el análisis de imágenes han posibilitado el desarrollo de múltiples aplicaciones para la identificación, localización, conteo, selección y en su caso a su recolección manual o automatizada mediante robots. Esta tecnología se ha implementado para frutas como naranjas y cítricos en general, manzanas, peras, melocotones, plátanos, fresas, papayas, nueces, patatas, tomates, aceitunas, uvas, etc.; y también para granos, como los guisantes, trigo, maíz o arroz (Brosnan & Sun, 2002; Narendra & Hareesha, 2010; Beltrán Ortega, et al., 2016; Kaur, et al., 2017). Las principales aplicaciones se centran

en la identificación, localización, conteo y separación de frutos según su madurez, facilitando tanto la recolección manual como la automatizada mediante robots.

Las técnicas empleadas varían desde las clásicas como la detección de contornos hasta la segmentación y clasificación de imágenes, aplicando métodos como k-medias, k-NN, mínima distancia y aprendizaje automático. Entre las primeras, (Ramos, et al., 2017) emplearon técnicas clásicas para clasificar frutos de café en cosechables o no, mediante segmentación de imágenes usando el detector de bordes Canny y análisis *watershed*, así como la clasificación de los frutos con clasificadores como Bayes y SVM. De manera similar (Li, et al., 2016) utilizaron un enfoque basado en correlación cruzada normalizada rápida (FNCC) y transformada de Hough para detectar y contar frutos cítricos verdes. Por su parte, (Lin, et al., 2019) desarrollaron una nueva transformada probabilística de Hough para detectar y agregar subfragmentos de interés en imágenes de cítricos y aplicaron un clasificador SVM basándose en características de color y textura. Igualmente, (Kurtulmus, et al., 2011) utilizaron el análisis de textura Gabor y un enfoque inspirado en la detección de rostros para identificar cítricos verdes. Por último, (Malik, et al., 2018) detectaron tomates maduros para un robot de recolección usando un algoritmo basado en el espacio HSV y segmentación de cuencas hidrográficas.

Sin embargo, los métodos tradicionales, que dependen de la extracción manual de características o de reglas generales, no son viables a gran escala (Lu, et al., 2020), por lo que generalmente se requiere la aplicación de técnicas avanzadas. Así, (Gatica, et al., 2013) aplicaron redes neuronales artificiales (ANN) para detectar aceitunas en los árboles, utilizando parámetros morfológicos y de color. (Asensio Jiménez, et al., 2022) detectaron copas de árboles de olivo mediante técnicas de aprendizaje automático (árboles de decisión) sobre imágenes multiespectrales adquiridas desde VARP. (Cano Marchal, et al., 2021b) e (Illana Rico, et al., 2022) hicieron una estimación de nutrientes de los olivos igualmente a partir de imágenes multiespectrales adquiridas desde VARP (Afonso, et al., 2020) detectaron tomates en invernaderos utilizando la red neuronal convolucional MaskRCNN, que identifica objetos y los píxeles correspondientes, obteniendo buenos resultados. Por su parte, (Aquino, et al., 2020) entrenaron una red neuronal convolucional (ResNet) con más

de 300.000 sub-imágenes de olivos, logrando una alta precisión en la clasificación de las aceitunas.

No obstante, es la combinación de imágenes hiperespectrales y algoritmos de aprendizaje profundo lo que permite analizar grandes volúmenes de datos espectrales que facilitan la extracción de características y la clasificación a nivel de plantas y frutos. Así, entre las técnicas aplicadas se encuentran las redes neuronales profundas (DNN) para la predicción del rendimiento de cultivos a partir de índices de vegetación (Fei, et al., 2022), para estimar el rendimiento y la biomasa en la soja (Yoosefzadeh-Najafabadi, et al., 2021), para el fenotipado del trigo (Moghimi, et al., 2020) o la estimación del contenido en clorofila en las hojas del trigo (Zhang, et al., 2021). También se aplican redes convolucionales (CNN) 1D a los espectros medios para estimar la concentración de nitrógeno del dosel y la biomasa aérea en césped, cebada y pastos (Pullanagari, et al., 2021; Patel, et al., 2023), el rendimiento del maíz (Yang, et al., 2021); o redes recurrentes (RNN) para predecir la humedad en canola y trigo (Torres-Tello & Ko, 2021) o detectar el contenido de clorofila en maíz y relacionarlo con su crecimiento (Zhao, et al., 2023). Finalmente hay trabajos que consideran la utilización de redes convolucionales 2D o *autoencoders*, que permiten la segmentación semántica de la imagen; estas se han empleado para la predicción del contenido de clorofila en el dosel en perales (Paul, et al., 2020), la humedad en canola y trigo (Torres-Tello & Ko, 2021), el rendimiento del maíz (Yang, et al., 2021) o el fenotipado del maíz (Nguyen, et al., 2023).

Sin embargo, migrar los algoritmos de aprendizaje profundo para monitorear el crecimiento de los cultivos presenta retos. Es probable que los algoritmos no se apliquen de manera directa a diferentes tipos de cultivos o entornos ambientales, lo que requiere un modelado adaptativo y una extracción efectiva de características a nivel de dosel. Además, el monitoreo de los cultivos a escala del dosel implica varias tareas, como la detección del contenido biológico y fisicoquímico, la evaluación del colapso de los cultivos, la predicción del rendimiento y el fenotipado de alto rendimiento. Coordinar todas estas tareas es esencial para mejorar la eficiencia del monitoreo agrícola y la toma de decisiones (Shuai, et al., 2024).

Un segundo conjunto importante de aplicaciones es la detección de estrés por plagas y enfermedades en los cultivos, que igualmente resulta esencial para mejorar el rendimiento y la calidad, mediante la aplicación de técnicas rápidas, no destructivas y eficientes, como la determinación de índices (NDVI y otros) o la clasificación de imágenes (Peñuelas & Filella, 1998; Mahlein, 2016; Lowe, et al., 2017; Thomas, et al., 2018; Neupane & Baysal-Gurel, 2021). Sin embargo, en los últimos años, el uso de imágenes hiperespectrales ha revolucionado este proceso de monitorización de las enfermedades en plantas o árboles, mediante el empleo de índices de vegetación específicos y técnicas de aprendizaje automático (Lowe, et al., 2017; Thomas, et al., 2018; Golhani, et al., 2018; Neupane & Baysal-Gurel, 2021; Wan, et al., 2022; Wang, et al., 2022b; Shuai, et al., 2024).

No obstante, la detección de estrés en plantas por plagas y enfermedades enfrenta aún varios desafíos (Shuai, et al., 2024), relacionadas con la elección de la plataforma desde donde adquirir las imágenes o con las técnicas de detección. Así, las plataformas terrestres portátiles y robóticas son efectivas para la detección temprana y detallada, aunque tienen una cobertura limitada; mientras que los VANT, que ofrecen una mayor cobertura, están influenciados por las condiciones climáticas y limitaciones legales de vuelo, además de las limitaciones que aún muestran para el empleo de cámaras hiperespectrales (Shuai, et al., 2024). Por otra parte, los factores cambiantes en exteriores pueden distorsionar las respuestas espectrales de los cultivos, lo que demanda algoritmos adaptativos. Además, la respuesta de las plantas a diferentes tipos de estrés puede parecer similar, lo que requiere datos más completos y métodos analíticos más precisos, que generalmente conducen a la utilización de técnicas de aprendizaje profundo (Golhani, et al., 2018; Neupane & Baysal-Gurel, 2021; Wan, et al., 2022; Wang, et al., 2022b; Shuai, et al., 2024)

Así, en algunos trabajos se emplean técnicas de aprendizaje automático o profundo, tanto redes profundas (DNN) como recurrentes (RNN) para detectar hongos como el tizón tardío en el arroz (Chen, et al., 2022; Bai, et al., 2022) o insectos en la soja (Osco, et al., 2022). En otros casos, se aplican redes convolucionales 1D a los espectros medios para predecir el contenido de agua en plantas de maíz (Rehman, et al., 2020) o detectar igualmente el mildiu de la patata (Shi, et al., 2022). Sin embargo, la mayor parte de los trabajos aplican redes convolucionales 2D para identificar y

localizar blanquita en repollos (Hu, et al., 2022); gorgoja en árboles de pino (Gao, et al., 2023); mildiu de la patata (Qi, et al., 2023) o virus Y (Polder, et al., 2019); tizón del arroz (Feng, et al., 2022), roya del trigo (Zhang, et al., 2019) y fusarium del trigo (Jin, et al., 2018; Zhang, et al., 2020b).

1.3.3.3. Escala de órganos

En general se puede decir que el control de calidad en los productos agrícolas es clave para satisfacer las expectativas de los consumidores y seleccionar adecuadamente los lotes para el procesamiento. Factores como las influencias físicas, biológicas y químicas pueden afectar a la calidad de los productos, influyendo en características como la madurez de los frutos, el vigor de las semillas, la apariencia y el precio (Shuai, et al., 2024).

El control de calidad y la selección de frutas una vez recolectadas, ya sea en laboratorio o directamente en las cadenas de producción, como en las almazaras, ha sido objeto de numerosos estudios en los que se utilizan diversas técnicas de segmentación y clasificación de imágenes. Estas técnicas han sido aplicadas a diferentes frutas, incluyendo plátanos, tomates, cítricos y aceitunas, con el objetivo de evaluar su madurez, defectos y calidad.

En trabajos pioneros se han empleado algoritmos de segmentación y clasificación (k-means, etc.), que han mostrado resultados variables según el tipo de fruta y los atributos analizados (Ojeda-Magaña, et al., 2010). Igualmente, desde hace tiempo se vienen utilizando las redes neuronales (Yang, 1996) o clasificadores k-NN y de distancia mínima (Patil & Yadahalli, 2011), para identificación y clasificación de granos alimentarios.

No obstante, es la introducción de las técnicas de aprendizaje automático y profundo junto a las imágenes multiespectrales e hiperespectrales las que han producido un avance más significativo de las aplicaciones en este ámbito. Así, se pueden encontrar distintas revisiones sobre el estado de la cuestión como las de (Wang, et al., 2022b; Guerri, et al., 2023; Shuai, et al., 2024) y de forma más específica la de (Wieme, et al., 2022). En este se citan trabajos específicos sobre el control de calidad de frutas y verduras, encontrando aplicaciones sobre la detección de

pigmentos (clorofila, caroteno, etc.), firmeza y elasticidad, contenido en humedad y solutos, acidez, desórdenes fisiológicos, daños físicos, infecciones y enfermedades, etc. Las técnicas hiperespectrales y de aprendizaje automático y profundo se han usado con éxito para clasificar la calidad, detectar la viabilidad de semillas, identificar daños y defectos, y clasificar variedades (Wang, et al., 2022b; Barbedo, 2023).

De forma concreta, estas técnicas se han aplicado en el reconocimiento y clasificación de granos de maíz a partir del espectro del infrarrojo cercano (Zhang, et al., 2020c; Wei, et al., 2023), así como los granos de trigo germinados que influyen en buena medida sobre la calidad de la harina (Zhu, et al., 2023). Igualmente se ha empleado en la clasificación de variedades de arroz (Onmankhong, et al., 2022) o el vigor de sus semillas (Wu, et al., 2022). Se han descrito, otras aplicaciones sobre la inspección de los granos de café (Chen, et al., 2022), la viabilidad de las semillas de pimienta (Hong, et al., 2023) o de sófora (Pang, et al., 2022) o la determinación del origen de las bayas de goji (Dong, et al., 2022),

No obstante, persisten aún desafíos importantes en la generalización de los modelos de calidad debido a las variaciones espectrales entre diferentes variedades, lotes y fuentes. Estas variaciones complican el desarrollo de modelos universales y, en muchos casos, se requieren modelos específicos para cada situación. Entrenar numerosos modelos de aprendizaje profundo aumenta la carga computacional, aunque los enfoques de aprendizaje por transferencia pueden ayudar a reducir esta carga (Shuai, et al., 2024). Otro obstáculo importante es cumplir con los requisitos de baja latencia en aplicaciones del mundo real, que exigen mayor eficiencia y precisión en la clasificación. La investigación futura debería centrarse en el desarrollo de plataformas hiperespectrales más eficientes, un preprocesamiento de datos más rápido, inspecciones de clasificación con baja latencia y modelos más livianos.

Por su parte, la monitorización del estado nutricional de los cultivos, las plantas y sus órganos es también crucial para mejorar su rendimiento. Tradicionalmente, los análisis de laboratorio para evaluar parámetros nutricionales han sido costosos y consumen mucho tiempo (Wang, et al., 2022a; Barbedo, 2023; Wieme, et al., 2022) Como alternativa, la combinación de imágenes hiperespectrales y aprendizaje profundo ha demostrado ser una solución prometedora, ya que permite monitorear

diferentes parámetros de los órganos de los cultivos, como el contenido de agua, nitrógeno, clorofila, materia seca y sólidos solubles (Shuai, et al., 2024). Así, se han desarrollado algunas aplicaciones para estimar la madurez de las fresas (Gao, et al., 2020) o de los granos de maíz cubiertos (Zhang, et al., 2020c).

A pesar de los avances, aún quedan pendientes aspectos en la monitorización del estado nutricional, especialmente en condiciones experimentales. Los cambios espectrales pueden deberse a muchos factores por lo que no es sencillo determinar su origen y relación. Además, la falta de conjuntos de datos diversos, de gran escala y bien etiquetados representa un obstáculo importante para construir modelos robustos (Shuai, et al., 2024).

Finalmente, y más relacionado con los objetivos de este trabajo, la identificación y cuantificación precisa del estrés en los cultivos es fundamental para una gestión agrícola eficaz y la selección de variedades adecuadas. De esta manera, las técnicas hiperespectrales y el aprendizaje profundo se han utilizado ampliamente para estudiar el estrés biótico y abiótico en diferentes órganos de los cultivos, tanto de organismos patógenos (Agarwal, et al., 2020) como de residuos de pesticidas (Han & Gao, 2019). Concretamente, se han desarrollado aplicaciones para identificar el estrés y enfermedades en granos alimenticios (Gao, et al., 2021) y particularmente en el arroz (Feng, et al., 2022; Zhang, et al., 2022) o el trigo (Zhang, et al., 2020b; Chu, et al., 2022). De igual forma se han utilizado para detectar virus asociados a la vid y las uvas (Sawyer, et al., 2023) o las manzanas (Cui, et al., 2022). De especial interés en este caso se ha de citar un trabajo que permite la detección de antracnosis en los olivos (Fazari, et al., 2018).

Sin embargo, todavía existen retos en la adquisición y procesamiento de datos, lo que afecta la precisión y la detección temprana de las tensiones en los cultivos. Esto se debe a que los síntomas de estrés suelen ser sutiles, incluso cuando se emplean imágenes hiperespectrales avanzadas. Además, el proceso de etiquetar datos para entrenar modelos de aprendizaje profundo es complejo y laborioso, lo que genera dificultades en la creación de modelos precisos (Shuai, et al., 2024). La mayoría de las investigaciones actuales se centran en el estudio de un solo tipo de estrés bajo condiciones controladas, lo que permite observar los cambios en el nivel de estrés a

lo largo del tiempo. Sin embargo, las interacciones entre diferentes variedades de cultivos, factores estresantes y condiciones ambientales provocan variaciones en las respuestas espectrales. Esto resalta la necesidad de desarrollar técnicas más avanzadas y precisas para la identificación y cuantificación del estrés en escenarios más complejos y realistas.

1.3.3.4. Otras aplicaciones en el sector oleícola

El análisis de imágenes y la visión por computador han sido profusamente aplicadas en diversas etapas del proceso de producción de aceite de oliva, así como en la inspección de aceitunas, tanto en el árbol como una vez recolectadas.

Así, desde hace años, se han desarrollado distintas aplicaciones. (Díaz, et al., 2004) desarrollaron un sistema de visión por computadora para clasificar aceitunas de mesa empleado distintos algoritmos. Por su parte, (Furferi & Carfagni, 2010) utilizaron la visión artificial y el algoritmo k-medias para evaluar defectos en la superficie de las aceitunas y determinar el índice de maduración. (Guzmán, et al., 2013a) emplearon un sistema de visión infrarroja para detectar defectos en las aceitunas y clasificar su daño en cuatro categorías, mientras que (Cano-Marchal, et al., 2021a) propusieron una metodología para evaluar defectos en las aceitunas en las almazaras mediante segmentación global y adaptativa.

Para la clasificación de lotes de aceitunas, (Aguilera Puerto, et al., 2015) y (Aguilera Puerto, et al., 2019) (Martínez Gila, et al., 2015a) desarrollaron métodos automáticos para distinguir entre aceitunas de árbol (sanas) y aceitunas de suelo (defectuosas). Utilizaron vectores de características que combinaban histogramas en diferentes espacios de color y valores de textura, y aplicaron clasificadores como análisis discriminante y redes neuronales, así como máquinas de soporte vectorial (SVM), logrando excelentes resultados con una precisión de hasta el 97%. Por su parte, (Fernández Ruiz, 2022) empleó diversos métodos de visión por computador y análisis de imágenes, especialmente de aprendizaje automático y profundo (redes convolucionales 2D preentrenadas) para identificar las aceitunas y diferenciarlas de los residuos (hojas) y suciedad de la cinta transportadora.

En cuanto la estimación del estado de maduración de las aceitunas, (Furferi & Carfagni, 2010) aplicaron redes neuronales para ello; mientras que (Guzmán, et al., 2013b) (Guzmán, et al., 2015) desarrollaron un método de segmentación y clasificación basado en imágenes NIR y el algoritmo k-NN para predecir el índice de maduración de las aceitunas. (Martínez Gila, et al., 2022) utilizaron distintos clasificadores como el de *Bayes Naive*, un perceptrón multicapa y máquinas de vectores de soporte para determinar la firmeza de las aceitunas. Otro aspecto crucial, más centrado en la producción, es la estimación del contenido de aceite, para lo que (Ram, et al., 2010) desarrollaron modelos de predicción basados en regresión lineal y redes neuronales, siendo estas últimas más precisas. Por su parte, (Carfagni, et al., 2008) combinaron el análisis de imágenes con redes neuronales para estimar parámetros como la acidez y el número de peróxidos del aceite de oliva, logrando también detectar defectos en las aceitunas. (Navarro Soto, et al., 2018) utilizaron técnicas de regresión de mínimos cuadrados parciales (PLS) para estudiar distintos parámetros en las aceitunas como la acidez, contenido en clorofila, etc.

También se han desarrollado técnicas para la clasificación de variedades de aceitunas. Así, (Vanloot, et al., 2014) desarrollaron un sistema de visión artificial para identificar cinco variedades principales de aceitunas francesas mediante el análisis de imágenes de huesos de aceituna, utilizando el análisis de componentes principales (PCA) para reducir la complejidad de los datos y el análisis discriminante como clasificador. (Agriopoulou, et al., 2021) emplearon un enfoque similar para diferenciar entre seis variedades de aceitunas basándose en la forma y dimensiones de los huesos.

Un ejemplo destacado es el trabajo de (Martínez Gila, et al., 2013) y (Martínez Gila, et al., 2018a), quienes se enfocaron en el control del proceso de batido y la pasta del aceite. Utilizando imágenes multiespectrales (en el espectro visible e infrarrojo), lograron ajustar parámetros del proceso, como la temperatura, el tiempo de batido, etc. Mediante un análisis de textura en el espectro visible y una segmentación infrarroja, detectaron variables clave: aceite flotante, viscosidad y granularidad de la pasta.

Adicionalmente, la visión por computador ha sido útil en la evaluación de la calidad del aceite tras su extracción, especialmente en la clasificación del aceite según el contenido de impurezas y humedad. (Cano Marchal, et al., 2013) propusieron una metodología que clasifica las muestras de aceite en tres categorías según el contenido de impurezas, utilizando imágenes en color en los espacios Lab y HSV y aplicando técnicas como el análisis de componentes principales, análisis discriminante, máquinas de vectores de soporte (SVM) y redes neuronales artificiales (ANN). En cuanto a la calidad química del aceite, (Martínez Gila, et al., 2015b) desarrollaron un método para estimar parámetros clave como la acidez, el índice de peróxido y la humedad, utilizando cámaras hiperespectrales. Tres algoritmos, entre ellos el algoritmo genético (GA) y el de proyección sucesiva (SPA), identificaron las bandas espectrales óptimas para el análisis. Finalmente (Martínez Gila, et al., 2018b) desarrollaron un modelo de regresión para identificar igualmente polifenoles en las muestras de aceite de oliva.

2. MATERIALES Y MÉTODOS

2.1.1. Hardware y software

Para la realización del presente trabajo se ha contado un ordenador adecuado para la ejecución de modelos complejos de aprendizaje automático y análisis de datos. Este equipo permitió llevar a cabo el procesamiento eficiente de grandes volúmenes de datos y la implementación de técnicas de aprendizaje profundo requeridas en el proyecto. A continuación, se detallan las especificaciones del equipo empleado:

- Procesador (CPU): 11th Gen Intel® Core™ i7-11375H a 3.30 GHz.
- Tarjeta gráfica (GPU): NVIDIA GeForce RTX 3060 Laptop GPU.
- Memoria RAM: 32 GB.

En cuanto al software, se han empleado los siguientes programas:

- Matlab
- Excel
- Word

Entrando en detalle, en Matlab además de las funciones y aplicaciones implementadas para este trabajo se han utilizado algunas de las cajas de herramientas (*Toolbox*) que el software contiene, como son la aplicación *Image Labeler*, *Classification Learner* y el entorno *App Designer*.

La aplicación *Image Labeler* de Matlab permite crear clases y etiquetar los píxeles de una imagen o un conjunto de ellas. Una vez etiquetadas, los datos se guardan en un archivo tipo *gTruth*, el cual se puede importar al entorno de trabajo, conteniendo los datos de cada píxel etiquetado. En este trabajo se utilizó para obtener el conjunto de datos con la que se entrenó la red convolucional 2D que realiza la segmentación semántica de la aceituna, separándola del fondo y eliminando los defectos de estas.

La aplicación *Classification Learner* permite importar datos etiquetados, dividirlos en los conjuntos de entrenamiento, validación y test y utilizarlos para entrenar modelos de aprendizaje automático, así como redes sencillas.

Adicionalmente, tiene funciones para realizar selección de características como los métodos de análisis de componentes principales (PCA) y ANOVA. En este trabajo se ha utilizado el *Classification Learner* para realizar el entrenamiento de la clasificación de las aceitunas mediante árboles de decisión, algoritmo SVM, análisis discriminante, k-NN, y métodos ensamblados y redes neuronales sencillas. Además, se ha realizado un estudio de los métodos de selección de características que ofrece.

El entorno *App Designer* es una herramienta para crear aplicaciones con interfaces gráficas (GUI). Permite diseñar visualmente la interfaz arrastrando y colocando componentes, y programar su funcionamiento mediante un editor de código integrado en MATLAB. Es útil para desarrollar aplicaciones interactivas que integren análisis, visualización y procesamiento de datos de forma sencilla. En este proyecto se utiliza este entorno para la creación de una interfaz de visualización de las imágenes y espectros medios por píxeles, otra para la definición de áreas de interés (ROI) y una última para la visualización de los resultados tras la clasificación.

2.1.2. Preparación de las muestras y adquisición de imágenes

Las imágenes han sido proporcionadas por los tutores del trabajo. Han consistido en dos conjuntos de imágenes, de aceitunas y de hojas, tratadas con distintos productos fitosanitarios en distintos niveles de concentración.

Con carácter previo a la adquisición de las imágenes, se ha realizado la aplicación de los productos fitosanitarios a las aceitunas y hojas. Las aceitunas se recolectaron mediante vibración en un olivar ecológico de la variedad Picual situado en la provincia de Jaén, para garantizar que las aceitunas estuvieran inicialmente libres de productos fitosanitarios. A continuación, se aplicaron distintos pesticidas con concentraciones que se encuentran comúnmente en las prácticas agrícolas diluidos con 500 ml de agua destilada como se indica en Tabla 2-1, creando un total de 25 diluciones (5 para cada producto) (Martínez-Gila, et al., 2024).

Para preparar las muestras de aceitunas contaminadas, los 20 kg de aceitunas recolectadas se dividieron en 26 subgrupos, sirviendo uno de ellos como muestra de control, mientras que cada uno de los 25 subgrupos restantes se trató con una de las 25 diluciones. Para cada muestra contaminada, las aceitunas se extendieron

uniformemente sobre una base de plástico y se aplicó la dilución correspondiente utilizando una boquilla rociadora para asegurar una cobertura uniforme. De cada muestra tratada, se seleccionaron aleatoriamente 10 aceitunas para ser inspeccionadas con la cámara hiperespectral según se indica más adelante; mientras que las aceitunas restantes se enviaron a un laboratorio externo acreditado de análisis de aceite, donde se aplicó un método estandarizado de detección destructiva de múltiples residuos basado en cromatografía de gases acoplada a espectrometría de masas (GC-MS/MS) para certificar el nivel de contaminación de cada muestra (Martínez-Gila, et al., 2024). Los resultados obtenidos en los análisis se muestran en la Tabla 2-2.

En el caso de las hojas, se ha seguido un procedimiento similar, obteniendo en este caso 12 muestras de hojas para cada concentración para el haz y otras tantas para el envés, con objeto de ser inspeccionadas por la cámara hiperespectral.

Producto	Dilución 1 Dilución madre 1%(p/v)	Dilución 2 Dilución 0.1% (p/v)	Dilución 3 Dilución intermedia 0.01% (p/v)	Dilución 4 Dilución 0.0001% (p/v)	Dilución 5 Dilución 0.00001% (p/v)
Diflufenican 50%	10 g/ 500 ml	50 g 1%/ 500 ml	5 g 1%/ 500 ml	5 g 0.01%/ 500 ml	0.5 g 0.01%/ 500 ml
Oxifluorfen 24%	20.83 g/ 500 ml	50 g 1%/ 500 ml	5 g 1%/ 500 ml	5 g 0.01%/ 500 ml	0.5 g 0.01%/ 500 ml
Piretrina (delta- metrina) 2.5%	200 g/ 500 ml	50 g 1%/ 500 ml	5 g 1%/ 500 ml	5 g 0.01%/ 500 ml	0.5 g 0.01%/ 500 ml
Piretrina (lambda- cihalotrina) 1.5%	333.33 g/ 500 ml	50 g 1%/ 500 ml	5 g 1%/ 500 ml	5 g 0.01%/ 500 ml	0.5 g 0.01%/ 500 ml
Tebuconazol 25%	20 g/ 500 ml	50 g 1%/ 500 ml	5 g 1%/ 500 ml	5 g 0.01%/ 500 ml	0.5 g 0.01%/ 500 ml

Tabla 2-1. Concentraciones de los productos fitosanitarios aplicados a las aceitunas en este trabajo.

Aceitunas		
Producto	Nivel	Concentración
Sin tratar	0	
Deltametrina	1	37,2 (LC=0,010 mg/Kg; GC MSMS QqQ) *LMR UE (mg/Kg): 0.6
	2	19,1 (LC=0,010 mg/Kg; GC MSMS QqQ) *LMR UE (mg/Kg): 0.6
	3	0,77 (LC=0,010 mg/Kg; GC MSMS QqQ) *LMR UE (mg/Kg): 0.6
	5	Todos los parámetros analizados han sido menores que el LC
Diflufenican	1	31,8 (LC=0,010 mg/Kg; GC MSMS QqQ) *LMR UE (mg/Kg): 0,6
	2	3,42 (LC=0,010 mg/Kg; GC MSMS QqQ) *LMR UE (mg/Kg): 0,6
	3	0,46 (LC=0,010 mg/Kg; GC MSMS QqQ) *LMR UE (mg/Kg): 0,6
	4	0,015 (LC=0,010 mg/Kg; GC MSMS QqQ) *LMR UE (mg/Kg): 0,6
	5	Todos los parámetros analizados han sido menores que el LC
Lambda-Cihalotrina	1	84,9 (LC=0,010 mg/Kg; GC MSMS QqQ) *LMR UE (mg/Kg): 0.5
	2	11,4 (LC=0,010 mg/Kg; GC MSMS QqQ) *LMR UE (mg/Kg): 0.5
	3	0,95 (LC=0,010 mg/Kg; GC MSMS QqQ) *LMR UE (mg/Kg): 0.5
	4	0,014 (LC=0,010 mg/Kg; GC MSMS QqQ) *LMR UE (mg/Kg): 0.5
	5	Todos los parámetros analizados han sido menores que el LC
Oxiflufen	1	72,3 (LC=0,010 mg/Kg; GC MSMS QqQ) *LMR UE (mg/Kg): 1
	2	11,5 (LC=0,010 mg/Kg; GC MSMS QqQ) *LMR UE (mg/Kg): 1
	3	1,12 (LC=0,010 mg/Kg; GC MSMS QqQ) *LMR UE (mg/Kg): 1
	5	Todos los parámetros analizados han sido menores que el LC
Tebuconazol	1	116 (LC=0,010 mg/Kg; GC MSMS QqQ) *LMR UE (mg/Kg): 0.5
	2	19,2 (LC=0,010 mg/Kg; GC MSMS QqQ) *LMR UE (mg/Kg): 0.5
	3	1,29 (LC=0,010 mg/Kg; GC MSMS QqQ) *LMR UE (mg/Kg): 0.5
	4	Todos los parámetros analizados han sido menores que el LC
	5	Todos los parámetros analizados han sido menores que el LC

Tabla 2-2. Concentraciones de los productos fitosanitarios en las aceitunas resultantes de los análisis.

Una vez, seleccionadas las aceitunas se procedió a la adquisición de las imágenes. Para ello, se ha utilizado una cámara hiperespectral PIKA L de la marca Resonon, que captura la radiación electromagnética en 300 bandas entre el ultravioleta, visible e infrarrojo cercano (381,76 nm – 1019,51 nm), lo que proporciona una distancia espectral de aproximadamente 2,54 nm. La cámara cuenta con una resolución espacial de 900 píxeles por línea y una resolución espectral de 300 canales por línea (Martínez-Gila, et al., 2024). La cámara se muestra en la Ilustración 2-1 y sus características en la Tabla 2-3.

El sistema de iluminación empleado consistió en dos focos halógenos de 100 W, dispuestos para proporcionar una iluminación uniforme durante la adquisición.



Ilustración 2-1. Cámara multispectral empleada.

A	Características
Rango espectral	400-1000nm
Canales espaciales	900
Resolución espectral (FWHM)	2.7 nm
Muestreo espectral por píxel	5.86 μ m
Tasa máxima de fotogramas	249 fps
Profundidad de bits	12
Dimensiones	115x104x66 mm
Peso	0.64 kg
Objetivo	C-mount
Tipo de sensor	CMOS
Interfaz de datos y consumo	USB 3.0
Consumo de batería	3.4 W (vía USB)
Temperatura de operación	0 a 50°C

Tabla 2-3. Características de la cámara multispectral.

En la Tabla 2-4 se muestran los conjuntos de imágenes empleadas en este trabajo, concretamente el número de imágenes que se han procesado y analizado, tanto para las aceitunas como para las hojas. Las imágenes están en formato BIL (*bands interleaved by lines* o bandas intercaladas por líneas).

Así, en el caso de las aceitunas, aunque inicialmente se contaba con 240 imágenes (no se dispone de las imágenes correspondientes a los niveles de concentración 4 de deltametrina y oxifluorfen), se han descartado algunas por presentar propiedades muy diferentes en el visible, probablemente por su estado de maduración, quedando finalmente un total de 234 imágenes. Se distribuyen en aproximadamente 10 imágenes para cada uno de los 5 niveles, de cada uno de los 5 tipos de fitosanitarios aplicados (salvo en deltametrina y oxifluorfen, donde solo se dispone de 4 niveles), más 10 imágenes de aceitunas sin tratar.

Por su parte, en el caso de las hojas se dispone de 46 imágenes para los haces de la hoja y 46 para el envés, que corresponden aproximadamente a 2 imágenes de haz y 2 de envés para cada uno de los 5 niveles de cada uno de los 5 tipos de fitosanitarios (excepto para tebuconazol, donde se solo se dispone de 3 niveles), más las 2 + 2 imágenes de las hojas sin tratar. Puesto que cada imagen contiene 6 hojas, en total se han capturado imágenes del haz y envés de 276 hojas.

Producto	Nivel	Aceitunas Nº imágenes	Hojas Nº imágenes
Sin tratar	0	10	2 / 2
Deltametrina	1	10	2 / 2
	2	10	2 / 2
	3	10	2 / 2
	4	0	2 / 2
	5	10	2 / 2
		40	10 / 10
Diflufenican	1	10	2 / 2
	2	8	2 / 2
	3	10	2 / 2
	4	10	0
	5	10	2 / 2
		48	8 / 8
Lambda-Cihalotrina	1	10	2 / 2
	2	10	2 / 2
	3	10	2 / 2
	4	10	2 / 2
	5	10	2 / 2
		50	10 / 10
Oxiflurfen	1	10	2 / 2
	2	9	2 / 2
	3	10	2 / 2
	4	0	2 / 2
	5	10	2 / 2
		39	10 / 10
Tebuconazol	1	9	2 / 2
	2	9	2 / 2
	3	10	0
	4	10	0
	5	9	2 / 2
		47	6 / 6
Total		234	46 / 46
			276 / 276

Tabla 2-4. Imágenes empleadas en el trabajo.

2.2. Metodología

La metodología seguida en este trabajo, en el que se parte de las imágenes de las aceitunas y hojas, proporcionadas por los tutores del trabajo, se esquematiza en el siguiente diagrama de flujo (Ilustración 2-2).

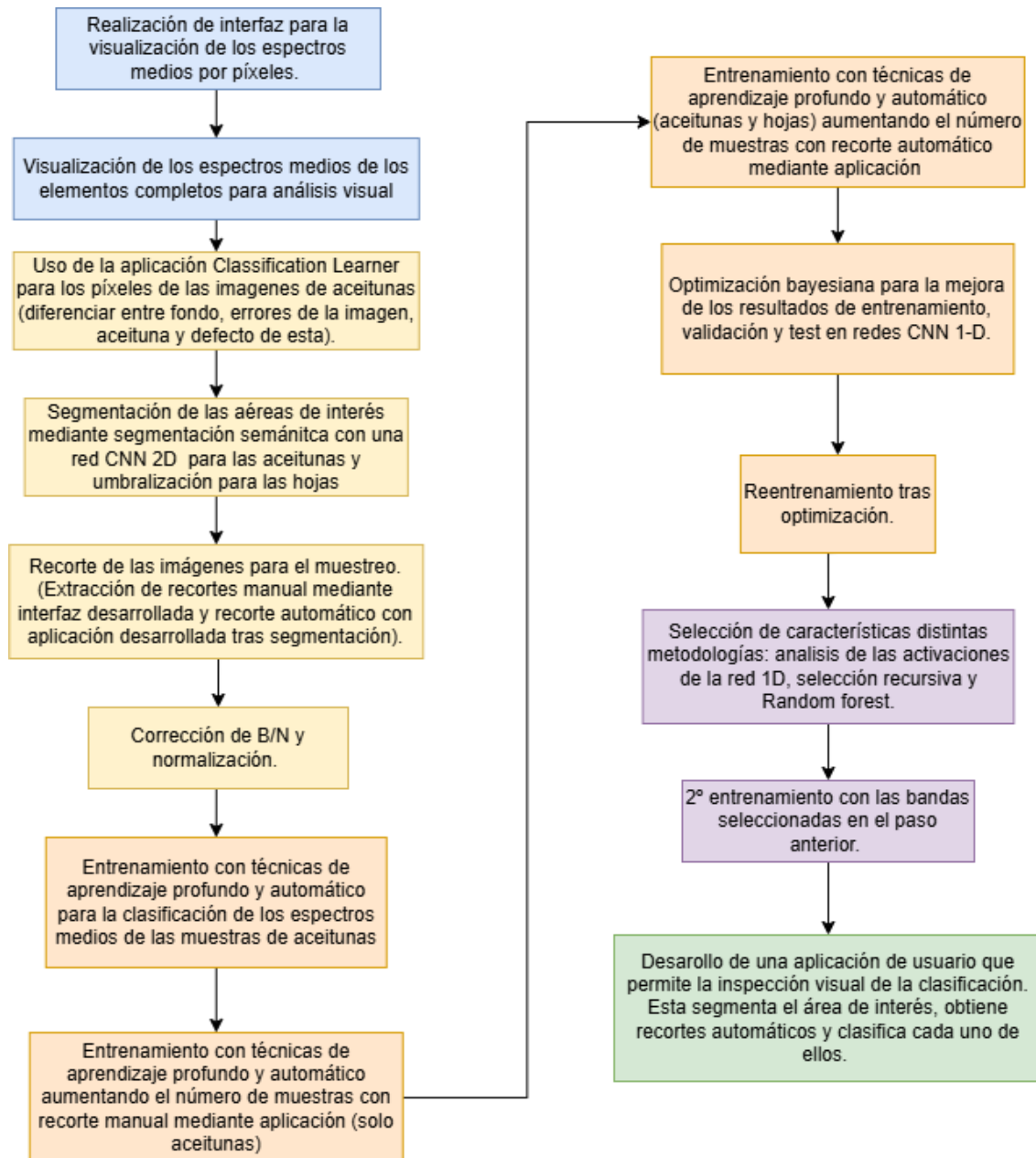


Ilustración 2-2. Diagrama general de flujo seguido en este trabajo.

2.2.1. Interfaz de visualización y consulta

Se ha creado una interfaz para la visualización de imágenes en Matlab, en la que se seleccionan las bandas con las que se pueden observar composiciones en color real (rojo, verde y azul en sus correspondientes posiciones RGB) y en falso color (cualquier otra combinación, incluyendo las bandas de infrarrojo y borde del rojo). Como bandas del rojo, verde y azul se consideran las bandas 128 (643,59 nm), 82 (547,21 nm) y 39 (458,70 nm), respectivamente. En la Ilustración 2-3 se muestran las distintas regiones del espectro electromagnético en el visible (aproximadamente entre 400 y 680 nm), borde del rojo (680-800 nm) e infrarrojo cercano (800-1500 nm).

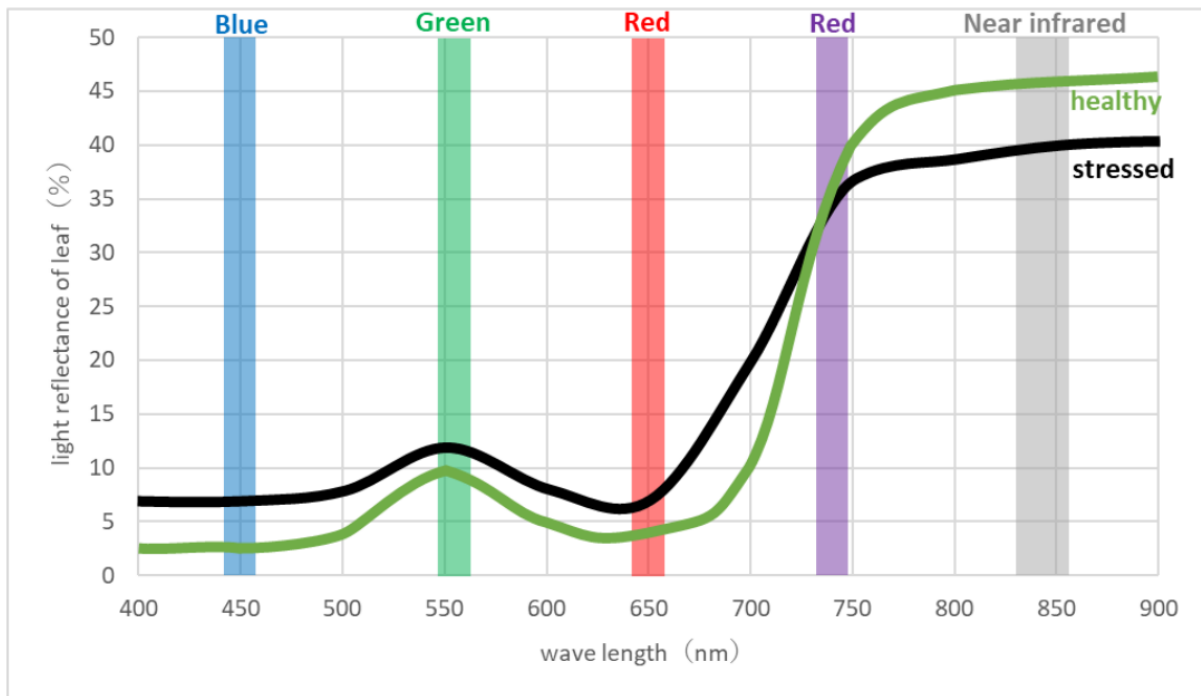


Ilustración 2-3. Regiones del espectro electromagnético entre el visible e infrarrojo cercano (tomada de (Ito, et al., 2023): <https://ceur-ws.org/Vol-3459/paper3.pdf>).

En la Ilustración 2-4 se muestra la interfaz de visualización, que además permite la consulta de los niveles digitales de los píxeles de la imagen en cada una de las bandas, mediante la obtención de la curva espectral. Por su parte en la Ilustración 2-5 se muestran algunas combinaciones de color real y falso color; y en Ilustración 2-6 las curvas espectrales extraídas de los píxeles de la imagen.

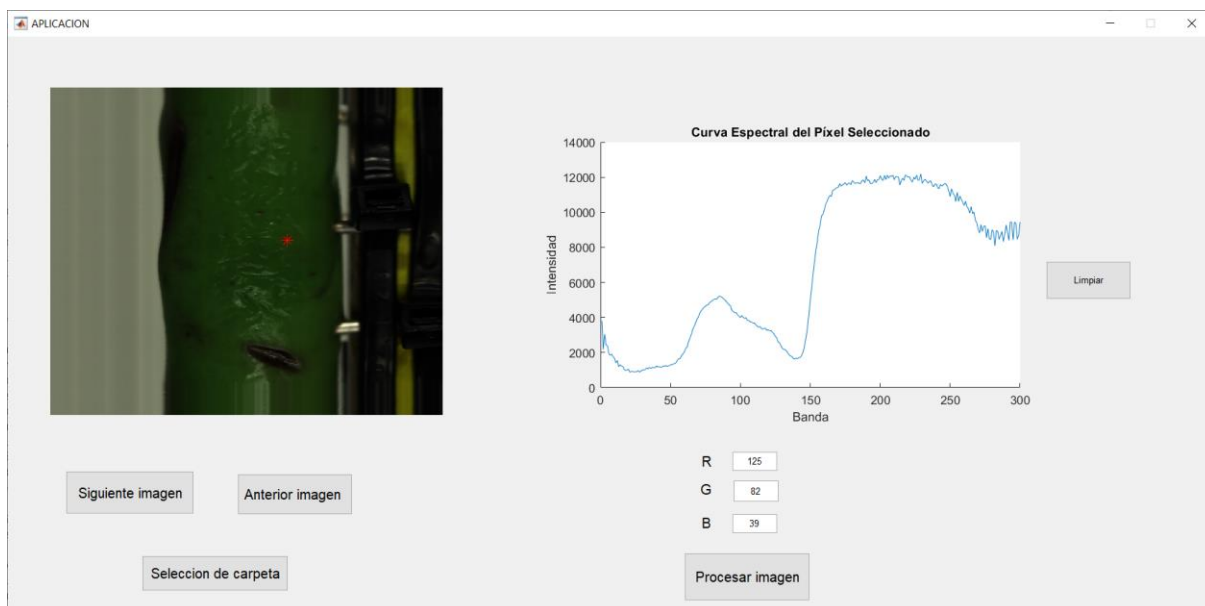


Ilustración 2-4. Interfaz de visualización y consulta de imágenes.

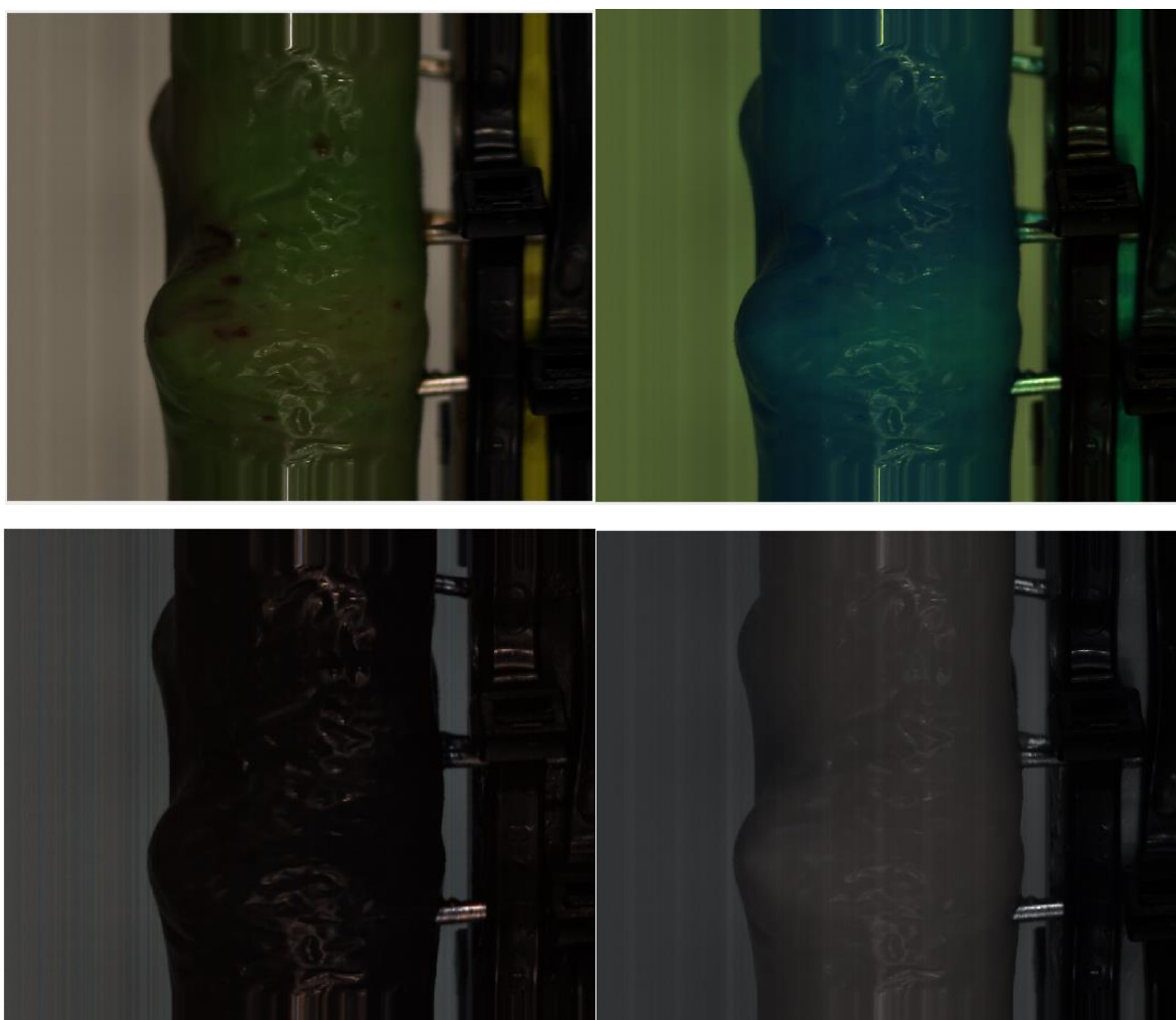


Ilustración 2-5. Diferentes combinaciones de las imágenes en color real y falso color.

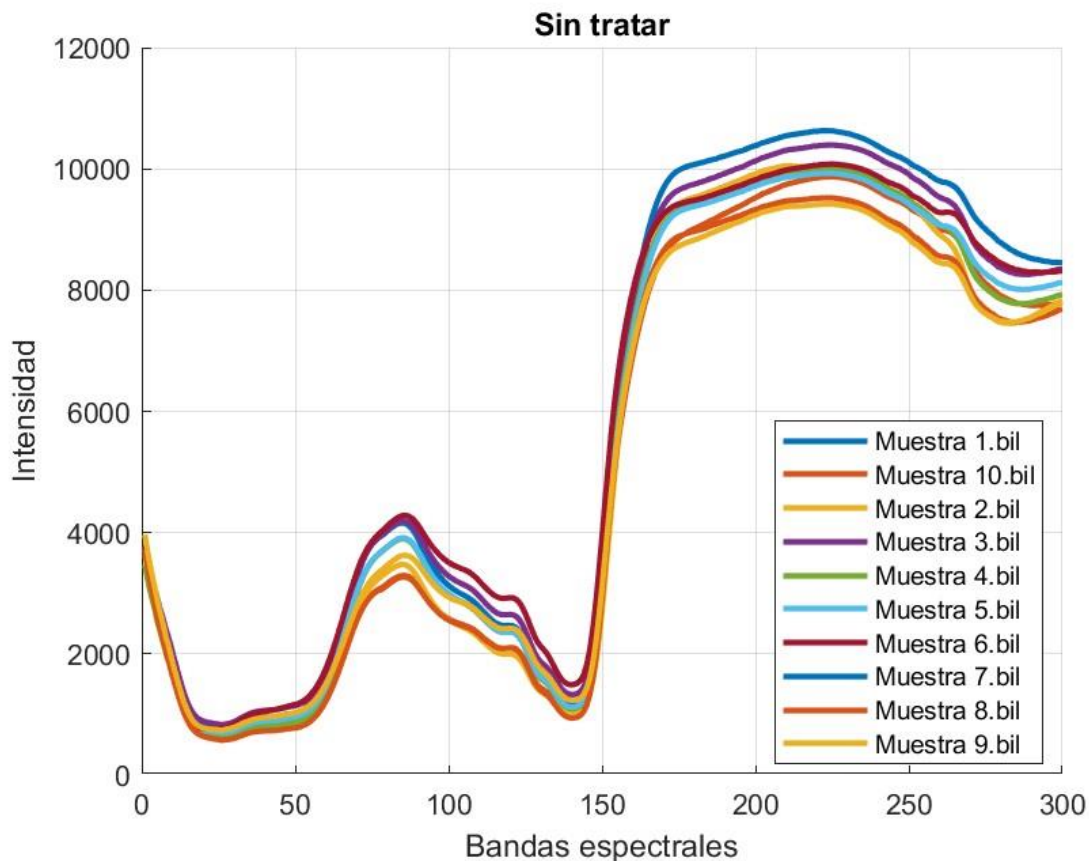


Ilustración 2-6. Curvas espectrales extraídas de los pixeles.

2.2.2. Preprocesado

Consta de dos fases principalmente: la obtención del área o región de interés de las imágenes (ROI) y la normalización a valores de B/N.

2.2.2.1. Obtención del área de interés de las imágenes

La obtención del área o región de interés (ROI) de las imágenes es una operación que se realiza habitualmente en el análisis de imágenes para centrar este en determinadas áreas, descartando el resto de la imagen, que no interesa o que directamente dificulta el análisis. Por ello es una labor de preprocesamiento que se realiza en la mayor parte de los trabajos relacionados con el análisis de imágenes hiperespectrales, en general (Paoletti, et al., 2019; Lv & Wang, 2020; Jia, et al., 2021); así como en el sector agrícola (Wang, et al., 2021; Wang, et al., 2022b; Guerri, et al., 2023; Shuai, et al., 2024) y en la detección de productos y enfermedades en las plantas y sus órganos en particular (Golhani, et al., 2018; Neupane & Baysal-Gurel, 2021).

Hay distintas estrategias para la obtención o definición de las áreas de interés dentro imágenes, una buena parte de ellas enfocadas a extraer de forma automática los objetos completos a analizar órganos completos como frutos, semillas, plantas, cultivos, etc.; o partes de ellos, como por ejemplo las áreas sanas dentro de un fruto. Las áreas de interés que permiten la extracción de frutos (aceitunas, uvas, soja, etc.), tallos o semillas (trigo, arroz, pimienta, etc.) se obtienen generalmente mediante la aplicación de umbrales en determinadas bandas, generalmente las bandas de infrarrojo cercano, aunque también en visible (Fazari, et al., 2018; Zhu, et al., 2018; He, et al., 2021; Chu, et al., 2022; Pang, et al., 2022; Wu, et al., 2022; Ye, et al., 2022; Hong, et al., 2023). En otros casos se ha aplicado una clasificación no supervisada (Sawyer, et al., 2023) o un clasificador sobre la imagen hiperespectral, como SVM (Nguyen, et al., 2021) para extraer las áreas de interés.

En este trabajo este tipo de estrategia se ha aplicado para extraer las áreas de interés de las aceitunas, que corresponden a aquellas zonas libres de defectos (picaduras, etc.), empleando una para ello una red convolucional bidimensional (CNN-2D), concretamente la DeepLabV3+ basada en la ResNet18.

La serie de redes DeepLab fueron propuestas por Liang-Chieh Chen y el equipo de Google en 2015-2017 (Chen, et al., 2015; Chen, et al., 2017; Chen, et al., 2018). Actualmente se han desarrollado cuatro versiones, que han sido sucesivas mejoras de la original:

- La primera versión (DeeplabV1) se basó en la VGG16, eliminando la última capa densamente conectada. Además, las dos últimas capas de *pooling* también se eliminaron, al no ser muy beneficiosas para la segmentación semántica, y se usa CRFs (*conditional random fields*) para mejorar la precisión de la clasificación.
- En la segunda versión (DeeplabV2) se sustituyó la VGG16 por una ResNet-101, que era más compleja y se propuso la estructura ASPP (*Atrous Spatial Pyramid Pooling*). Esta estructura realiza la operación de convolución en paralelo con diferentes tasas de muestreo para el mapa de características (Ilustración 2-7), es decir, hace una captura de la información del mapa de características multiescala, mejorando las prestaciones de la versión anterior.

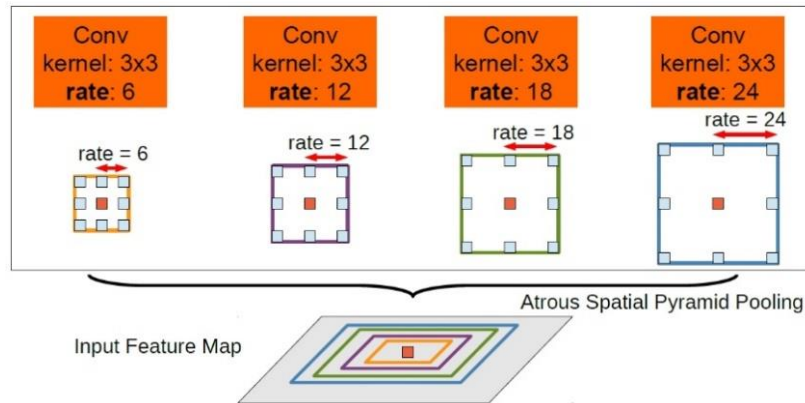


Ilustración 2-7. Red Deeplab (tomada de (Chen, et al., 2017)).

- La versión DeeplabV3 mejora la estructura ASSP y descarta el uso de CRFs, debido a que la exactitud de la clasificación ha mejorado tanto que no es necesario usarlo. Se sustituyó la capa de convolución de 3x3 por una de 1x1.
- La siguiente y última versión es la DeeplabV3+, actualizando de ResNet-101 a Xception. Normalmente se implementa con una estructura codificador-decodificador. El codificador está dividido en una red de convolución profunda y capas ASPP.

En el proyecto se ha programado en Matlab una estructura DeepLabV3+ con base ResNet18. Esta combinación se encuentra en diversos trabajos y también está recomendada en la propia web *Mathworks* de Matlab. La salida de la ResNet será la entrada de la estructura ASSP. Además, esta salida pasará a través de una capa de convolución con *kernel* 1x1 para alimentar el decodificador. En el decodificador se concatenan las funciones de extracción Resnet y ASPP para recuperar la información de entrada. Esta fusión ayuda al modelo a recuperar los bordes finos de los objetos durante el muestreo ascendente (*upsampling*), lo que mejora considerablemente el proceso de segmentación detallado. En la Ilustración 2-8 se muestra un ejemplo de la estructura completa.

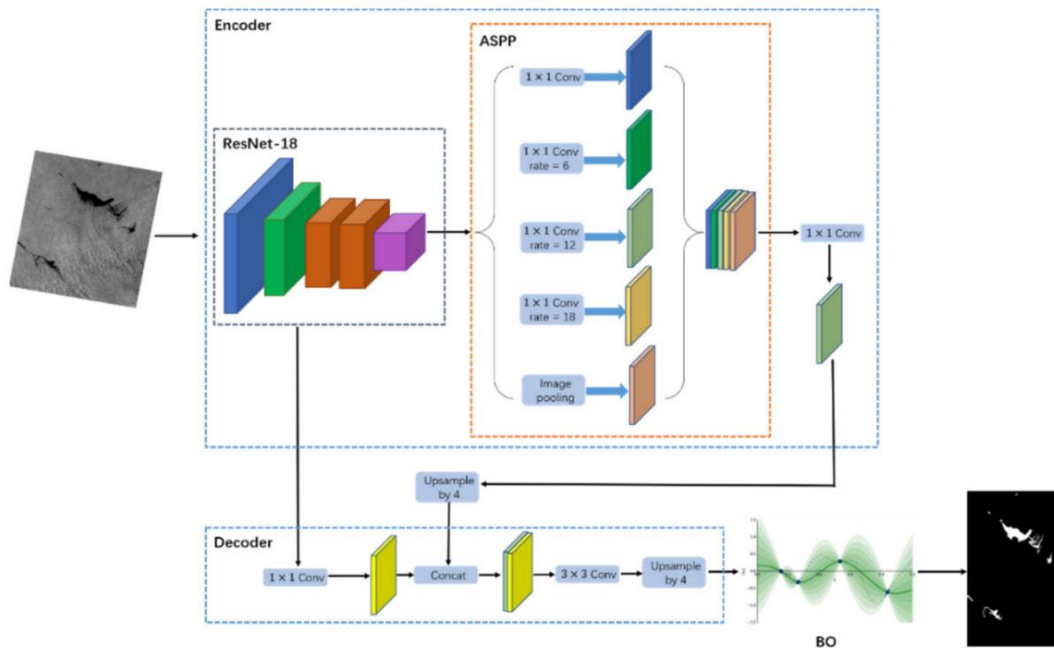


Ilustración 2-8. DeeplabV3+ (tomada de (Wang, et al., 2022c)).

La red utilizada está pre-entrenada, de manera que se podrá aprovechar el conocimiento adquirido (*Transfer Learning*), aunque para usarla en cada nueva aplicación es necesario añadir o modificar las últimas capas de las CNN de manera que se puedan ajustar los datos. Al crear la estructura es necesario además indicar el tamaño de la imagen de entrada, el número de clases y la CNN base que se utilizará, en este caso ResNet18. La estructura DeepLabV3+ con base ResNet-18 se muestra en la Ilustración 2-9, donde se puede observar cómo tras la capa de entrada se encuentra la red ResNet con sus capas y saltos característicos (bloques residuales); y, tras esta estructura, las capas ASSP seguidas de las capas decodificadoras. Se introducen el número de clases y el tamaño de la imagen de entrada de forma que la primera capa convolucional, la de entrada, llamada data y marcada en naranja, queda definida con el tamaño de entrada correspondiente.

Para esta red el tamaño de entrada será de $900 \times 750 \times 3$, es decir, el tamaño original de las imágenes en una composición RGB, puesto que el entrenamiento de esta es soportado computacionalmente, sin necesidad de hacer particiones. Se ha elegido una composición RGB puesto que el entrenamiento se ha realizado con ella, ya que permite identificar sin duda los distintos elementos a clasificar. Se valoró la utilización de la banda infrarroja, pero puesto que el resultado fue muy satisfactorio, finalmente se descartó.

Finalmente, hay que crear y reemplazar la capa de salida de clasificación de píxeles. En esta capa quedará definido el nombre de las clases y el peso (frecuencia de cada clase) que compensa el desbalance. Para esta red habrá que reemplazar la última capa *classification*, que se marca en verde en la Ilustración 2-9.

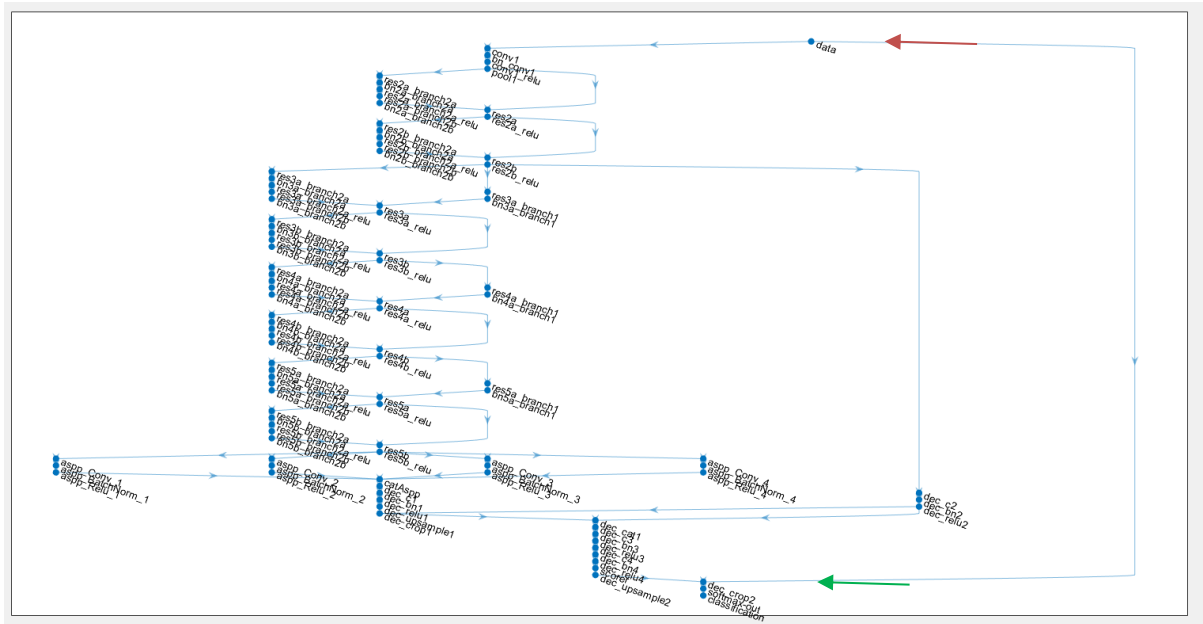


Ilustración 2-9. Estructura de la red DeepLabV3+ con base ResNet 18.

Una vez definida la red, hay que introducir algunas opciones de entrenamiento. Entre las más importantes se encuentran:

- Datos de validación
- Momentum
- Número máximo de épocas
- Tamaño del *mini-batch*, que estará limitado a la GPU que se utiliza para entrenar, por lo que se irá variando dependiendo de los requerimientos.
- Tasa de aprendizaje
- Factor de pérdida de la tasa de aprendizaje.
- Periodo para pérdida de la tasa de aprendizaje (número de épocas para que la tasa de aprendizaje caiga según el factor de caída definido).
- Frecuencia de Validación: Frecuencia para realizar el proceso de validación expresada en número de iteraciones.

- *ValidationPatience*: Indica el número de veces que la función de pérdida en el conjunto de datos de validación puede ser mayor o igual que la pérdida más pequeña anterior antes de que se detenga el entrenamiento de la red.

En este caso, se van a considerar los hiperparámetros recogidos en la Tabla 2-5.

Hiperparámetros	DeepLabV3+ con ResNet-18
Solver	Sgdm
LearnRateDropPeriod	10
LearnRateDropFactor	0.1
Momentum	0.9
InitialLearnRate	10^{-3}
L2Regularization	0.005
MaxEpochs	50
MiniBatchSize	4
ValidationPatience	8

Tabla 2-5. Hiperparámetros seleccionados en la red empleada.

Tras configurar la red y realizar el aumento de los datos, se introducirán los datos de entrenamiento para que la red entrene. Para obtener los datos de entrenamiento es necesario etiquetar una gran cantidad de imágenes. Esto debe realizarse píxel a píxel y para ello se utiliza la aplicación *Image Labeler* de Matlab (Ilustración 2-10). En ella se importa el conjunto de imágenes que se van a etiquetar y se crean 7 clases:

- Aceituna, que es la clase de mayor interés.
- Defecto de la aceituna.
- Fondos blanco, negro y amarillo, que son los distintos fondos que se pueden apreciar en las imágenes.
- Enganche, que es la pequeña vara metálica que sujeta a la aceituna.
- Error de adquisición, que se trata de la parte superior e inferior de la imagen, en los que la imagen de la aceituna aparece movida.

Con la herramienta pincel se colorean los diferentes objetos de las tres clases. Se realizó el proceso con un conjunto de datos de 115 imágenes. El etiquetamiento no tiene que ser exhaustivo, debido a la gran cantidad de imágenes que hay para etiquetar, es decir, no hay que etiquetar cada uno de los píxeles de cada una de las imágenes

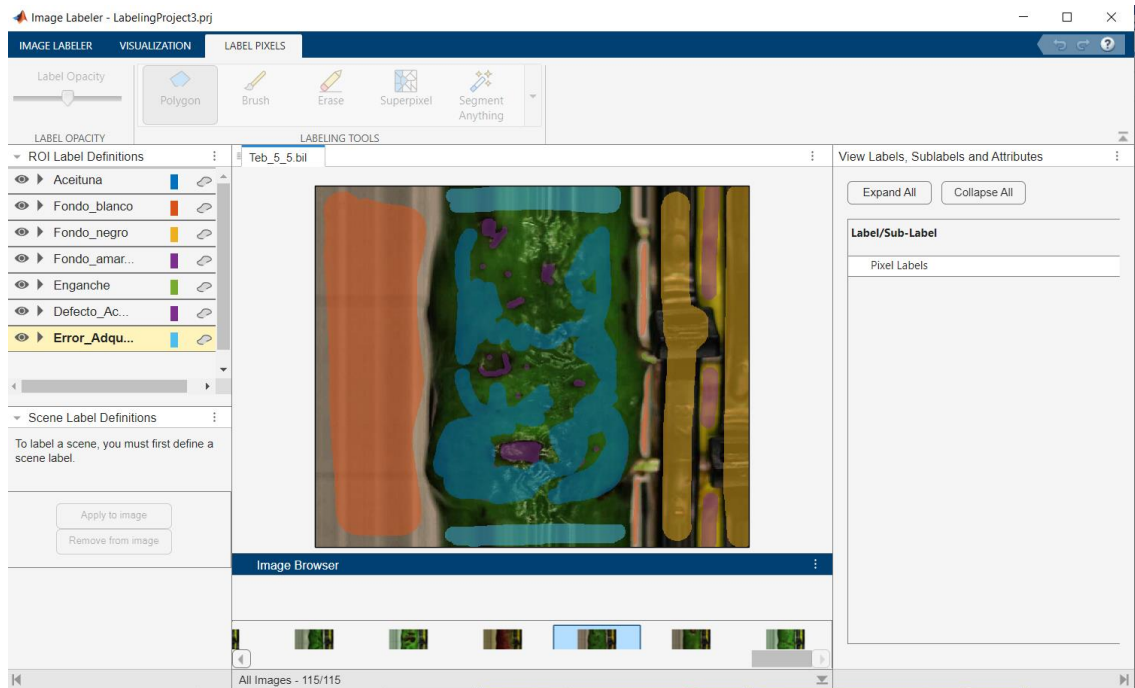


Ilustración 2-10. Interfaz Image Labeler de Matlab para etiquetamiento de píxeles.

Una vez se inicia el entrenamiento, Matlab muestra una nueva ventana llamada *Training Progress* con el proceso de entrenamiento de la red. En ella se podrá ir observando una estimación del gradiente en cada iteración y la actualización de los parámetros de la red. En los gráficos, se muestra lo siguiente:

- La exactitud de entrenamiento, es decir, la exactitud de la clasificación en cada *mini-batch*. Se muestra en la parte superior en color azul.
- Exactitud de entrenamiento suavizada, que se obtiene aplicando un algoritmo de suavizado a la exactitud de entrenamiento. Es menos ruidosa, facilitando así la detección de tendencias.
- Exactitud de validación, es decir, la exactitud en el conjunto entero de validación. Se muestra como puntos negros (cada 50 iteraciones) y línea negra discontinua, superpuestas a las anteriores en el gráfico.
- LOSS (función de coste o pérdida) de entrenamiento, LOSS de entrenamiento suavizado y LOSS de validación. Se muestra superpuestos como los tres anteriores.

Asimismo, en el panel de la derecha se muestran algunos parámetros:

- Número de iteraciones realizadas

- Tiempo de entrenamiento: comienzo y tiempo transcurrido.
- Número de épocas procesadas respecto del total.
- La frecuencia de validación expresada en número de iteraciones.
- La tasa de aprendizaje que será actualizada dependiendo de la configuración de la red.
- Recursos hardware: uso de GPU o CPU y número.

En la figura se muestra la ventana que aparece en Matlab, con los gráficos y datos indicados, así como el proceso de entrenamiento correspondiente.

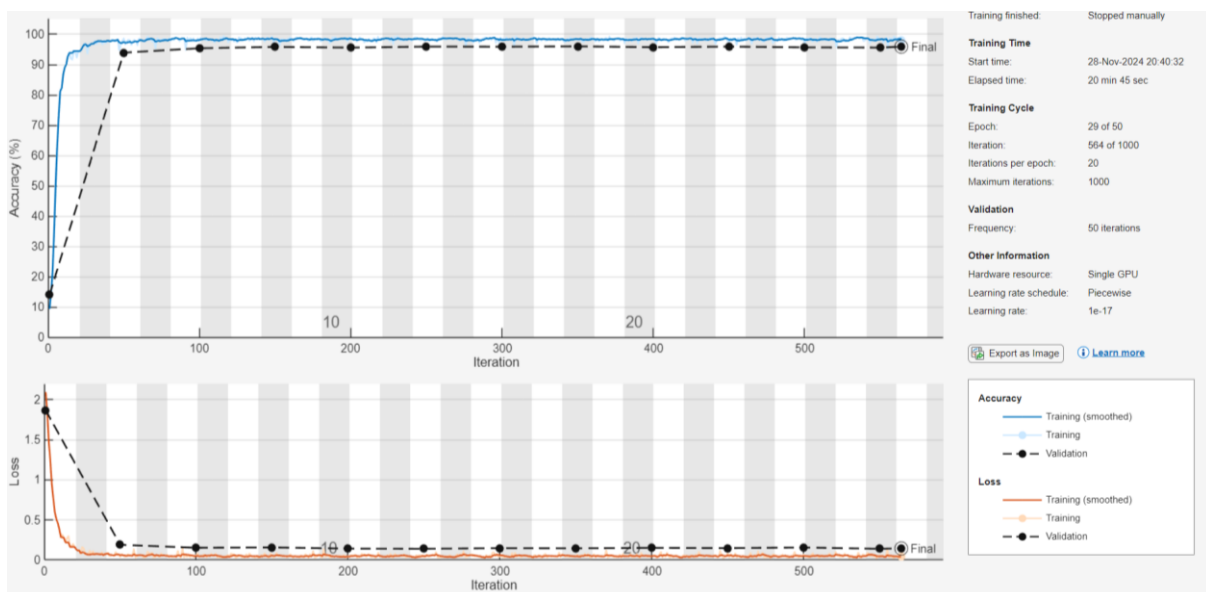


Ilustración 2-11. Progreso de entrenamiento en la red DeepLabV3+.

La curva de aprendizaje es una curva típica exponencial que se estabiliza temprano. La curva tiene cierto rizado pero las oscilaciones no son excesivas. Así, tras 29 épocas (564 iteraciones), se cumple el criterio de validación y finaliza el entrenamiento con los resultados que se muestran en la Tabla 2-6.

	Exactitud de Validación	95.91%
Tiempo de entrenamiento		20 minutos
Tasa de aprendizaje		$1 \cdot 10^{-7}$

Tabla 2-6. Datos del entrenamiento de la red DeepLabV3+.

El resultado de la clasificación de la imagen para algunas aceitunas de ejemplo se muestra en la Ilustración 2-12.

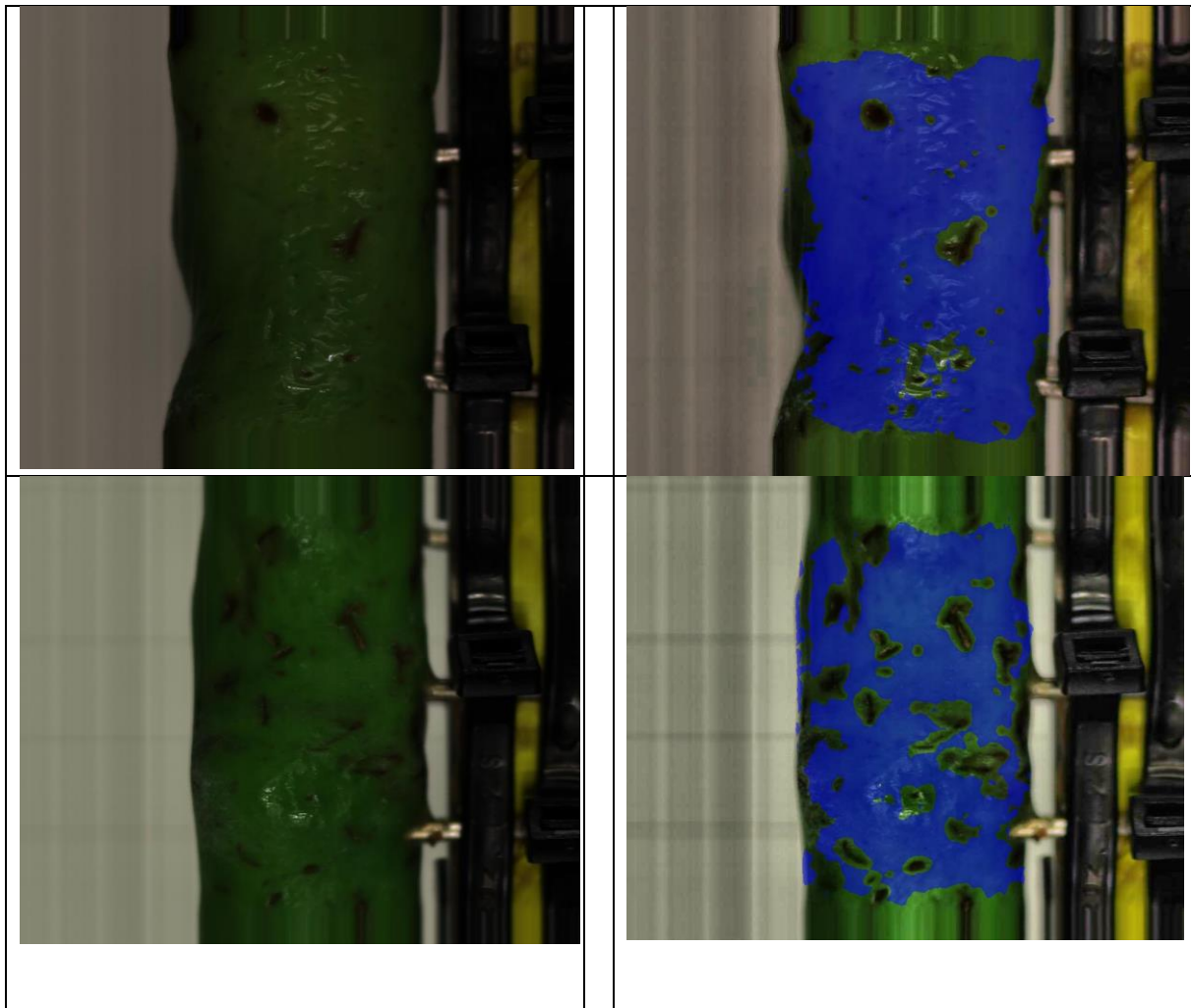


Ilustración 2-12 Imágenes clasificadas de aceitunas mediante la red DeepLabV3+ Resnet 18, se muestra en azul lo considerado como el área de interés de las aceitunas (zonas sanas sin defectos).

Visualmente se observa que los resultados son prácticamente perfectos. Los píxeles de la aceituna quedan prácticamente perfectamente etiquetados. Se destaca que incluso la parte de defecto de adquisición, donde además aparece la aceituna, es clasificada correctamente y no como aceituna, tal y como se entrenó.

En cuanto a los resultados cuantitativos, en la Tabla 2-7 se muestran los resultados globales de la exactitud para los datos de entrenamiento, validación y test; mientras que en la Ilustración 2-13 se presentan las matrices de confusión completas, con los correspondientes valores de sensibilidad y precisión. Como sucedía en el análisis visual, los resultados son excelentes.

	Entrenamiento	Validación	Test
DeepLabV3+	98.46%	95.91%	97.33%

Tabla 2-7. Exactitud o tasa de acierto para los tres conjuntos de datos.

Matriz de confusión datos de entrenamiento

True Class	Aceituna	5357490	42296	1435	117469	248	3474	739	97.0%	3.0%
	Defecto_Aceituna	2350	2246530	24	1965		34	5	99.8%	0.2%
	Enganche	64	487	185631	87	9	23	145	99.6%	0.4%
	Error_Adquisicion	102190	8231	325	1039144	51	1989	512	90.2%	9.8%
	Fondo_amarillo	26	54	2774		1024255	36	1004	99.6%	0.4%
	Fondo_blanco	10444	14016	34712	1275	8470	10961187	13984	99.2%	0.8%
	Fondo_negro	3	2197	3838	514	8301	714	3918531	99.6%	0.4%

97.9%	97.1%	81.2%	89.5%	98.4%	99.9%	99.6%
2.1%	2.9%	18.8%	10.5%	1.6%	0.1%	0.4%

Aceituna
 Defecto_Aceituna
 Enganche
 Error_Adquisicion
 Fondo_amarillo
 Fondo_blanco
 Fondo_negro

Predicted Class

Matriz de confusión datos de validación

True Class	Aceituna	1200043	17224	473	124632		2810	97	89.2%	10.8%
	Defecto_Aceituna	24157	482068	159	4942		2659		93.8%	6.2%
	Enganche	24		25534	51	76	142	846	95.7%	4.3%
	Error_Adquisicion	22290	4118	86	166628		4	355	86.1%	13.9%
	Fondo_amarillo			192		184848	1253	1316	98.5%	1.5%
	Fondo_blanco	1169	4312	7692	189	395	2370521	1689	99.4%	0.6%
	Fondo_negro		1768	3293	43	324	2576	1000062	99.2%	0.8%

96.2%	94.6%	68.2%	56.2%	99.6%	99.6%	99.6%
3.8%	5.4%	31.8%	43.8%	0.4%	0.4%	0.4%

Aceituna
 Defecto_Aceituna
 Enganche
 Error_Adquisicion
 Fondo_amarillo
 Fondo_blanco
 Fondo_negro

Predicted Class

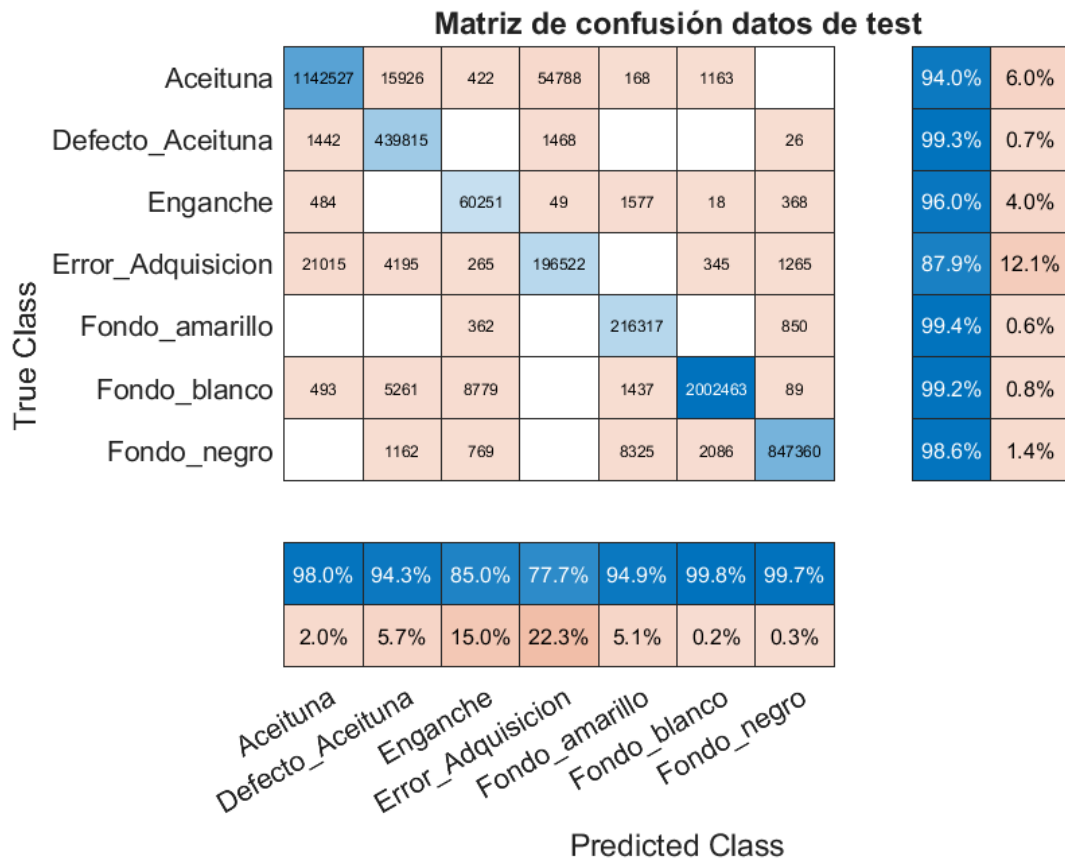


Ilustración 2-13. Matrices de confusión para los datos de entrenamiento, validación y test en la red DeepLabV3+ empleada para la obtención de las áreas de interés de las imágenes.

En el caso de las hojas se ha seguido una estrategia diferente, consistente en realizar una umbralización, dado su mayor definición en la imagen frente al fondo y otros elementos que aparecen en la imagen (sujeciones, etc.).

Tras probar con algunas bandas, se decidió utilizar el índice NDVI (Rouse, et al., 1974; Bannari, et al., 1995), que ofreció muy buenos resultados visualmente. Este índice es ampliamente utilizado para caracterizar el estado (vigor, verdor, humeda, etc.) de la vegetación y se define mediante la siguiente expresión:

$$NDVI = \frac{\rho NIR - \rho Rojo}{\rho NIR + \rho Rojo}$$

Ecuación 2-1

Donde:

ρ NIR: Reflectancia en infrarrojo cercano (NIR); ρ Rojo: Reflectancia en la banda del rojo.

Valores positivos indican cierto vigor vegetal, tanto mayor cuanto más se acerque a 1; mientras que los valores próximos a 0 son característicos del suelo o áreas urbanas y los negativos de suelos húmedos o algunos tipos de rocas.

La definición de las áreas de interés de las imágenes de las hojas a partir de la umbralización del NDVI, se realiza mediante una sencilla aplicación de Matlab en la que se introducen las imágenes de las hojas, se aplica la expresión anterior del NDVI y se fija un umbral 0,3, determinado que todos los píxeles con valores superiores de este umbral son hojas. Este valor se determinó mediante la realización de pruebas de ensayo-error, pero tiene significación física, puesto que como se ha apuntado los valores positivos de NDVI son indicativos de presencia de una vegetación con un cierto vigor (Rouse, et al., 1974; Bannari, et al., 1995; Jensen, 2009).

En la Ilustración 2-14 se muestra la definición de las áreas de interés de las hojas, habiendo varias hojas por imagen.

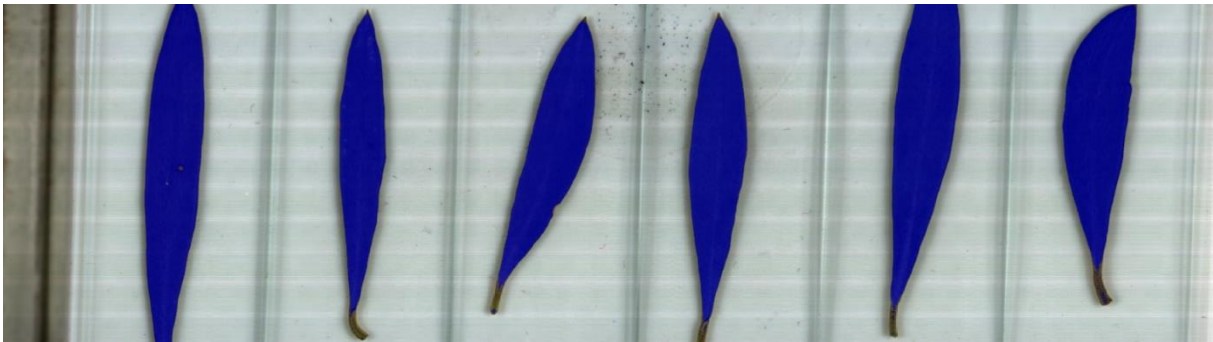


Ilustración 2-14. Obtención de las áreas de interés en las hojas mediante umbralización del NDVI.

2.2.2.2. Corrección B/N y normalización

Normalmente, en el análisis de imágenes se suele trabajar con valores relativos de la respuesta de cada área iluminada por una radiación electromagnética. El parámetro que se emplea habitualmente es la reflectancia que es la relación entre el flujo reflejado por una superficie y el flujo incidente, es decir una variable relativa y adimensional que varía entre 0 y 1 (Jensen, 2009). En imágenes de exterior, la reflectancia se calcula a partir de la respuesta o radiancia que llega al sensor, reflejada por la superficie terrestre o los objetos situados sobre ella y la radiación incidente procedente del sol. Para ello se han de conocer una serie de parámetros como la radiación solar extraterrestre de cada día, el ángulo de incidencia, la posición del satélite o plataforma aérea, así como parámetros de calibración del sensor en cada una de las longitudes de onda (referencias), por lo que las correcciones y calibraciones son realizadas usualmente por las agencias espaciales (NASA o ESA) o los servicios que suministran las imágenes (USGS, Copernicus, etc.). En este caso, estas operaciones suelen ir acompañadas de correcciones atmosféricas que tienen en cuenta el estado de la atmósfera en cada zona y momento (Jensen, 2009).

En el caso de las imágenes tomadas en laboratorio, en los que la iluminación se proporciona mediante un sistema propio en cada caso, y en condiciones diferentes, se suele optar por calcular la reflectancia de forma relativa comparando el valor de la respuesta obtenido en cada punto de la superficie (píxeles) y el valor obtenido en una superficie negra y blanca. Es la conocida como corrección a blanco y negro o corrección al valor oscuro (*dark correction*), que se emplea en numerosos trabajos, entre algunos de los consultados (Barbedo, et al., 2015; Zhu, et al., 2019; Gao, et al., 2020; Liu, et al., 2020a; He, et al., 2021; Su, et al., 2021; Chen, et al., 2022; Chu, et al., 2022; Onmankhong, et al., 2022; Pang, et al., 2022) (Ye, et al., 2022; Wu, et al., 2022; Zhang, et al., 2022; Domínguez-Cid, et al., 2023; Zhu, et al., 2023). Para el negro se suele emplear la respuesta del sensor cuando se tapa, con una reflectancia próxima a 0; y para el blanco, alguna superficie blanca con reflectancia próxima a 1.

Así, la expresión que suele emplearse es la siguiente:

$$I_c = \frac{I_o - N}{B - N}$$

Ecuación 2-2

Donde: I_c es la imagen corregida; I_o es la imagen original (*raw image*); N es la imagen negra, generalmente la imagen adquirida al cubrir la lente de la cámara con una tapa opaca con una reflectancia o reflectividad de casi el 0 %; B es la imagen adquirida al escanear una superficie blanca (por ejemplo, una placa de teflón blanca) con una reflectancia de casi el 100 %.

Otra alternativa que emplean algunos trabajos es aplicar una expresión equivalente, cuando no se ha adquirido una imagen sobre una superficie blanca (teflón o similar) y negra (tapa); en este caso se utilizan los valores máximos y mínimos de la imagen en cada una de las bandas espectrales, que asegura una cierta normalización, aun cuando las condiciones de iluminación no sean las mismas para todas las imágenes (Sheela & Deepa, 2013; Chu, et al., 2022). En concreto, la expresión aplicada es:

$$V_N = \frac{V_o - Min}{Max - Min}$$

Ecuación 2-3

Donde: I_c/o es la imagen corregida o en su caso original; Max es el valor máximo en cada imagen; Min es el valor mínimo en cada imagen.

Esta operación de normalización es la que se ha aplicado en este trabajo, correspondiendo los V_o a los valores medios de la reflectancia en cada banda de la imagen en las aceitunas o las hojas, mientras que el máximo corresponde a los valores medios de la placa blanca del fondo y el mínimo a los valores medios de las partes más oscuras, como los cinturones de sujeción o sombras acusadas en la imagen. La operación se ha programado en Matlab y permitirá la obtención de las curvas espectrales normalizadas.

2.2.3. Obtención de las curvas espectrales

La obtención de las curvas espectrales, curvas de reflectancia o firmas espectrales es una operación básica en los estudios en los que se emplean técnicas multiespectrales o hiperespectrales, especialmente en estos últimos (Paoletti, et al., 2019; Lv & Wang, 2020; Jia, et al., 2021). Específicamente, los estudios relacionados con el sector agrícola también parte de esta operación (Wang, et al., 2021; Wang, et al., 2022b; Guerri, et al., 2023; Shuai, et al., 2024).

Consecuentemente, en este estudio, la obtención de las curvas espectrales de cada pixel o las curvas medias de un conjunto de ellos es la operación más importante, ya que todos los análisis derivan de ella. A continuación, se explican las técnicas seguidas para llevar a cabo esta operación, tanto a nivel de pixel, como de áreas de interés o ventanas.

2.2.3.1. Consulta y extracción de las curvas espectrales de cada pixel

Como se ha apuntado anteriormente, la consulta de los valores de reflectancia de los píxeles de la imagen en cada banda se realiza mediante la interfaz de visualización y consulta, que permite observar y analizar la curva espectral asociada a cada pixel de forma interactiva, a medida que se recorre la imagen con el cursor.

En la Ilustración 2-4 se mostró la interfaz de consulta desarrollada en Matlab y en la Ilustración 2-6 un ejemplo de las curvas espectrales asociadas a cada pixel.

En algunos casos, lo que se hace es extraer los valores de reflectancia en cada pixel o conjunto de píxeles, lo que permite su análisis más exhaustivo, mediante la utilización de técnicas de aprendizaje automático y profundo. Este análisis permite detectar propiedades de los objetos observados en la imagen, como una enfermedad (Nguyen, et al., 2021); o clasificarlas de acuerdo a un criterio determinado, como variedades de aceitunas (Domínguez-Cid, et al., 2023). La extracción se puede hacer sobre toda la imagen o una parte de ella, o bien seleccionando manualmente los píxeles en la imagen (Domínguez-Cid, et al., 2023) o de forma aleatoria (Nguyen, et al., 2021).

2.2.3.2. Obtención de las curvas espectrales medias de un área de interés

Esta práctica es empleada en multitud de trabajos, bien para analizar de forma visual las curvas espectrales, o bien como base para la aplicación de redes convolucionales, generalmente 1D (aunque también 3D), que permiten realizar labores de identificación o clasificación. La mayor parte de las veces lo que se obtienen es la curva espectral media de toda imagen o de una región de interés. Así, las curvas espectrales medias han sido obtenidas en diversos trabajos relacionados con la detección de fitosanitarios (He, et al., 2021; Chu, et al., 2022; Ye, et al., 2022; Zhang, et al., 2022) en diversos tipos de granos y plantas, enfermedades en frutos

(Cui, et al., 2022), clasificación de variedades en granos y bayas (Zhu, et al., 2019; Zhang, et al., 2020c; Gao, et al., 2021; Dong, et al., 2022; Onmankhong, et al., 2022) u otros criterios de clasificación como el vigor, la madurez, etc. (Gao, et al., 2020; Liu, et al., 2020a; Su, et al., 2021; Pang, et al., 2022; Wu, et al., 2022; Hong, et al., 2023; Nguyen, et al., 2023; Wei, et al., 2023; Zhu, et al., 2023).

En este trabajo, los valores de reflectancia media de cada banda en un área de interés representativa del total de las imágenes de las muestras de aceitunas se obtienen mediante una aplicación de Matlab, que exporta los resultados a una hoja de Excel. El área de interés corresponde en este caso a la parte sana y libre de defectos de una aceituna o a una hoja, que había sido extraída automáticamente mediante la red Deeplabv3+.

Otras formas de proceder son la obtención del espectro medio de una ventana que se defina de forma interactiva sobre la imagen (Ren, et al., 2018; Yang, et al., 2021; Zhang, et al., 2022; Zhang, et al., 2020c) o de forma aleatoria sobre un conjunto de píxeles aislados o agrupados (muestreo de conglomerados). Estos procedimientos se explican en los dos siguientes apartados.

2.2.3.3. Obtención de las curvas espectrales medias de una ventana definida por el usuario

La obtención de las curvas espectrales medias en una ventana definida por el usuario se realiza en este trabajo mediante una aplicación en Matlab que permite definir interactivamente las ventanas sobre una interfaz, en la que se van visualizando las imágenes de las aceitunas y las hojas. Una vez definida la ventana, los valores medios de reflectancia en cada banda se envían a una hoja Excel.

El objetivo de crear estas ventanas fue generar un aumento de datos (*data augmentation*) ante el bajo tamaño de la muestra con las imágenes completas (o sus áreas de interés), que hacía que el proceso de entrenamiento no se llevara a cabo de forma correcta y menos aún la validación y test, para lo que se disponía de una muestra menor aún. Se trató de hacer un aumento de datos típico añadiendo ruido a los datos de entrenamiento, pero no fue suficiente, por lo que se decide de hacer un aumento de datos mediante la extracción de los espectros medios de ventanas

definidas manualmente en las imágenes. Con la creación de estas ventanas se consiguió aumentar la muestra por 10, ya que se crearon 10 ventanas por imagen, aunque lógicamente se podrían las que se considere oportuno.

Para ello se desarrolló una interfaz mediante la aplicación de *Matlab App Designer*, en las que el usuario puede etiquetar ventanas de la imagen, calculándose los valores medios de reflectancia en cada banda, que se almacenan etiquetados en un Excel, todo de manera automática. La interfaz de la aplicación se muestra en la Ilustración 2-15, junto con un ejemplo de creación de una ventana. Esta incluye unos botones para seleccionar la carpeta donde se encuentran las imágenes de cada tipo de fitosanitario y nivel de concentración, para definir las ventanas (etiquetar) y para grabar los resultados en la hoja Excel, lo que se realiza secuencialmente a medida que van visualizándose y creándose las ventanas. También se incluyen distintas opciones para etiquetar las aceitunas, así como los fondos negro y blanco que se identifiquen en la imagen.

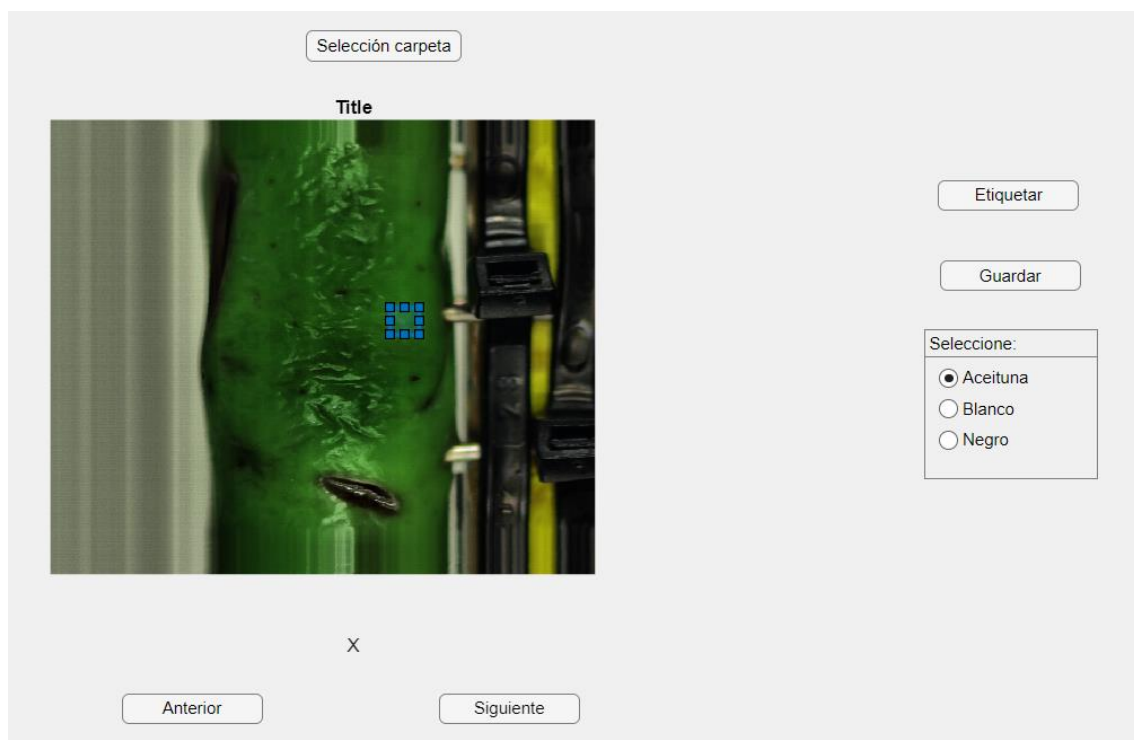


Ilustración 2-15. Interfaz de la aplicación para la extracción de los valores medios de reflectancia en ventanas definidas manualmente.

2.2.3.4. Obtención de las curvas espectrales en cuadrículas creadas automáticamente

En este caso, se desarrolla una aplicación de Matlab en la que se definen de forma automática las cuadrículas a partir de unos datos de entrada. Esto, permitirá aumentar el tamaño de la muestra, pues no será necesario realizar el recorte de cada imagen, además de agilizar el proceso. Para su aplicación es necesario disponer de la definición de las áreas de interés, con el objeto de definir las cuadrículas dentro de ellas y no capturar datos que no interesen (fondo, defectos, errores, etc.).

Una vez obtenida el área de interés es esta la que se divide en cuadrículas, a las que se les calculan las curvas espectrales o más bien los valores medios de reflectancia de cada banda, que se envían a una hoja Excel.

Para el caso de las aceitunas se obtuvieron cuadrículas de 25 x 25 píxeles, lo que dio lugar a un total de 63750 muestras. En el caso de las hojas, se decidió reducir el número de cuadrículas, por lo que se utilizan unas de 50 x 50 píxeles, que dan lugar a 6970 cuadrículas. Las cuadrículas obtenidas para aceitunas y hojas se muestran la Ilustración 2-16.

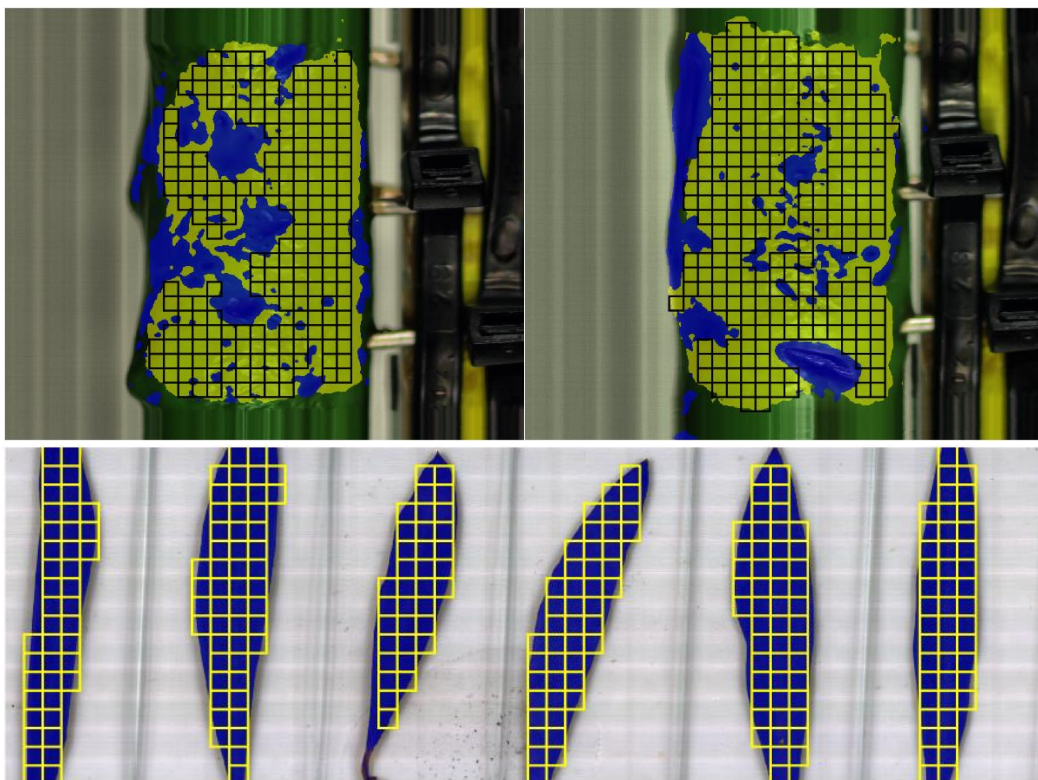


Ilustración 2-16. Creación de las cuadrículas definidas automáticamente para la extracción de los valores medios de reflectancia en aceitunas y hojas.

2.2.4. Selección de características

Los métodos de selección de características consisten en una serie de procesos por los que se eligen un subconjunto de variables relevantes de un conjunto de datos para mejorar el rendimiento de modelos de aprendizaje automático (Guyon & Elisseeff, 2003), buscando reducir la dimensionalidad, minimizar el riesgo de sobreajuste y mejorar la interpretabilidad de los modelos.

La selección de características es una práctica muy habitual en los estudios que emplean imágenes hiperespectrales, para reducir sus dimensiones, ya que al número de píxeles (matriz 2D) hay que añadir en cada una de ellas un número importante de bandas. La selección se aplica generalmente para reducir el número de bandas, ya que muchas de ellas presentan una elevada correlación entre sí, por lo que podría hacerse una selección sin reducir excesivamente la información. Se suele aplicar como paso previo a los métodos de identificación y clasificación que se realizan mediante métodos de aprendizaje automático.

Según se indicó en la introducción, hay diversos métodos de selección de características (Guyon & Elisseeff, 2003):

- Métodos de filtro: Se basan en realizar la selección de forma independiente del modelo, empleando métricas estadísticas como la correlación. Entre ellos el más conocido es el factor de inflación de la varianza (VIF).
- Métodos de envoltura: Evalúan el rendimiento del modelo usando diferentes subconjuntos de características. Como ejemplo se puede citar la eliminación recursiva de características (RFE).
- Métodos de incrustación: Integran la selección de características en el proceso de entrenamiento del modelo. Dentro de ellos está el LASSO o los métodos de clasificación de variables, que permiten evaluar además su importancia, como los árboles de decisión o los bosques aleatorios.

En los estudios consultados en este trabajo, centrados en la aplicación de técnicas hiperespectrales al sector agrícola, se han encontrados numerosos ejemplos de aplicación de estas técnicas de selección de características. Los más sencillos son aquellos en los que se seleccionan directamente algunas bandas, especialmente las

RGB o el infrarrojo de cercano para hacer composiciones en color o como paso a otros análisis en cultivos, variedades de frutos (como aceitunas) y granos (Yang, et al., 2021; Onmankhong, et al., 2022; Domínguez-Cid, et al., 2023).

Sin embargo, el método más aplicado es el análisis de componentes principales, que realmente no es puramente un método de selección de características, sino de reducción de la dimensionalidad (número de variables) a través de la transformación de las bandas originales de la imagen en unas nuevas bandas que no presentan correlación entre sí, a través de una rotación de los ejes dimensionales (Jolliffe, 2002). Como muestra, en los trabajos consultados se ha aplicado este análisis a la detección de pesticidas (Ye, et al., 2022); detección de enfermedades y patógenos (Barbedo, et al., 2015; Nguyen, et al., 2021; Qi, et al., 2023); identificación y clasificación de granos y frutos (Zhu, et al., 2019; Gao, et al., 2020; Zhang, et al., 2020c; Gao, et al., 2021; Chen, et al., 2022; Onmankhong, et al., 2022; Wei, et al., 2023); o la clasificación de cultivos y usos del suelo (Xu, et al., 2018; Haut, et al., 2019; Ladi, et al., 2022; Patel, et al., 2023). En algunos casos se ha aplicado el método de incrustación de vecinos estocásticos distribuidos en t (t-SNE), que es una especie de análisis de componentes principales no lineal para clasificación de usos del suelo (Zhong, et al., 2017; Meng, et al., 2021).

También son numerosos los trabajos en los que se aplican métodos secuenciales de selección como la selección secuencial de características o SFS (Gao, et al., 2020); la eliminación de variables no informativas, UVE (Pang, et al., 2022); el algoritmo de proyección sucesiva, SPA (Pang, et al., 2022; Wu, et al., 2022; Wei, et al., 2023); los algoritmos genéticos (Wei, et al., 2023); o el muestreo reponderado adaptativo competitivo, CARS (Liu, et al., 2020a; Dong, et al., 2022). Estos métodos se aplican a la identificación y clasificación de distintos tipos de granos y frutos.

Finalmente, se han consultado trabajos que aplican métodos de incrustación basados en árboles aleatorios (Sawyer, et al., 2023), o las propias redes convolucionales 1D (Zhang, et al., 2020c; Chu, et al., 2022) y 2D-3D (Nguyen, et al., 2021) a la identificación de granos o la detección de fitosanitarios y enfermedades.

Asimismo, en algún trabajo se aplican técnicas novedosas de aprendizaje profundo como los mapas de saliencia (Su, et al., 2021).

En este trabajo se han probado distintos métodos, como el análisis de componentes principales (PCA), el análisis de la varianza (ANOVA), la eliminación recursiva de características (RFE), los árboles aleatorios (RF), y un método basado en el aprendizaje profundo (red CNN-1D), concretamente el conocido como técnica de explicaciones independientes del modelo y localmente interpretables (LIME). Los dos primeros se aplican mediante el *Classification Learner* de Matlab, como paso previo a los métodos de aprendizaje automático, y los tres últimos se han programado como aplicaciones igualmente en Matlab.

Con carácter previo, se ha de indicar que se ha realizado una eliminación por precaución de las primeras y últimas bandas, siguiendo la recomendación de muchos trabajos consultados, entre otros (Zhang, et al., 2020c; Chu, et al., 2022; Onmankhong, et al., 2022; Wu, et al., 2022; Ye, et al., 2022; Hong, et al., 2023). La razón es que estas bandas suelen introducir mucho ruido e interferencias.

2.2.4.1. Análisis de componentes principales

El análisis de componentes principales, como ya se ha indicado no es realmente un método de selección de características, sino de reducción de la dimensionalidad, basado en la transformación de las bandas originales de la imagen en unas nuevas bandas que no presentan correlación entre sí (Jolliffe, 2002).

En este caso, se ha utilizado el análisis de componentes principales incluido en el *Classification Learner* de Matlab. Una vez introducidos los datos, mediante un archivo Excel o un archivo de texto (CSV), se elige la herramienta PCA en el menú, abriéndose una ventana como la que se muestra en la Ilustración 2-17.

En la ventana, se ha de seleccionar el criterio de reducción de las componentes, que puede ser la varianza explicada por las componentes principales o el número de componentes principales a obtener. En este caso, se ha empleado la varianza explicada (pruebas con 90, 95 y 99%); una vez seleccionado, ya se puede aplicar el modelo de clasificación, que tendrá en cuenta las componentes principales.

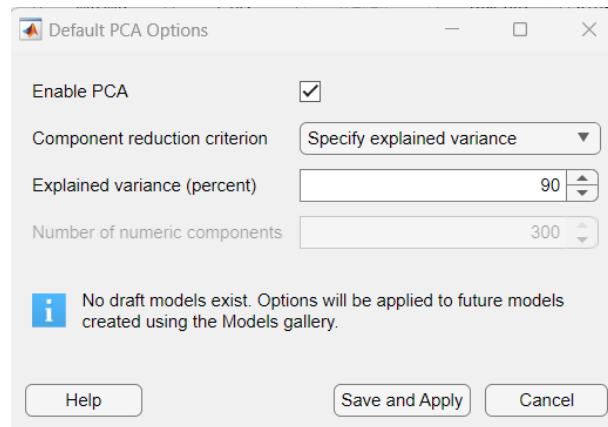


Ilustración 2-17. Ventana del análisis de componentes principales (PCA) de Matlab.

Entre los resultados del modelo se pueden consultar el número de componentes principales suficientes para alcanzar la varianza explicada (4 componentes para alcanzar el 90% en el ejemplo) y el porcentaje que explica cada una de las componentes, como se observa en la Ilustración 2-18.

Model 5: SVM
Status: Trained

Training Results
Accuracy (Validation) 55.7%
Total cost (Validation) 851
Prediction speed ~8800 obs/sec
Training time 5.4207 sec

▶ **Model Hyperparameters**

▶ **Feature Selection: 300/300 individual features selected**

▼ **PCA: Explaining at least 95% variance**

For PCA training, 150 of 1920 observations were ignored because they contained Infs or NaNs.
PCA is keeping enough components to explain 95% variance.
After training, 4 components were kept.
Explained variance per component (in order): 69.0%, 19.2%, 4.2%, 3.8%, 1.9%, 0.7%, 0.4%, 0.3%, 0.2%, 0.1% (variances of least important components hidden)

▶ **Misclassification Costs: Default**

▶ **Optimizer: Not applicable**

Ilustración 2-18. Componentes principales y varianza explicada.

2.2.4.2. Análisis de la varianza

El análisis de la varianza (ANOVA) es un método iterativo de clasificación o agrupación (*cluster*) de variables que consiste básicamente en minimizar la varianza entre los elementos de una clase (intraclase o intragrupo), maximizando la varianza entre los valores medios de las distintas clases (interclase o intergrupo).

Se puede aplicar como método de selección de características, evaluando para cada característica la relación entre ella y la variable objetivo, es decir, determinando la significancia estadística, usualmente a través de un p-valor. Así, se seleccionarán

las características con p-valores inferiores a un umbral (como 0,05), al considerarse estadísticamente significativas (James, et al., 2017).

Como se ha indicado, el análisis de la varianza se aplica mediante el *Classification Learner* de Matlab, eligiendo la herramienta *Feature Selection* desde el menú principal. Una vez introducidos los datos y seleccionada la herramienta, se abre una ventana como la que se muestra en la Ilustración 2-19.

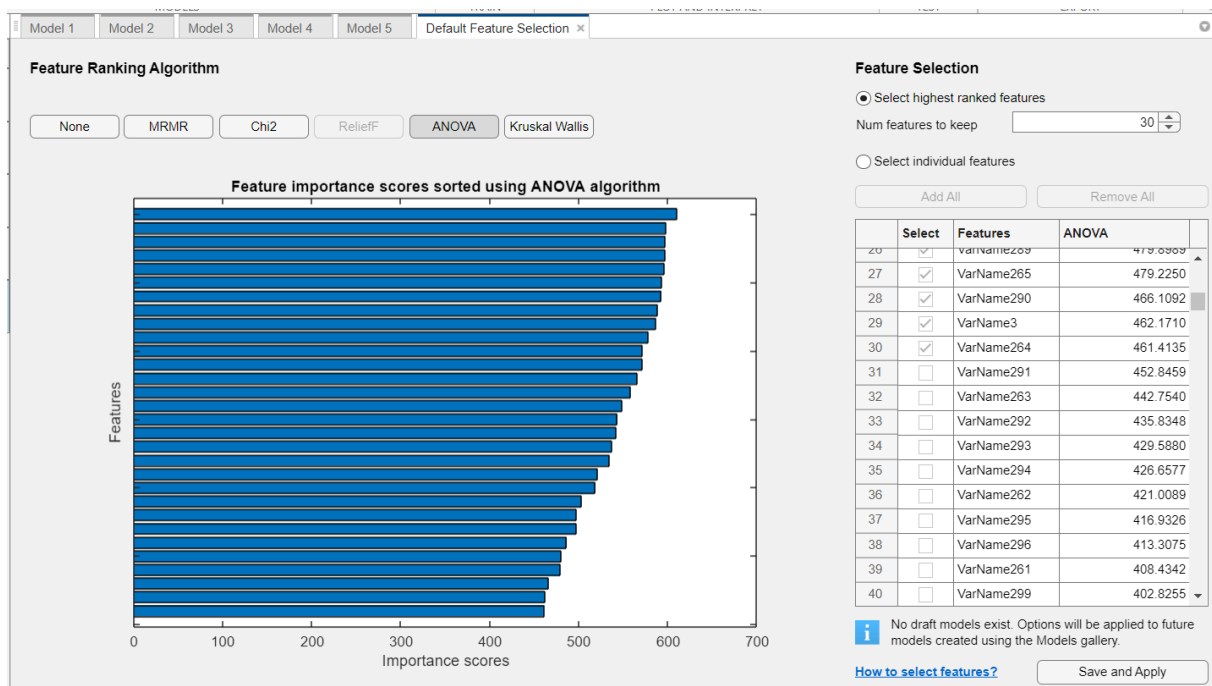


Ilustración 2-19. Ventana de Selección de Características (Feature Selection) en Matlab.

En la ventana, lo primero que hay que elegir es el método o algoritmo de establecer el ranking de características, en este caso el análisis de la varianza, ANOVA (aunque hay otros como Chi-2, Kruskal-Wallis, etc.). Además, se puede elegir el número de características a considerar (por ejemplo, 30) y las características concretar a mantener a la hora de aplicar el modelo, normalmente las (30) primeras. Aparece una tabla con las características (bandas) ordenadas decrecientemente en función de su importancia, así como su valor de importancia.

Una vez salvado, estas serán las características que se utilicen en el modelo a aplicar. Al hacerlo, aparece la información correspondiente, con las características a y su importancia considerar donde nuevamente se pueden cambiar (Ilustración 2-20).

The screenshot displays the configuration for 'Model 8: SVM'. The status is 'Draft'. Under 'Model Hyperparameters', the following settings are visible:

- Kernel function: Linear
- Box constraint level: 1
- Kernel scale mode: Auto
- Manual kernel scale: 1
- Multiclass method: One-vs-One
- Standardize data: Yes

Under 'Feature Selection: Top 30/300 features selected using ANOVA', the option 'Select highest ranked features' is chosen, with 'Num features to keep' set to 30. Below this, there are 'Add All' and 'Remove All' buttons, and a table showing the selected features:

	Select	Features	ANOVA
1	<input checked="" type="checkbox"/>	VarName1	610.1217
2	<input checked="" type="checkbox"/>	VarName275	597.8974
3	<input checked="" type="checkbox"/>	VarName272	597.2572

Ilustración 2-20. Características seleccionadas para el modelo.

2.2.4.3. Eliminación recursiva de características

La eliminación recursiva de características (*recursive feature elimination*, RFE) es un método de selección de características ampliamente utilizado en el aprendizaje automático, con el objetivo de identificar el subconjunto óptimo de características relevantes para mejorar el rendimiento del modelo mientras se descartan las irrelevantes o redundantes. El proceso sigue un enfoque iterativo con los siguientes pasos (Chen & Jeong, 2007):

1. Entrenamiento inicial: Se entrena un modelo usando todas las características disponibles.
2. Evaluación de importancia: Se mide la importancia de cada característica (por ejemplo, mediante coeficientes en modelos lineales o puntuaciones de importancia en árboles de decisión).
3. Eliminación de características menos importantes: Se elimina la característica menos importante o un subconjunto de características con menor contribución.
4. Repetición: Se entrena nuevamente el modelo con el subconjunto restante de características y se repite el proceso hasta alcanzar un número deseado de características o un umbral de rendimiento aceptable.

En este caso, el método se ha programado en Matlab mediante la función *sequentialfs*, en la que se han de introducir los datos (en un archivo Excel o de texto) y seleccionar la opción de *backward feature selection*, que es la que aplica el RFE (puede también aplicarse la opción de *forward feature selection*). Los resultados se proporcionan en una tabla o en un vector con las bandas más importantes y un coeficiente, que es una medida de dicha importancia.

2.2.4.4. Aprendizaje automático: Bosques aleatorios

El método de los bosques aleatorios o *Random Forest* (Breiman, 2001) se utiliza comúnmente para la selección de características debido a su capacidad para medir la importancia de cada característica en la predicción de la variable objetivo. El proceso de selección se consigue mediante los pasos siguientes:

1. Entrenamiento del modelo: Se entrena un modelo *Random Forest* con todas las características disponibles.
2. La importancia de cada característica se evalúa en el modelo observando la reducción de la impureza (como Gini o entropía) o el decremento en el error cuando se utiliza esa característica en cada árbol.
3. Se calcula la importancia promedio en todos los árboles.
4. La selección de características relevantes se establece mediante la conservación de las características con una importancia alta y la eliminación de las que tienen poca o ninguna relevancia.

En este trabajo, el método se ha implementado con una aplicación de Matlab mediante la función *impOOB*. Esta función requiere previamente un entrenamiento de los datos con árboles aleatorios (*random forest*, RF, cuya función es *fitrensemble*). La salida es una tabla o vector con el coeficiente de importancia de cada característica y un gráfico que la muestra visualmente.

2.2.4.5. Basado en activaciones internas de la red 1D.

El método presentado tiene como objetivo identificar las bandas espectrales más relevantes para la clasificación de datos hiperespectrales mediante el análisis de las activaciones internas de una red neuronal convolucional unidimensional (CNN-1D).

Este enfoque proporciona una interpretación del modelo al resaltar las características más importantes aprendidas durante el entrenamiento.

1. El análisis comienza con el entrenamiento de una red CNN-1D utilizando datos hiperespectrales. Cada observación es una curva espectral representada como un vector de 300 valores, correspondiente a diferentes bandas espectrales. La red se entrena para clasificar estas curvas, aprendiendo características relevantes en sus capas internas.
2. Se selecciona la última capa convolucional de la red para el análisis. Esta capa captura características que reflejan la influencia de las bandas espectrales en las predicciones del modelo. Las activaciones de esta capa se utilizan como base para evaluar la importancia de las características.
3. Para cada observación, se pasa la curva espectral a través de la red y se extraen las activaciones de la capa seleccionada. Estas activaciones indican cómo la red responde a cada banda espectral de la entrada. Posteriormente, se calcula la importancia promedio de cada banda espectral al promediar las activaciones a través de los canales de salida de la capa.
4. Se acumulan las importancias de cada banda espectral a lo largo de todas las observaciones en el conjunto de datos. La acumulación global se promedia, obteniendo un vector que representa la importancia promedio de cada banda espectral en el modelo.
5. Con base en el mapa de importancia global, se establece un umbral (por ejemplo, el percentil 90) para identificar las bandas espectrales más relevantes. Estas bandas se consideran las características principales utilizadas por el modelo para tomar decisiones.
6. Finalmente, se genera un gráfico que muestra la importancia promedio de cada banda espectral. Este gráfico permite interpretar visualmente cuáles son las bandas más influyentes en las predicciones del modelo.

En este caso, lo primero que se ha de hacer es entrenar la red CNN-1D, en este caso con los datos correspondientes a las curvas hiperespectrales, según se verá en el apartado siguiente. Una vez entrenada y ajustada la red, se selecciona la característica más importante para cada categoría según el mapa de puntaje de importancia.

Como en los casos anteriores, se ha realizado el programa en Matlab. Como se ha explicado, será necesario la introducción de los datos etiquetados, además del modelo de la red CNN-1D.

2.2.5. Clasificación de las muestras

Para la clasificación de las imágenes completas de muestra de las aceitunas y las hojas según el tipo y el nivel de fitosanitario aplicado, se emplean en primer lugar diversos métodos de aprendizaje automático disponibles en el *Classification Learner* de Matlab; y a continuación se aplican las redes neuronales convolucionales 1D, que se implementan mediante programación, igualmente en Matlab.

2.2.5.1. Técnicas de aprendizaje automático

El aprendizaje automático (*Machine Learning*, ML) consiste en desarrollar un algoritmo de predicción para la resolución de un problema particular, mediante un proceso de aprendizaje a partir de los datos que permite encontrar patrones o tendencias en ellos (Torres, 2020). Hay infinidad de técnicas de aprendizaje automático y variantes de las mismas, que han sido aplicadas en diversos estudios relacionados con el sector agrícola en el que se parte de imágenes hiperespectrales adquiridas con diferentes sistemas, tanto en el exterior como en laboratorio.

Entre ellas, el método más utilizado son las máquinas de vectores de soporte (SVM), que han sido aplicados en inspección y clasificación de frutos y granos (Gao, et al., 2020; Zhang, et al., 2020c; Gao, et al., 2021; He, et al., 2021; Onmankhong, et al., 2022; Pang, et al., 2022; Wu, et al., 2022; Wei, et al., 2023); y de forma más específica para evaluar el estrés por pesticidas en frutos (Ye, et al., 2022) y hojas (Zhang, et al., 2022). Además, se ha utilizado muy ampliamente en otros estudios relacionados con la agricultura, como las clasificaciones de cultivos y usos del suelo a las que nos referiremos más adelante.

También profusamente utilizadas, aunque en menor medida, están otras técnicas muy conocidas de aprendizaje automático como los bosques aleatorios (RF), que se han aplicado en reconocimiento y clasificación de frutos y semillas (He, et al., 2021; Pang, et al., 2022) o para estudiar residuos de fitosanitarios y enfermedades de frutos y plantas (Ye, et al., 2022; Aggarwal, et al., 2023; Sawyer, et al., 2023; Qi, et al.,

2023); o los árboles de decisión, que también se emplean en reconocimiento (Hong, et al., 2023; Wei, et al., 2023). Estas técnicas y otras como los k-vecinos más próximos (k-NN) se emplean en estudios generales de clasificación de cultivos y usos del suelo y en otras aplicaciones que se recogen distintas revisiones (Lowe, et al., 2017; Lu, et al., 2020; Neupane & Baysal-Gurel, 2021; Wan, et al., 2021; Wan, et al., 2022; Moharram & Sundaram, 2023). Recientemente se ha publicado un trabajo por parte de los tutores de este TFM, entre otros, que aplican diversas técnicas de aprendizaje automático como las citadas hasta ahora a la detección de residuos de pesticidas en aceitunas (Martínez-Gila, et al., 2024).

Otro conjunto de técnicas que proporcionan buenos resultados son las basadas en la estadística multivariante clásica, aunque incorporadas a los algoritmos de aprendizaje automático, como el análisis discriminante, empleado para el reconocimiento de plantas y semillas (He, et al., 2021; Wu, et al., 2022; Hong, et al., 2023) o igualmente para la detección de pesticidas (Martínez-Gila, et al., 2024); la regresión logística en reconocimiento de granos (Zhang, et al., 2020c), detección de fitosanitarios (Aggarwal, et al., 2023) o enfermedades en las plantas (Ye, et al., 2022); o técnicas de agrupación no supervisada como las k-medias (Neupane & Baysal-Gurel, 2021).

Finamente, y ya en transición a las técnicas de aprendizaje profundo que se tratarán en el apartado siguiente, hay que citar las redes neuronales poco profundas de tipo perceptrón o perceptrón multicapa (MLP), que igualmente se han empleado ampliamente para reconocimiento y clasificación de granos y plantas (Li, et al., 2017; Hong, et al., 2023; Wei, et al., 2023) o detección de enfermedades en hojas (Aggarwal, et al., 2023), además de otras muchas aplicaciones que se indican en diversas revisiones (Lowe, et al., 2017; Golhani, et al., 2018; Ang & Seng, 2021 ; Neupane & Baysal-Gurel, 2021; Wan, et al., 2022).

En este trabajo, los métodos de clasificación de las muestras de aceitunas mediante aprendizaje automático se han llevado a cabo mediante la aplicación *Classification Learner* de Matlab, con la idea de comparar los resultados con los obtenidos mediante el aprendizaje profundo. Esta aplicación aparece en la ventana general del programa dentro de las herramientas o aplicaciones (Ilustración 2-21).

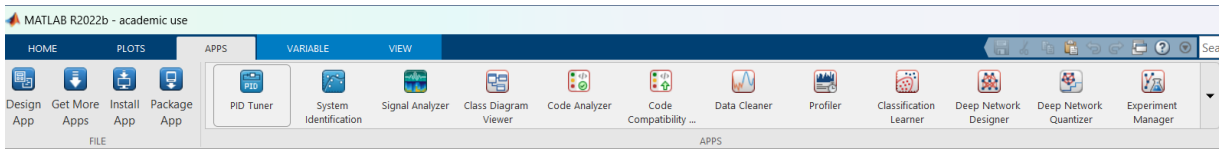


Ilustración 2-21. Classification Learner en el menú principal de Matlab.

Una vez abierto el *Classification Learner* (Ilustración 2-22), hay distintos aspectos que considerar. El principal son los distintos modelos a aplicar, que se muestran de forma más detallada en la Ilustración 2-23.

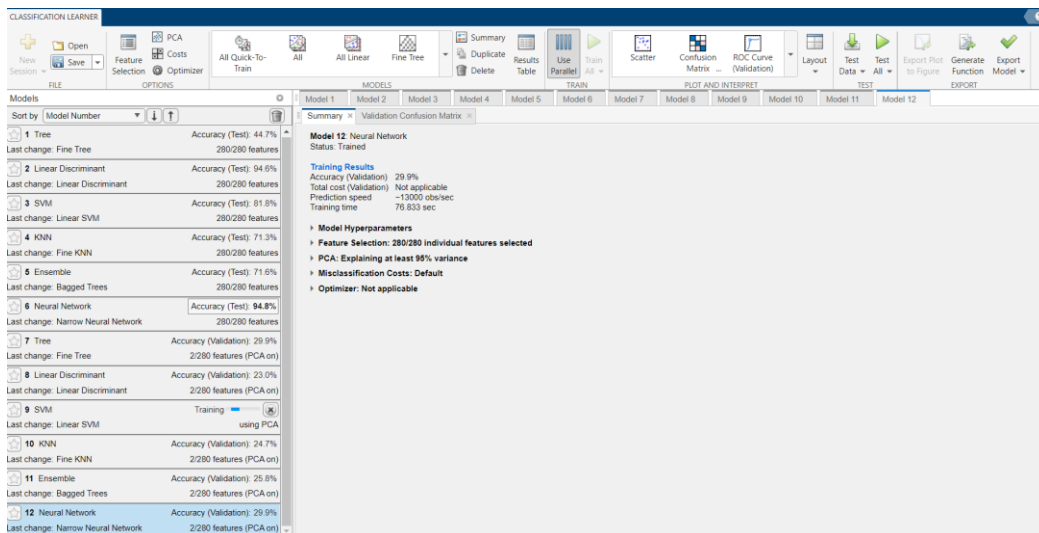


Ilustración 2-22. Ventana principal del Classification Learner de Matlab.

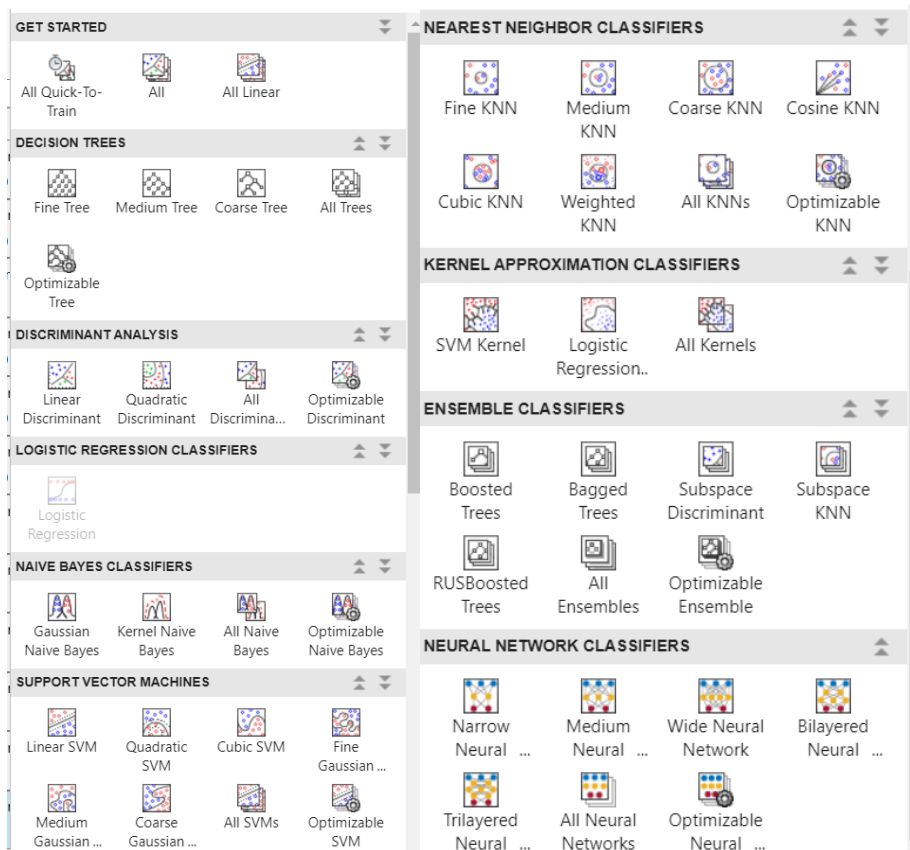


Ilustración 2-23. Clasificadores disponibles en la versión empleada de Matlab.

En primer lugar, se disponen los clasificadores de tipo árbol (de decisión) con varias modalidades; a continuación, se sitúa el análisis discriminante, la regresión logística y los clasificadores de probabilidad bayesiana, es decir, métodos que podrían incluirse en la estadística más o menos clásica. Tras ellos, se disponen los métodos más conocidos de aprendizaje automático, como el k-vecinos más próximos y máquinas de vectores de soporte. El siguiente grupo es el de clasificadores ensamblados o ensamblaje de clasificadores, por ejemplo, mediante técnicas de “bootstrap” o “bagging”. Bootstrap es una técnica estadística de muestreo con reemplazo, que es utilizada para estimar propiedades estadísticas (como la media, la varianza, etc.) de un conjunto de datos; mientras que el bagging (o agregación bootstrap) es una técnica de ensamblaje que utiliza bootstrap (reemplazo) para construir múltiples modelos y combina sus predicciones (Breiman, 2001); cuando se emplean árboles de decisión (bagged trees), conducen a la técnica conocida como bosques aleatorios o random forest. Finalmente, la aplicación implementa distintos tipos de redes neuronales, generalmente de tipo poco profundo.

A la vista de las técnicas disponibles, en este trabajo se van utilizar distintos tipos representativos de los distintos grupos mencionados. Concretamente, se han empleado:

- Análisis discriminante: Es una herramienta clásica del análisis multivariante (Johnson & Wichern, 2007; James, et al., 2017) que se basa en maximizar la separación entre grupos, representados por funciones multivariantes, lo que permite predecir a qué grupo pertenece un determinado caso en función de sus características. En este caso se va a aplicar el análisis discriminante lineal (LDA), que emplea funciones lineales para la clasificación.
- Árboles de decisión (DT): Son un modelo de predicción muy utilizado en diversos ámbitos (Breiman, et al., 1986; Hastie, et al., 2009). Dado un conjunto de datos se fabrican diagramas de construcciones lógicas que sirven para representar y categorizar una serie de condiciones que ocurren de forma sucesiva, en este caso para la clasificación de objetos. Los árboles están formados por nodos que se van dividiendo de acuerdo a criterios, como por ejemplo el coeficiente de diversidad de Gini. En este caso se van a emplear los árboles de ajuste fino (*fine trees*, FDT).
- Bosques aleatorios o *random forest* (RF): Son un método de aprendizaje automático utilizado principalmente para problemas de clasificación y regresión (Breiman, 2001; Hastie, et al., 2009). Se basan en la idea de crear un conjunto de árboles de decisión y combinar sus predicciones para mejorar la precisión y robustez del modelo. Cada árbol se entrena sobre una muestra aleatoria de los datos (*bootstrap*), lo que permite que cada árbol aprenda patrones diferentes de los datos.
- K-vecinos más próximos (k-NN): Es uno de los algoritmos de clasificación más conocidos dentro del aprendizaje supervisado (Hastie, et al., 2009). El algoritmo clasifica cada nuevo ejemplo calculando la distancia de ese ejemplo con todos los del conjunto de entrenamiento, siendo la clase predicha aquella a la que pertenezcan los ejemplos más cercanos.
- Máquinas de vectores de soporte (SVM) es igualmente un método de clasificación múltiple y regresión, desarrollado dentro de campo de la ciencia computacional (Vapnik, 1995; Hastie, et al., 2009). En este método se parte

de un conjunto de entrenamiento, que son etiquetadas en diferentes clases y representadas en puntos del espacio que separen las diferentes clases mediante un espacio lo más amplio posible, permitiendo así clasificar las nuevas muestras de test en función de su proximidad.

- Red neuronal superficial (tipo perceptrón multicapa, MLP). Las redes neuronales (ANN) se definen como como unas estructuras algorítmicas que permiten crear modelos formados por múltiples capas de procesamiento que realizan una serie de transformaciones lineales y no lineales sobre unos datos de entrada, para generar una salida próxima a la esperada (Bosch, et al., 2019; Torres, 2020). Un perceptrón, como unidad básica, funciona como un modelo de regresión lineal simple, que calcula los parámetros de pesos y sesgos (w y b), que va ajustando iterativamente a partir de los datos de entrada persiguiendo que el error se minimice.

Además de los propios clasificadores, otros aspectos a tener en cuenta para realizar el aprendizaje en el *Classification Learner* son:

- La posibilidad de obtener una muestra de test, además de las de entrenamiento y validación. En este caso, se ha aplicado esta partición, adoptando un porcentaje del 15% para test y 85% para entrenamiento/validación. Una vez se ha seleccionado esta posibilidad se debe aplicar el análisis de la muestra test.
- El tipo de validación, que puede ser validación de retención (*hold-out*) o cruzada, que es la que se eligen con *5-folds*. En cada validación se utiliza el 80% (del 85%) para entrenamiento y el 20% restante para la validación.
- Los parámetros a introducir en el modelo, que serán propios de cada uno de ellos. En este caso, se han utilizado los valores por defecto en cada uno de los modelos, puesto que no se pretende hacer un análisis de sensibilidad de los mismos, sino comparar sus resultados con los obtenidos con el aprendizaje automático.
- La posibilidad de aplicar análisis de componentes principales (PCA) o selección de características, mediante diversos métodos, según se ha mencionado anteriormente. En este caso, se han aplicado el PCA,

considerando el alcanzar el 99% de la varianza y una selección de los 30 factores de mayor importancia mediante el análisis de la varianza (ANOVA).

- Los resultados que se van a mostrar, siendo los más importantes la matriz de confusión, que amplía el dato de la exactitud (accuracy) para la validación, que se muestra directamente; se pueden mostrar además las métricas de la matriz, como la tasa de verdaderos positivos (TPR) o sensibilidad; la tasa de falsos negativos (FNR); el valor predictivo positivo (PPV), también conocido como precisión o recuperación de información (recall); y la tasa de falsos descubrimientos (FDR).
- Además, se pueden mostrar los valores del área bajo la curva ROC (AUC-ROC) para cada clase, que es un índice muy completo para valorar la exactitud y la precisión de las predicciones. En este caso, a pesar de ser un índice muy útil y completo, se descarta, ya que los valores son para cada tipo o nivel de producto, lo que complicaría mucho su descripción. Si se ha considerado la obtención de una muestra de test, se pueden obtener la exactitud de la misma, la matriz de confusión y sus parámetros y el valor de AUC-ROC. Finalmente se puede obtener una tabla resumen de los resultados, que puede ser exportada como archivo de texto.

2.2.5.2. Técnicas de aprendizaje profundo

El término aprendizaje profundo (DL) se reserva para aquellos métodos de aprendizaje automático, fundamentalmente redes neuronales, que están compuestas de múltiples capas de procesamiento, con múltiples niveles de abstracción. A partir de los datos de partida y a través de las múltiples capas neuronales se llevan a cabo una serie de transformaciones lineales y no lineales para generar una salida que se aproxima iterativamente a la esperada (etiqueta o *label*) (Torres, 2020).

Este tipo de técnicas se han empleado en numerosos trabajos en el ámbito de la agricultura y la industria asociada, que parten de la adquisición de imágenes hiperespectrales a diversas escalas de trabajo, como se ha mencionado en la introducción. Los tipos de redes empleadas son muy diversas, dependiendo de la aplicación, destacando las redes convolucionales (CNN), tanto 1D como 2D e incluso 3D, además de redes recurrentes (RNN) y otras.

Las redes 1D se han desarrollado para analizar fenómenos que varían en una dimensión como pueden ser las señales de telecomunicaciones, series temporales o como en este caso el espectro asociado a un pixel, a un grupo de pixeles o a una región dentro de una imagen, como ya se ha descrito en el apartado anterior (Shuai, et al., 2024). En este último sentido, se han utilizado sobre todo cuando se pretende reconocer o clasificar muestras de plantas, u órganos o partes de las mismas (frutos, granos, semillas, hojas, tallos, etc.), por lo que son las utilizadas de forma principal en este trabajo.

Se han descrito en trabajos relacionados con el reconocimiento y clasificación de variedades de granos, frutos y tallos (Zhu, et al., 2019; Zhang, et al., 2020c; He, et al., 2021; Dong, et al., 2022; Onmankhong, et al., 2022; Zhu, et al., 2023); o para estudiar su rendimiento, viabilidad o estado de maduración (Li, et al., 2017; Su, et al., 2021; Yang, et al., 2021; Pang, et al., 2022; Wu, et al., 2022; Hong, et al., 2023). Más específicamente relacionados con los objetivos de este trabajo se han empleado para analizar el estrés por herbicidas y pesticidas en granos y frutos (Chu, et al., 2022; Ye, et al., 2022; Zhang, et al., 2022). Finalmente se han aplicado para la clasificación de cultivos y usos del suelo (Mou, et al., 2017; Haut, et al., 2019; Cao & Guo, 2020; Meng, et al., 2021; Patel, et al., 2023), aunque en este último caso, son las redes 3D las más empleadas, como ahora se indicará.

Por su parte, las redes CNN 2D, debido a su potencia de análisis y en el reconocimiento de objetos y patrones en las imágenes, también se han aplicado en estudios relacionados con el reconocimiento y clasificación de granos y frutos (Zhu, et al., 2019; Gao, et al., 2020; Chen, et al., 2022; Dong, et al., 2022) y más específicamente para la detección de estrés por fitosanitarios (Chu, et al., 2022; Ye, et al., 2022) y enfermedades (Fazari, et al., 2018; Nguyen, et al., 2021; Qi, et al., 2023). Aunque las aplicaciones más numerosas de este tipo de redes hay que buscarlos en la clasificación de cultivos y usos del suelo (segmentación) en otras escalas de trabajo, así como en otras aplicaciones recogidas en distintas revisiones (Golhani, et al., 2018; Wan, et al., 2022; Wang, et al., 2022b). En este trabajo se han utilizado sobre todo para la segmentación de la imagen y consiguiente selección o recorte de las áreas de interés, como se ha indicado en el apartado correspondiente.

Las redes 3D ofrecen la ventaja de trabajar con datos de 2D (por ejemplo, la imagen de las aceitunas y las hojas de este trabajo) más una tercera dimensión, como pueden ser las curvas espectrales de las imágenes hiperespectrales. Así, se han empleado para clasificar granos, fenotipado o control de calidad (Su, et al., 2021; Karila, et al., 2022 ; Nguyen, et al., 2023; Qi, et al., 2023; Zhu, et al., 2023; Chen, et al., 2022; Gao, et al., 2021). Más relacionados con los objetivos de este trabajo, se han aplicado a la detección de enfermedades en plantas (Nguyen, et al., 2021; Shi, et al., 2022; Cui, et al., 2022; Sawyer, et al., 2023). Tienen el inconveniente de que requieren una capacidad de procesamiento que excede habitualmente de las que ofrecen los ordenadores personales, por lo que, dados los objetivos de este trabajo, no se han aplicado específicamente, ya que quedan cubiertos por otros tipos de redes.

Finalmente hay que referirse a la aplicación de otros tipos de redes como las redes convolucionales denominadas profundas (DCNN), principalmente en clasificación de cultivos y usos del suelo, aunque también en problemas como el reconocimiento de semillas (Wu, et al., 2022) o la presencia de organismos patógenos en las plantas (Agarwal, et al., 2020). Las redes recurrentes también se han aplicado al reconocimiento de semillas y granos (Zhang, et al., 2020c; Pang, et al., 2022), además de en clasificación de usos del suelo. Por su parte, en trabajos de revisión se hace referencia a redes generativas (GAN) y aprendizaje por transferencia (TL) en este ámbito de trabajo (Wang, et al., 2021; Moharram & Sundaram, 2023; Wang, et al., 2022b; Guerri, et al., 2023).

La metodología desarrollada consiste generalmente en obtener el espectro medio de los objetos estudiados, sus partes o una muestra de píxeles, y analizar el comportamiento del mismo con vistas a la clasificación de los objetos, en este caso de las muestras de aceitunas y hojas a los que se han aplicado distintos productos fitosanitarios en diferentes concentraciones. Para el desarrollo de la red se ha utilizado la función *trainNetwork* que permite el control de la red mediante diversos parámetros. Asimismo, la red se ha empleado para la selección de características como se ha indicado en el apartado anterior.

El tipo de red implementada para todos los casos es del mismo tipo, una red convolucional 1-D, aunque se modifican algunos parámetros de esta como el número

de capas, el *dropout* o los hiperparámetros de entrenamiento para los distintos casos, obteniendo una buena relación entre la complejidad del modelo y el sobreajuste. Una mayor complejidad en el modelo permite que la red alcance una exactitud mayor en los datos de entrenamiento; pero, al mismo tiempo, puede dar lugar a un sobreajuste a estos datos, proporcionando unas tasas de exactitud para validación y test considerablemente inferiores. Se considera que una red está sobreajustada cuando la exactitud de los datos de entrenamiento es superior en aproximadamente un 10% respecto de la exactitud en los datos de validación y test.

Finalmente, tras realizar distintas pruebas, modificando el número de capas convolucionales, número y tamaño de filtro y capas de *dropout*, se obtiene la siguiente red. La red es una red convolucional 1-D, con tres capas convolucionales. A la red se le introducen los espectros medios correspondientes y a partir de ahí clasificará entre los tipos de productos fitosanitarios y las distintas concentraciones. En la figura Ilustración 2-24 se muestra el esquema de la red utilizada.

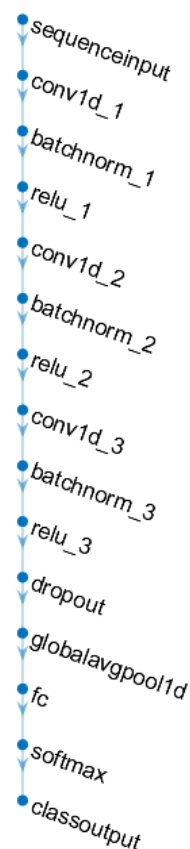


Ilustración 2-24. Esquema de las capas de la red convolucional (CNN) 1D implementada.

A continuación, se detallan cada una de las capas utilizadas, comenzando por la capa de entrada de secuencias, *sequenceInputLayer*. Esta capa define que la red acepta secuencias; en este caso se introducen secuencias de 280 características, que corresponden con las 280 bandas que se analizan. Es la puerta de entrada de los datos hacia las capas posteriores.

La red incluye además varias capas convolucionales (*convolution1dLayer*) para detectar patrones en las secuencias. Cada capa tiene filtros de tamaño 5, lo que significa que analiza ventanas de 5 elementos consecutivos en la secuencia. La primera capa extrae 64 mapas de características, mientras que las dos siguientes extraen 256 cada una, aumentando la capacidad de la red para detectar patrones más complejos. Todas las capas tienen la opción de "*padding*", '*same*', lo que garantiza que la salida tenga el mismo tamaño que la entrada, preservando la longitud de la secuencia.

Después de cada capa convolucional, hay una capa de normalización por lotes (*batchNormalizationLayer*) y una capa de activación ReLU (*reluLayer*). La normalización estabiliza los valores de las características y acelera el entrenamiento al reducir el riesgo de desvanecimiento o explosión del gradiente. La activación ReLU introduce no linealidad, permitiendo que la red modele relaciones complejas en los datos.

Para reducir el riesgo de sobreajuste, se incluye una capa de abandono (*dropoutLayer*), que desactiva aleatoriamente el porcentaje indicado de las neuronas durante cada iteración de entrenamiento. Esto fuerza a la red a ser más robusta y a no depender excesivamente de conexiones específicas. Este parámetro se modificará para evitar el sobreajuste en cada una de las redes entrenadas, pero su valor se encuentra entre el 30 y 60%.

La capa de *globalAveragePooling1dLayer* condensa todas las características generadas por las capas convolucionales en un solo valor promedio por canal. Esto reduce significativamente la dimensionalidad y permite que la red pase a la clasificación con un vector compacto que resume toda la secuencia. La capa completamente conectada (*fullyConnectedLayer*) reduce las características

resumidas al número de categorías. Posteriormente, la capa softmax (*softmaxLayer*) convierte estos valores en probabilidades, asegurando que su suma sea igual a 1. Finalmente, la capa de clasificación (*classificationLayer*) calcula la pérdida durante el entrenamiento y ajusta los pesos de la red en función de las etiquetas verdaderas.

Una vez creada la red básica, se procede al entrenamiento de las diferentes redes para la determinación de la exactitud en los datos de entrenamiento y validación. Al igual que en el caso anterior, estos se toman de manera completamente aleatoria y en una proporción 70/15/15. La red termina proporcionando los resultados de exactitud para el entrenamiento, la validación y el test. Adicionalmente muestra una gráfica con la evolución de la exactitud en entrenamiento y validación (Ilustración 2-25), junto a otros datos (tiempo, ciclos, épocas, etc.), así como las correspondientes matrices de confusión en test, entrenamiento y validación (Ilustración 2-26).

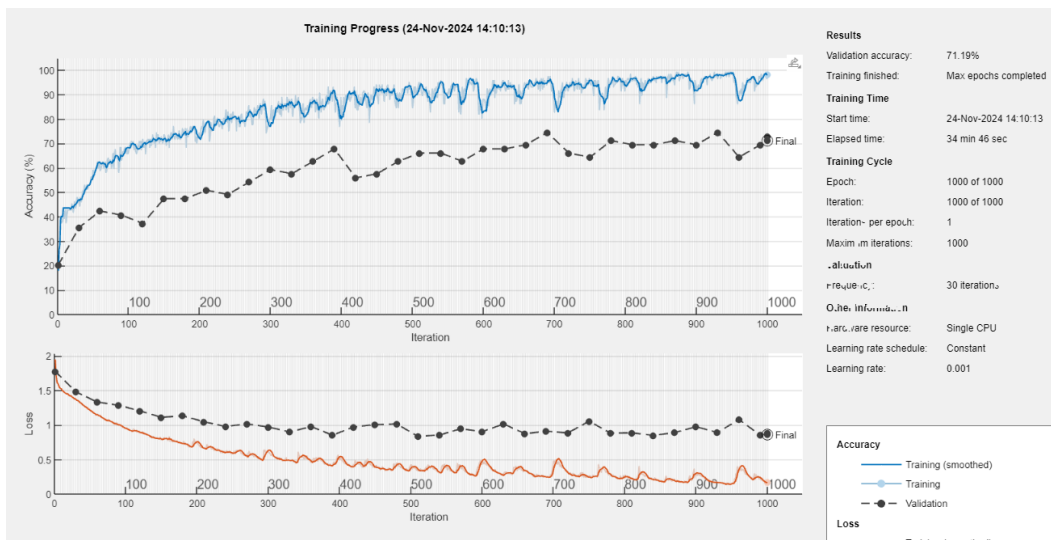


Ilustración 2-25. Curvas de exactitud de entrenamiento y validación.

	Del	Difu	Lam	Oxi	Sin_tratar	Teb
Del	7			1		2
Difu		8	4			1
Lam			14			
Oxi			1	5		
Sin_tratar		1			2	
Teb			1			11
	Del	Difu	Lam	Oxi	Sin_tratar	Teb

Ilustración 2-26. Matriz de confusión de los datos de tipo test tras la aplicación de la red convolucional.

En algunos casos de los analizados, los resultados que se obtienen con la red no son satisfactorios como se describirá en el apartado correspondiente (entrenamiento con una exactitud de hasta el 80% con las muestras extraídas de las ventanas definidas manualmente). Por ello, durante el entrenamiento se probaron diferentes opciones para las capas, filtros e hiperparámetros de la red. En primer lugar, se obtuvo una red muy sobreentrenada, por lo que se redujo el *dropout*, para permitir a la red generalizar mejor. Debido a este ajuste fue necesario cambiar algunos hiperparámetros de entrenamiento como la tasa de aprendizaje, que se aumentó. Se probó también a complicar la red, lo que llevaba a un mejor porcentaje de entrenamiento, pero menor de validación; y el caso contrario, con el que se consigue menor sobreajuste, pero menor exactitud. Sin embargo, pese a las múltiples pruebas realizadas manualmente, comprendiendo el funcionamiento de la red, no se pudo hallar una red con una mejor tasa de acierto que la mostrada anteriormente.

Por ello, se decide realizar otro entrenamiento utilizando un método de optimización bayesiana. La optimización bayesiana se utiliza para ajustar los hiperparámetros clave de la red neuronal. Este método construye un modelo probabilístico que predice cómo los hiperparámetros influyen en el rendimiento del modelo, basándose en evaluaciones previas. Se definen los siguientes hiperparámetros para optimizar:

- Número de capas convolucionales (1 a 6 capas).
- Número de filtros en cada capa convolucional (32 a 256).
- Tamaño de los filtros (3 a 15).
- Tasa de abandono (*DropoutRate*) para regularización (0 a 0.5).
- Tasa de aprendizaje inicial (*InitialLearnRate*) con transformación logarítmica.
- Regularización L2 para controlar el sobreajuste, también con transformación logarítmica.

La función objetivo evalúa diferentes configuraciones de estos hiperparámetros entrenando y validando redes con configuraciones específicas, devolviendo la precisión en el conjunto de validación como métrica de desempeño.

El programa ejecuta la optimización bayesiana para un número limitado de evaluaciones (10 en este caso). Durante cada evaluación, la optimización ajusta los hiperparámetros según las configuraciones anteriores, explorando de manera inteligente los valores que probablemente mejoren el desempeño. Esto se logra mediante la función de adquisición, que busca un equilibrio entre explorar nuevas configuraciones y explotar aquellas que ya han mostrado buenos resultados.

Con los mejores hiperparámetros seleccionados por la optimización bayesiana, se construye la arquitectura final de la red neuronal. El número de capas convolucionales, el tamaño y cantidad de filtros, y la tasa de abandono se configuran según los valores óptimos encontrados. Posteriormente, los datos de entrenamiento y validación se combinan para maximizar la información disponible durante el entrenamiento final. Este modelo se entrena utilizando el optimizador *Adam*, configurado con la mejor tasa de aprendizaje inicial y regularización L2 obtenidas durante la optimización.

Se obtiene con este proceso la red que se muestra a continuación (Ilustración 2-27), similar a la que se tenía inicialmente, pero añadiéndole una capa convolucional extra.



Ilustración 2-27. Red obtenida tras la optimización.

En la Ilustración 2-28 se muestran las iteraciones realizadas para la obtención de la red óptima. En cada iteración, la red calcula el error en la predicción de los datos de validación, y se toma la red con un menor error. En este caso, se toma la red obtenida en la sexta iteración.

Iter	Eval result	Objective	Objective runtime	BestSoFar (observed)	BestSoFar (estim.)	NumConvLayers	NumFilters	FilterSize	DropoutRate	InitialLearnRate	L2Regularization
1	Best	0.5814	148.24	0.5814	0.5814	3	244	3	0.098313	0.018899	0.00012462
2	Accept	0.63953	130.76	0.5814	0.58463	4	179	12	0.050631	0.024348	0.0029265
3	Accept	0.65116	84.528	0.5814	0.5814	2	235	8	0.37097	0.0048614	0.0074186
4	Accept	0.61628	44.895	0.5814	0.5814	2	96	11	0.33184	0.0084396	1.9582e-05
5	Accept	0.59302	40.427	0.5814	0.58141	2	72	13	0.18671	0.0016636	0.00023735
6	Best	0.33721	127.91	0.33721	0.33723	4	174	15	0.35792	0.00038407	8.4235e-05
7	Accept	0.33721	176.9	0.33721	0.33721	4	254	12	0.10661	0.00012011	1.1134e-05
8	Accept	0.52326	73.836	0.33721	0.33722	3	147	9	0.44247	0.0001821	0.00098961
9	Accept	0.67442	42.389	0.33721	0.33721	4	34	5	0.14885	0.00016291	0.0047979
10	Accept	0.33721	182.95	0.33721	0.33711	4	236	15	0.223	0.00010204	1.1407e-05

Ilustración 2-28. Iteraciones realizadas para obtener la red óptima.

Los hiperparámetros de la nueva red son los que se muestran a continuación (Ilustración 2-29). Se observa que el tamaño del filtro se aumenta considerablemente respecto a la red anteriormente utilizada.

Best observed feasible point:

NumConvLayers	NumFilters	FilterSize	DropoutRate	InitialLearnRate	L2Regularization
4	174	15	0.35792	0.00038407	8.4235e-05

Ilustración 2-29. Hiperparámetros de la red optimizada.

Adicionalmente se obtienen gráficos como los que se muestran en la Ilustración 2-30.

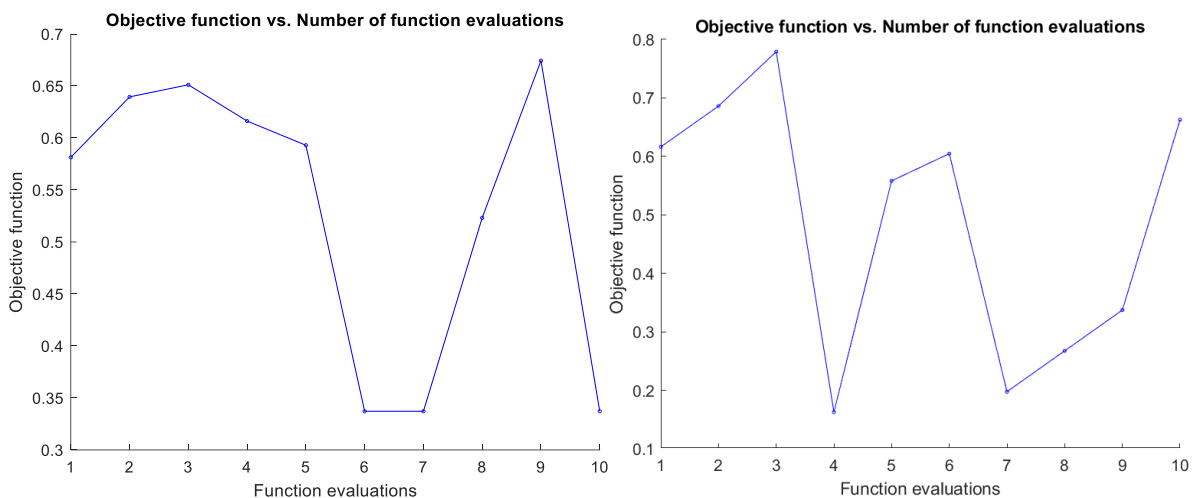


Ilustración 2-30. Gráficos obtenidos con información sobre la optimización.

Se realiza una prueba más debido a que en el caso del tebuconazol y lambda, no se llega a alcanzar el 90% de exactitud como en el resto de casos, por lo que se hace un análisis bayesiano más exhaustivo, en el que, en lugar de evaluar 10 redes, se evaluarán 30. Adicionalmente, el tiempo de evaluación de cada red se aumenta, aumentando el número de épocas, permitiendo que las redes lleguen a estabilizarse y así, se evalúa el funcionamiento real de la arquitectura. Se obtiene así la siguiente red (Ilustración 2-31):

```

Optimization completed.
MaxObjectiveEvaluations of 30 reached.
Total function evaluations: 30
Total elapsed time: 31592.7403 seconds
Total objective function evaluation time: 31585.1473

Best observed feasible point:
  NumConvLayers  NumFilters  FilterSize  DropoutRate  InitialLearnRate  L2Regularization
  -----
           4           245           13           0.1259           0.00036019           0.0004473

Observed objective function value = 0.13953
Estimated objective function value = 0.16407
Function evaluation time = 2211.9874

```

Ilustración 2-31. Características de la red optimizada.

Sin embargo, tras entrenarla no se obtienen los resultados deseables para la exactitud de los diferentes grupos de datos. Se observa que, pese a que el error de clasificación es algo menor, al entrenar la red y evaluar sobre los datos no mejora en exceso, y en test incluso disminuye su exactitud. Por tanto, se determina que con 10 iteraciones es suficiente para encontrar una configuración óptima de la red, reduciendo el tiempo de procesamiento considerablemente.

Finalmente, tras realizar el ajuste manual y, en la mayoría de casos la optimización bayesiana, se obtienen las siguientes redes neuronales e hiperparámetros de entrenamiento para aceitunas (Tabla 2-8) y hojas (Tabla 2-9).

Fitosanitario	Nº capas conv.	Nº Filtros	Tamaño del filtro	Dropout	Tasa de aprendizaje	L2 Regularization
Deltametrina	3	64	5	0.5	0.001	0.005
Diflufenican	3	64	5	0.5	0.001	0.005
Lambda-cihalotrina	4	174	15	0.35792	0.00038407	$8.4235 \cdot 10^{-5}$
Oxifluorfen	3	64	5	0.5	0.001	0.005
Tebuconazol	3	65	15	0.2256	0.0035656	0.002008
Tipo	4	243	13	0.49031	0.0012356	$2.2834 \cdot 10^{-5}$

Tabla 2-8. Arquitectura de la red e hiperparámetros de entrenamiento para las distintas redes de clasificación de aceitunas.

Fitosanitario	Nº capas conv.	Nº Filtros	Tamaño del filtro	Dropout	Tasa de aprendizaje	L2 Regularization
Deltametrina	6	192	15	0.3499	0.00038411	0.0081179
Diflufenican	6	130	15	0.20514	0.0002277	0.00025227
Lambda-cihalotrina	5	256	7	0.49255	0.0002125	0.00016527
Oxifluorfen	5	85	9	0.3516	0.0011525	0.0043032
Tebuconazol	6	32	14	0.014525	0.0001281	0.0003571
Tipo	6	220	11	0.3644	0.0001597	0.0034523

Tabla 2-9 Arquitectura de la red e hiperparámetros de entrenamiento para las distintas redes de clasificación de hojas.

2.2.6. Segmentación y clasificación de las imágenes

La segmentación semántica de imágenes es una operación muy empleada, ya que permite no solo identificar objetos o clases y sus características, si no localizarlos dentro de la imagen. El resultado es que cada uno de los píxeles de la imagen quedará asignado a una clase, es decir que la segmentación (semántica) da lugar a una clasificación de los píxeles de la imagen. Este tipo de operación es muy utilizado dentro del sector agrícola dentro de la escala de cultivos (clasificación de cultivos y usos del suelo) e incluso de plantas, y mucho menos a la escala de órganos.

Así, distintas técnicas de aprendizaje automático se han aplicado para la clasificación de cultivos y usos del suelo (cartografía) como el SVM (He, et al., 2018; Zhong, et al., 2018; Zhu, et al., 2018; Xu, et al., 2018; Haut, et al., 2019; Meng, et al., 2021; Wan, et al., 2021), los bosques aleatorios, árboles de decisión y k-vecinos más próximos (Meng, et al., 2021; Patel, et al., 2023; Haut, et al., 2019).

Por su parte, el aprendizaje profundo mediante redes convolucionales, especialmente 2D, ofrece significativas mejoras en este ámbito, ya que estas redes tienen en cuenta no solo el valor de los píxeles sino su entorno, por lo que muchos rasgos texturales son identificados por ellas. Junto con su extensión a redes CNN 3D, que incorporan la componente hiperespectral de cada píxel, y las profundas (DCNN) se está empleando de forma creciente en numerosos trabajos (Li, et al., 2017; Zhong, et al., 2017; He, et al., 2018; Mou, et al., 2017; Xu, et al., 2018; Zhong, et al., 2018; Zhu, et al., 2018; Haut, et al., 2019; Cao & Guo, 2020; Meng, et al., 2021) (Ladi, et al., 2022; Moharram & Sundaram, 2023). De igual forma las redes recurrentes (RNN) también se aplican en este tipo de estudios (Mou, et al., 2017; Moharram & Sundaram, 2023).

La segmentación para estudios a escala de órganos es mucho menos empleada por la dificultad que sugiere la detección de características a estas escalas, aunque se aplican en algunos ámbitos como la identificación de defectos en granos (Chen, et al., 2022) o la detección de fitosanitarios en frutos (Han & Gao, 2019). También se ha empleado para la separación de aceitunas, residuos (hojas) y suciedad en lotes de aceitunas en almazaras (Fernández Ruiz, 2022).

En este trabajo se ha empleado la segmentación semántica, como se ha indicado en el preprocesado, concretamente en la definición y recorte del área de interés en las aceitunas, es decir, la zona de las aceitunas que están libres de defectos y aisladas de su entorno.

Adicionalmente, se ha desarrollado una aplicación para la visualización de los resultados basada igualmente en la segmentación semántica. Como se demostrará en la siguiente sección, tras el análisis de los resultados de la clasificación de las muestras por el espectro medio de las imágenes completas y el de las ventanas o cuadrículas que permiten el aumento de datos, se determina que los resultados son más precisos con estos últimos, tanto si se definen manualmente como si se hace de forma automática. Por otra parte, las aceitunas reales pueden haberse tratado con distintos productos y mostrar distintas concentraciones, lo que justifica este proceso de segmentación que permita una evaluación más fina de los residuos de pesticidas.

Debido a ello, se desarrolló una aplicación para mostrar visualmente el proceso de clasificación cuando se realiza con las cuadrículas definidas automáticamente. Mediante la aplicación se podrá seleccionar una carpeta, donde se encuentran las imágenes en formato “.bil”, que serán leídas y tratadas. Con carácter previo, se aplicará la definición de las áreas de interés en el caso de las aceitunas mediante una segmentación semántica con la red convolucional previamente entrenada; y, para las hojas, se utiliza la segmentación por umbral previamente implementada. Una vez se separa el fondo de la parte de la imagen a tratar, se crean las cuadrículas de forma automática, calculándose el espectro medio de cada uno de ellas. A continuación, serán clasificadas en primer lugar por el tipo de producto con la red entrenada para este propósito y, posteriormente, se determina su concentración con la red correspondiente. Con esta aplicación el usuario podrá visualizar el pesticida con el que se ha tratado la muestra, qué zonas tienen una mayor concentración o menor, si esta es homogénea, etc.

Además de la clasificación, la interfaz permite la estimación del porcentaje de defectos en la aceituna. Se incluye esta utilidad, ya que, al hacer la segmentación semántica de la aceituna, como se explicó anteriormente, se obtuvo con gran exactitud

la clasificación de los píxeles correspondientes a esta clase. En la Ilustración 2-32 se muestran dos ejemplos de la aplicación de esta operación.

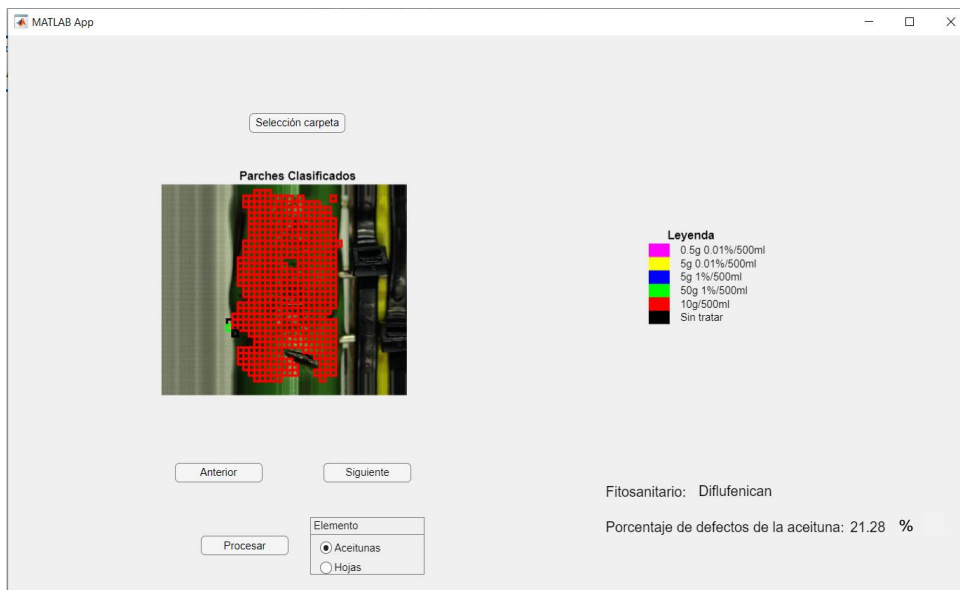
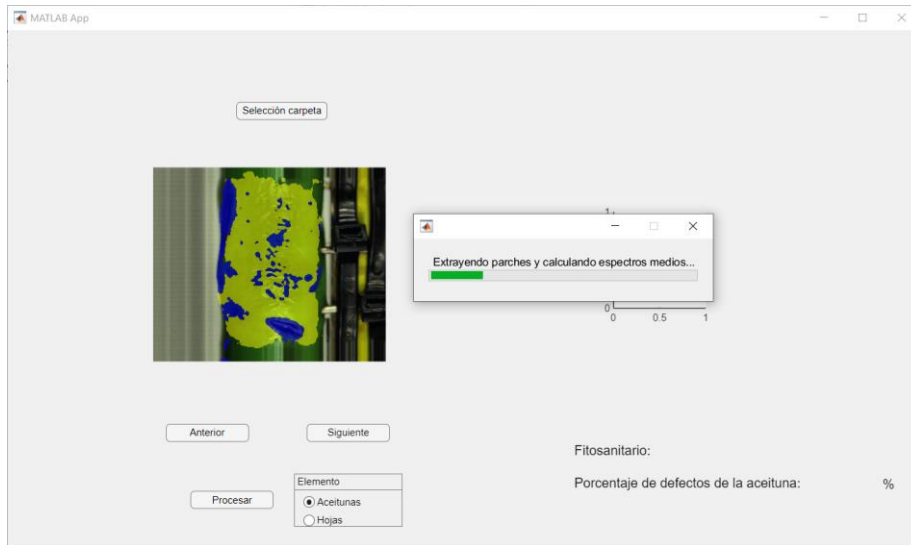


Ilustración 2-32 Interfaz de visualización por ventanas de la concentración de pesticida fitosanitarios de la imagen y porcentaje de defectos de la aceituna.

3. RESULTADOS Y DISCUSIÓN

Este apartado de resultados se estructura en varios subapartados, comenzando por la obtención de las curvas espectrales medias de las aceitunas, por tipos y nivel de concentración de productos fitosanitarios. A continuación, se presentarán los resultados de procesos de clasificación basados en el aprendizaje automático y profundo (red convolucional) en el caso de las aceitunas y únicamente en la red en el caso de las hojas. Finalmente, se presentan los resultados de la selección de características, tanto para las aceitunas como para las hojas.

3.1. Curvas espectrales de las aceitunas

De inicio se muestran las curvas espectrales medias de las aceitunas, para fundamentar todos los análisis posteriores. Concretamente, las que se muestran aquí son las medias de los, a su vez, espectros medios de las ventanas extraídas manualmente (Ilustración 3-1 a Ilustración 3-6).

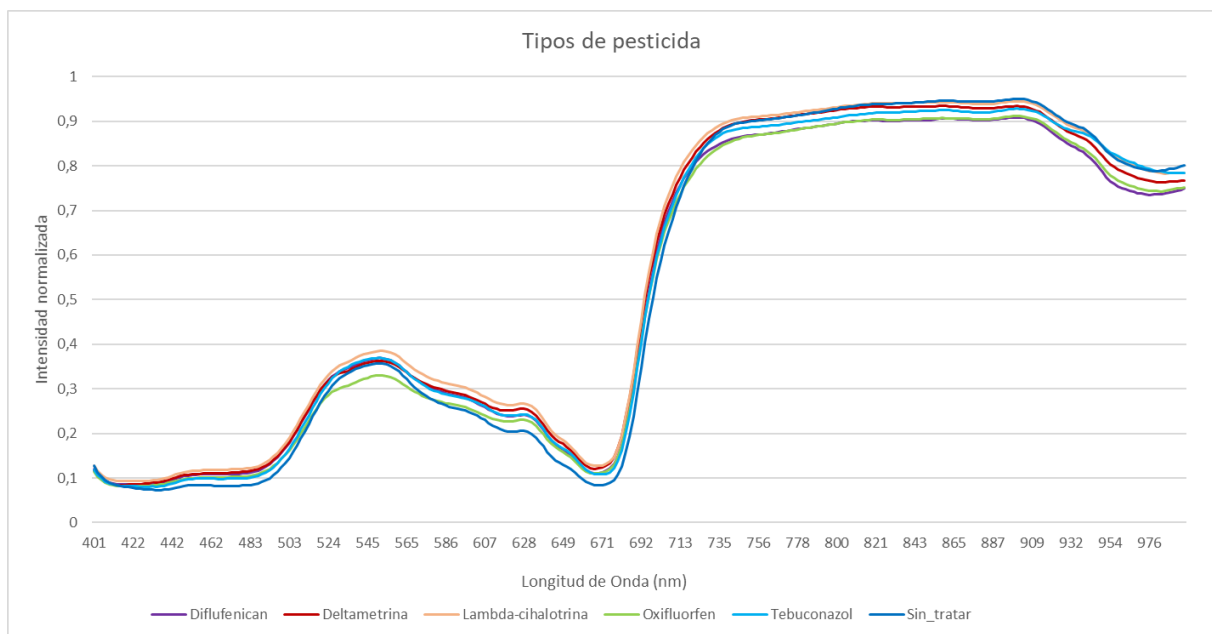


Ilustración 3-1. Curvas espectrales medias de las aceitunas tratadas con los distintos tipos de producto.

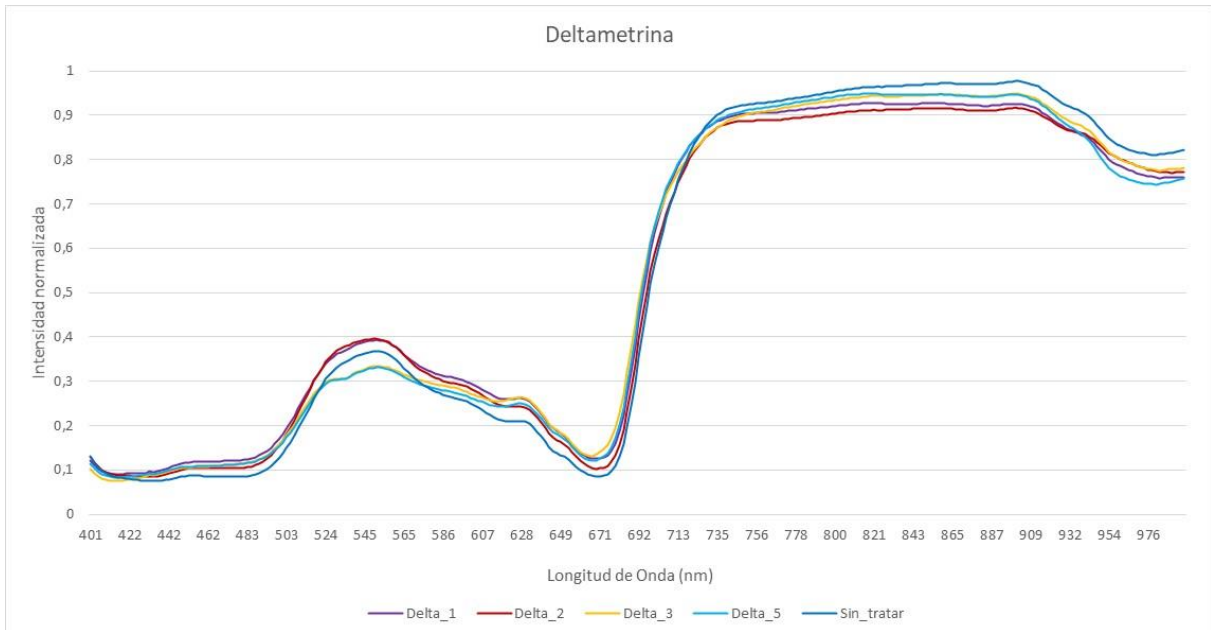


Ilustración 3-2. Curvas espectrales medias de las aceitunas tratadas con deltametrina.

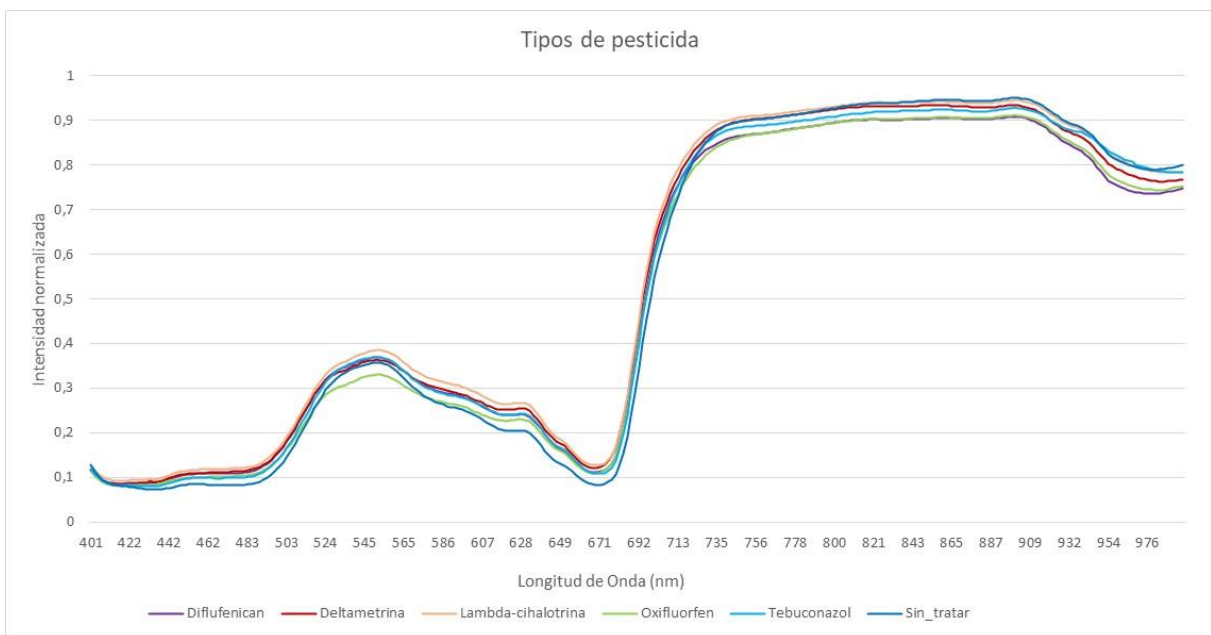


Ilustración 3-3. Curvas espectrales medias de las aceitunas tratadas con diflufenican.

En todos los casos se observa una curva típica de algún elemento de color verde relacionado con la vegetación. En la región visible del espectro (aprox. 400-700 nm) se observa el pico relativo precisamente en la región del verde (547 nm) y a continuación la meseta de altos valores de reflectancia del infrarrojo cercano (a partir de 780 nm aproximadamente), relacionada con la estructura de las paredes vegetales. Desde ahí hay un descenso hacia el infrarrojo medio (1500 nm aproximadamente), donde se producen mínimos relacionados con la absorción de agua (Jensen, 2009).

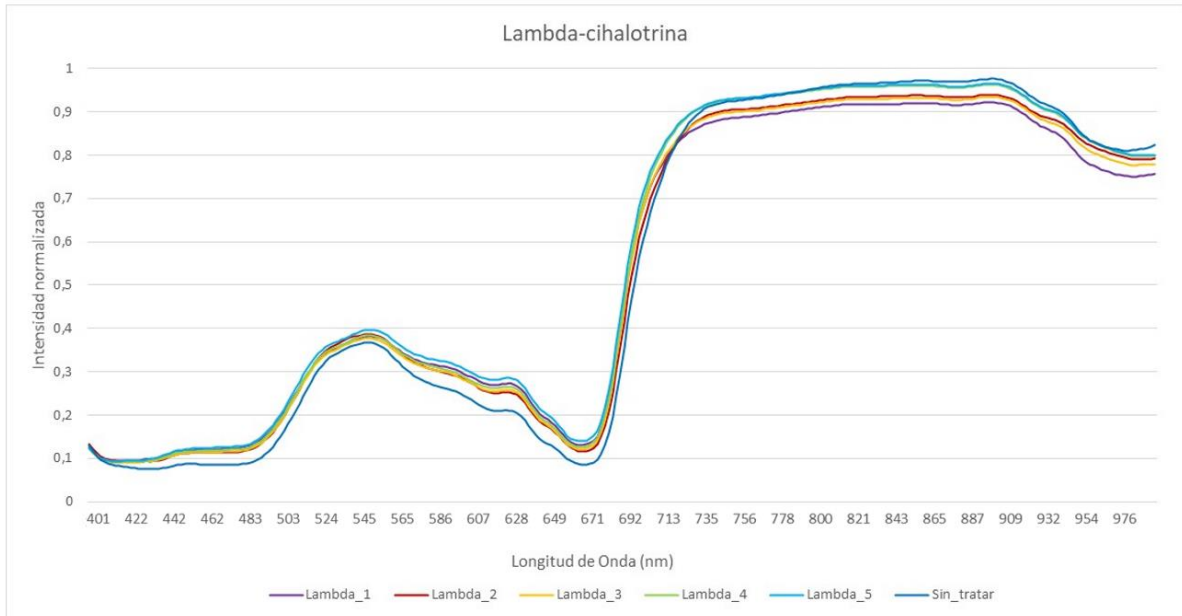


Ilustración 3-4. Curvas espectrales medias de las aceitunas tratadas con lambda-cihalotrina.

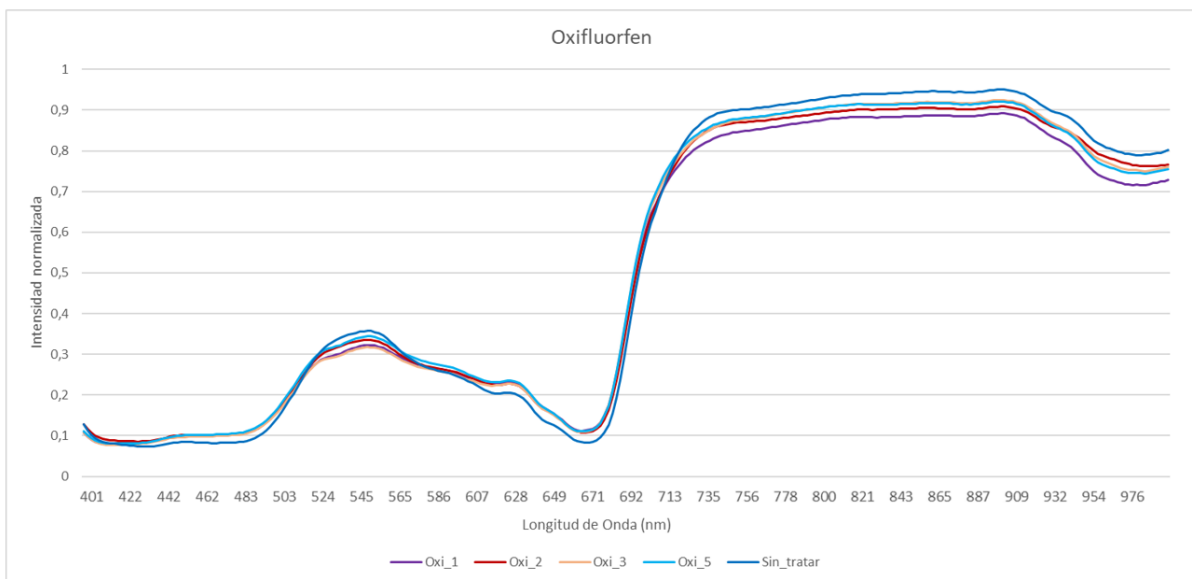


Ilustración 3-5. Curvas espectrales medias de las aceitunas tratadas con oxifluorfen.

Los mínimos que se observan asociados a la región del azul (459 nm) y rojo (643 nm), junto al máximo relativo ya comentado del verde, se deben a la absorción de los pigmentos de la vegetación, especialmente de la clorofila, por lo que cambios en el estado de la vegetación se traducirán en cambios en esta región y en consecuencia en la coloración de la misma (Jensen, 2009). Otra región interesante y muy sensible al estado de la vegetación es la rampa que se observa entre la región del rojo y la del infrarrojo cercano, conocido como borde o límite del rojo (red edge).

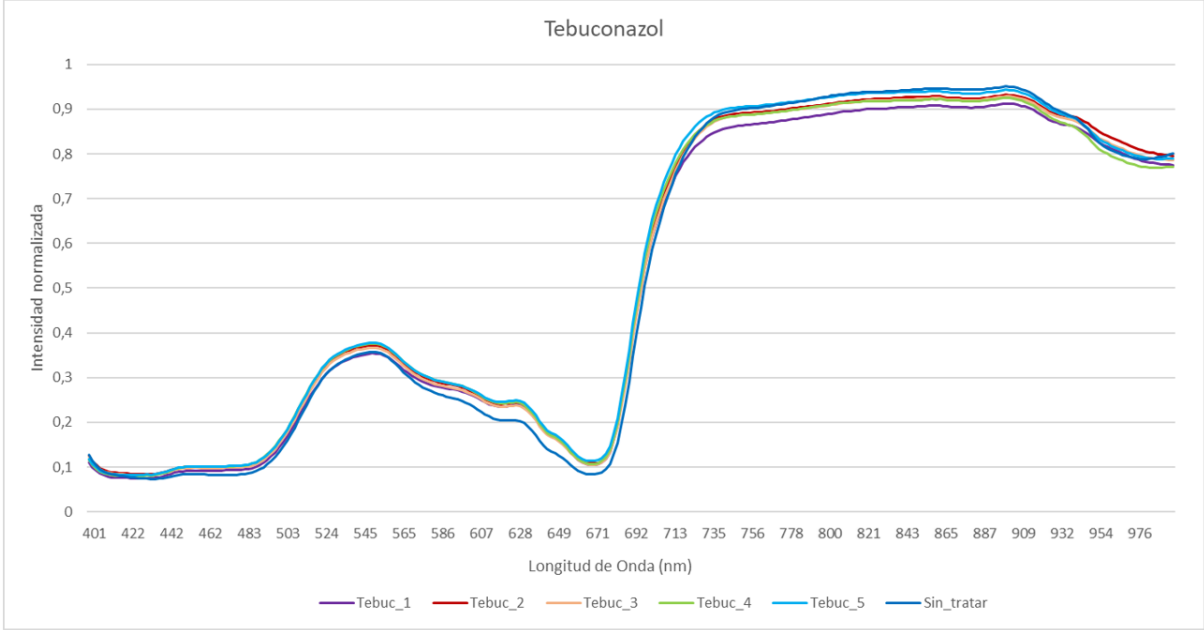


Ilustración 3-6. Curvas espectrales medias de las aceitunas tratadas con tebuconazol.

3.2. Clasificación de las muestras de aceitunas

En primer lugar, se presentan los resultados tras la aplicación de los métodos de aprendizaje automático para la clasificación de las muestras de imágenes de aceitunas, donde el análisis es más exhaustivo; y a continuación, los resultados de la clasificación en las hojas.

3.2.1. Técnicas de aprendizaje automático

En este apartado se analizan en primer lugar los resultados obtenidos a partir de las curvas espectrales medias de las aceitunas; en segundo lugar, los correspondientes a las curvas medias de las ventanas definidas manualmente; y en tercer lugar, las curvas medias de las cuadrículas definidas automáticamente.

En todos los tipos de muestras, se comenzará describiendo los resultados de analizar la diferenciación entre los tipos de pesticida y posteriormente se procederá al análisis por niveles de concentración en cada tipo de producto: deltametrina, diflufenican, lambda-cihalotrina, oxifluorfen y tebuconazol.

En cada uno de los casos, se presenta una tabla con los valores de exactitud (entrenamiento, validación y test), para cada método considerado, aplicando el análisis de componentes principales y la selección de factores. Además, se presentan las matrices de confusión para los métodos que mejores y peores resultados presenta, siempre sin aplicar el ACP y la selección de características.

3.2.1.1. Curvas espectrales medias

La primera clasificación se realizó a partir de las curvas espectrales medias de las aceitunas obteniendo los resultados que se muestran en las siguientes tablas (exactitud en validación y test) y figuras (matrices de confusión).

En la Tabla 3-1 se muestran los resultados del **análisis por tipos de pesticida**, siendo el método del análisis discriminante y la red neuronal los que obtienen mejores resultados, en torno al 80%. El resto de métodos presentan menores valores de la exactitud, siendo los árboles de decisión los que presenta los resultados más bajos de hasta 50% en test.

El análisis de componentes principales no mejora los resultados, sino que en realidad los empeora en la mayor parte de los casos, salvo en el k-NN y los bosques aleatorios, donde mejora ligeramente, a pesar de alcanzar el 99% de la varianza explicada (6 CP). En cuanto a la selección de factores, la exactitud en general disminuye en torno a un 10-15%, incluso manteniendo un número importante de factores (30). Los tiempos de ejecución no superan en general los 10 (20) segundos.

Las matrices de confusión (Ilustración 3-7), tomando como referencia el análisis discriminante, muestra que el producto que se distingue mejor es el diflufenican, seguido por las muestras de tebuconazol y sin tratar, y la que más se confunde es la deltametrina.

Modelo	Validac.	Test	Tiempo	Caract.	Técnica	ACP
Árboles Decisión	57,79	48,57	5,91	280/280		
An. Discriminante	81,43	79,90	10,61	280/280		
SVM	52,26	51,43	8,06	280/280		
KNN	73,87	65,71	4,64	280/280		
Bagged / Bosques Al.	63,32	68,57	18,02	280/280		
Red Neuronal	83,92	74,29	4,09	280/280		
Árboles Decisión	56,78	62,86	6,51	6 CP		99% Var.
An. Discriminante	45,73	40,00	7,27	6 CP		99% Var.
SVM	48,74	42,86	8,68	6 CP		99% Var.
KNN	83,42	80,00	5,49	6 CP		99% Var.
Bagged / Bosques Al.	74,87	60,00	13,59	6 CP		99% Var.
Red Neuronal	74,87	68,57	3,63	6 CP		99% Var.
Árboles Decisión	42,71	60,00	10,33	30/280	ANOVA	
An. Discriminante	67,34	68,57	9,20	30/280	ANOVA	
SVM	43,22	51,43	9,63	30/280	ANOVA	
KNN	51,76	54,29	15,53	30/280	ANOVA	
Bagged / Bosques Al.	49,25	54,29	12,59	30/280	ANOVA	
Red Neuronal	67,34	71,43	14,57	30/280	ANOVA	

Tabla 3-1. Resultados de la exactitud (accuracy) en validación y test del aprendizaje automático por tipos de producto en los espectros medios de las aceitunas.

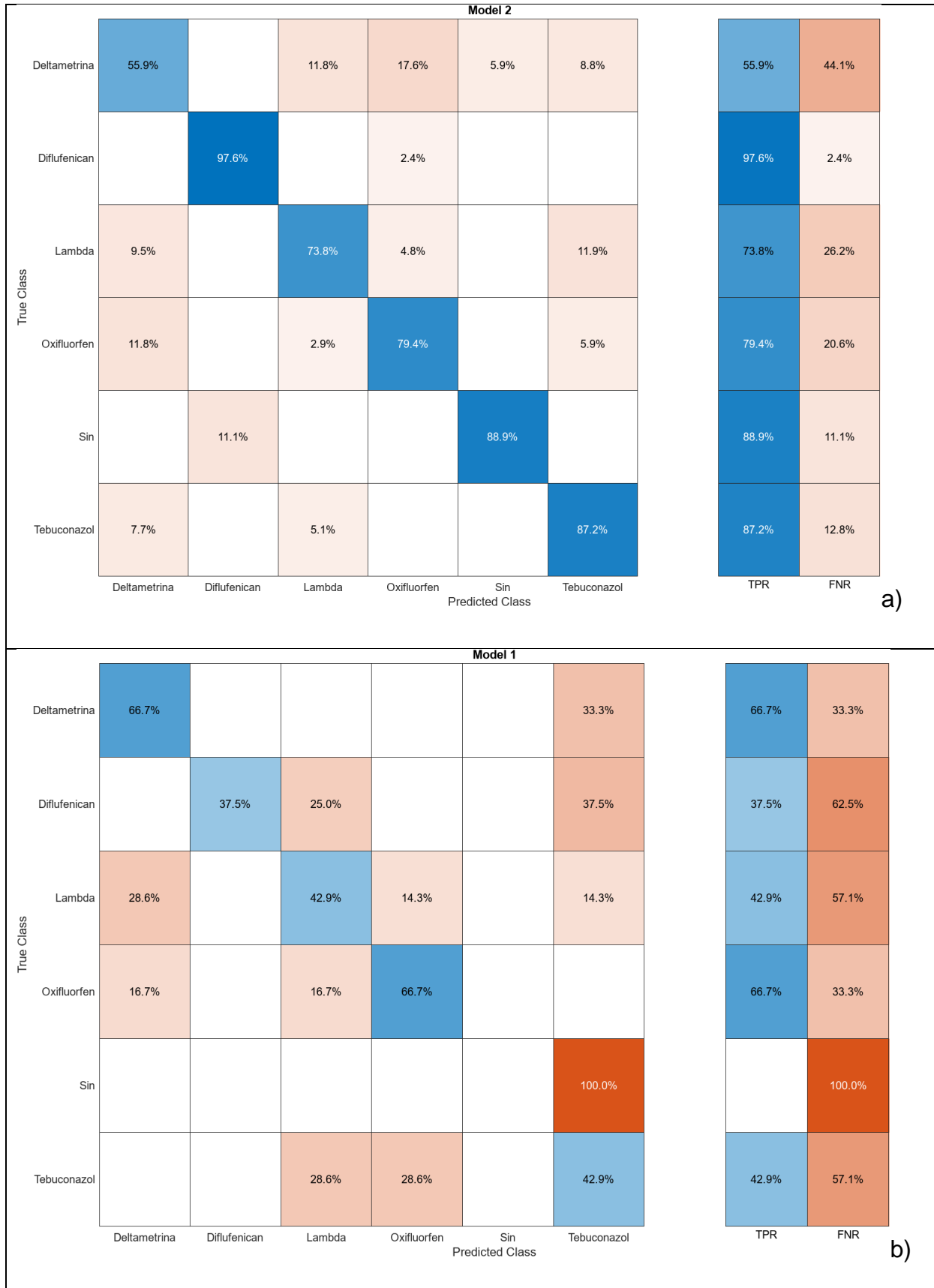


Ilustración 3-7. Matrices de confusión de los datos de test en el aprendizaje automático por tipos de producto en los espectros medios de las aceitunas: a: Análisis discriminante; b: Árbol de decisión.

En cuanto al análisis por niveles de producto aplicado, comenzando por la **deltametrina** (Tabla 3-2), se observa que éste en general muestra valores moderados o bajos en la mayor parte de los casos (no suelen superar el 60%), siendo SVM el que presenta el valor más elevado en test (80%) y los bosques aleatorios el que presentan el valor más bajo (40%). Los tiempos de ejecución suelen ser inferiores a 10 segundos.

El análisis de componentes principales muestra en este caso resultados similares en términos medios para los distintos métodos, respecto al análisis de todas las bandas, reteniéndose 4 componentes. La selección de factores, aun considerando 30 factores, presenta valores bastante mediocres, en torno a 10 puntos inferiores al análisis con todos factores.

Las matrices de confusión (Tabla 3-2 Ilustración 3-8) muestran que la mayor confusión se produce entre los niveles 2 y 3 (moderados-altos), ambos por encima del límite europeo, por lo que no supone ningún problema; mientras que las muestras sin tratar o los niveles bajos o altos de producto se diferencian mejor.

Modelo	Validac.	Test	Tiempo	Caract.	Técnica	ACP
Árboles Decisión	55,56	40,00	6,63	280/280		
An. Discriminante	60,00	60,00	8,03	280/280		
SVM	55,56	80,00	9,30	280/280		
KNN	62,22	60,00	5,61	280/280		
Bagged / Bosques Al.	53,33	40,00	13,12	280/280		
Red Neuronal	64,44	40,00	9,70	280/280		
Árboles Decisión	42,22	40,00	7,01	4 CP		99% Var.
An. Discriminante	62,22	80,00	7,66	4 CP		99% Var.
SVM	60,00	80,00	7,50	4 CP		99% Var.
KNN	48,89	60,00	8,41	4 CP		99% Var.
Bagged / Bosques Al.	48,89	40,00	13,72	4 CP		99% Var.
Red Neuronal	55,56	60,00	5,03	4 CP		99% Var.
Árboles Decisión	33,33	40,00	9,04	30/280	ANOVA	
An. Discriminante	64,44	60,00	8,09	30/280	ANOVA	
SVM	55,56	60,00	14,14	30/280	ANOVA	
KNN	40,00	40,00	11,51	30/280	ANOVA	
Bagged / Bosques Al.	46,67	40,00	12,88	30/280	ANOVA	
Red Neuronal	53,33	40,00	16,52	30/280	ANOVA	

Tabla 3-2. Resultados de la exactitud (accuracy) en validación y test del aprendizaje automático, para los niveles de deltametrina en los espectros medios de las aceitunas.9

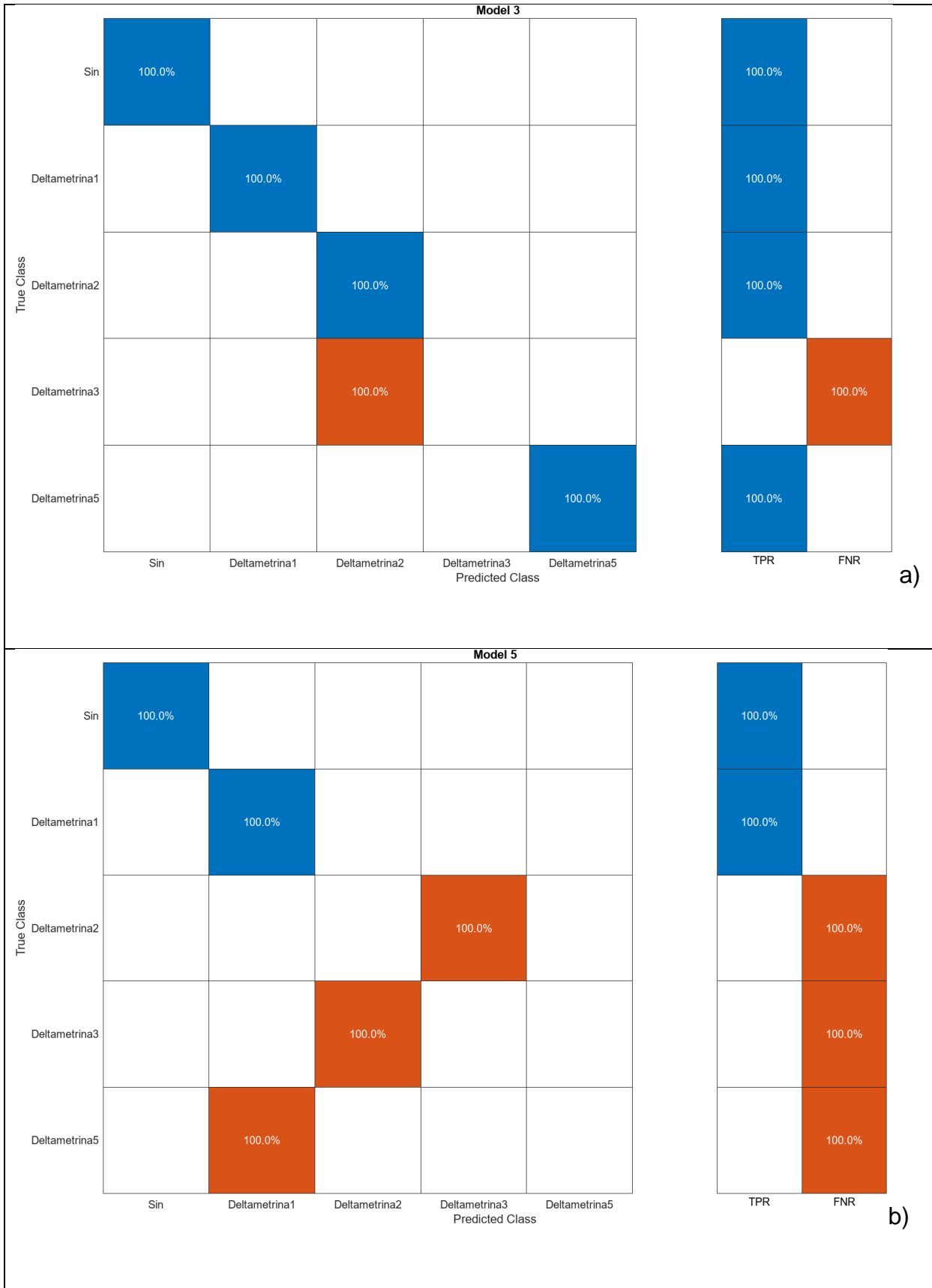


Ilustración 3-8. Matrices de confusión de los datos de test en el aprendizaje automático para los niveles de deltametrina en los espectros medios de las aceitunas: a: SVM; b: Bosques aleatorios.

Los valores de exactitud para los niveles de **diflufenican** (Tabla 3-3) son en general bastante elevados, en especial para las muestras de test, donde llegan a veces al 100%. No obstante, destacan sobre otros métodos el análisis discriminante, SVM, kNN y la red neuronal. Los tiempos de ejecución suelen ser inferiores a 15 segundos.

Al introducir el análisis de componentes principales, en este caso no se produce una reducción excesiva al operar con las 5 componentes que se mantienen al 99% de la varianza. Sin embargo, en la selección de factores sí se produce una cierta disminución de la exactitud, especialmente en el test en algunos de los métodos.

Las matrices de confusión (Ilustración 3-9) no resultan muy ilustrativas, por los pocos casos existentes en la muestra de test, aunque en general puede deducirse una buena discriminación del nivel de este producto en las distintas metodologías consideradas.

Modelo	Validac.	Test	Tiempo	Caract.	Técnica	ACP
Árboles Decisión	82,69	66,67	7,22	280/280		
An. Discriminante	86,54	100,00	13,71	280/280		
SVM	84,62	100,00	10,02	280/280		
KNN	75,00	100,00	5,54	280/280		
Bagged / Bosques Al.	76,92	83,33	14,78	280/280		
Red Neuronal	78,85	100,00	4,21	280/280		
Árboles Decisión	76,92	83,33	7,54	5 CP		99% Var.
An. Discriminante	90,38	100,00	12,11	5 CP		99% Var.
SVM	75,00	100,00	13,17	5 CP		99% Var.
KNN	82,69	100,00	12,04	5 CP		99% Var.
Bagged / Bosques Al.	71,15	83,33	13,97	5 CP		99% Var.
Red Neuronal	84,62	100,00	14,68	5 CP		99% Var.
Árboles Decisión	46,15	33,33	11,22	30/280	ANOVA	
An. Discriminante	75,00	66,67	13,09	30/280	ANOVA	
SVM	59,62	33,33	18,69	30/280	ANOVA	
KNN	69,23	83,33	11,17	30/280	ANOVA	
Bagged / Bosques Al.	55,77	50,00	22,09	30/280	ANOVA	
Red Neuronal	67,31	66,67	11,40	30/280	ANOVA	

Tabla 3-3. Resultados de la exactitud (accuracy) en validación y test del aprendizaje automático, para los niveles de diflufenican en los espectros medios de las aceitunas.

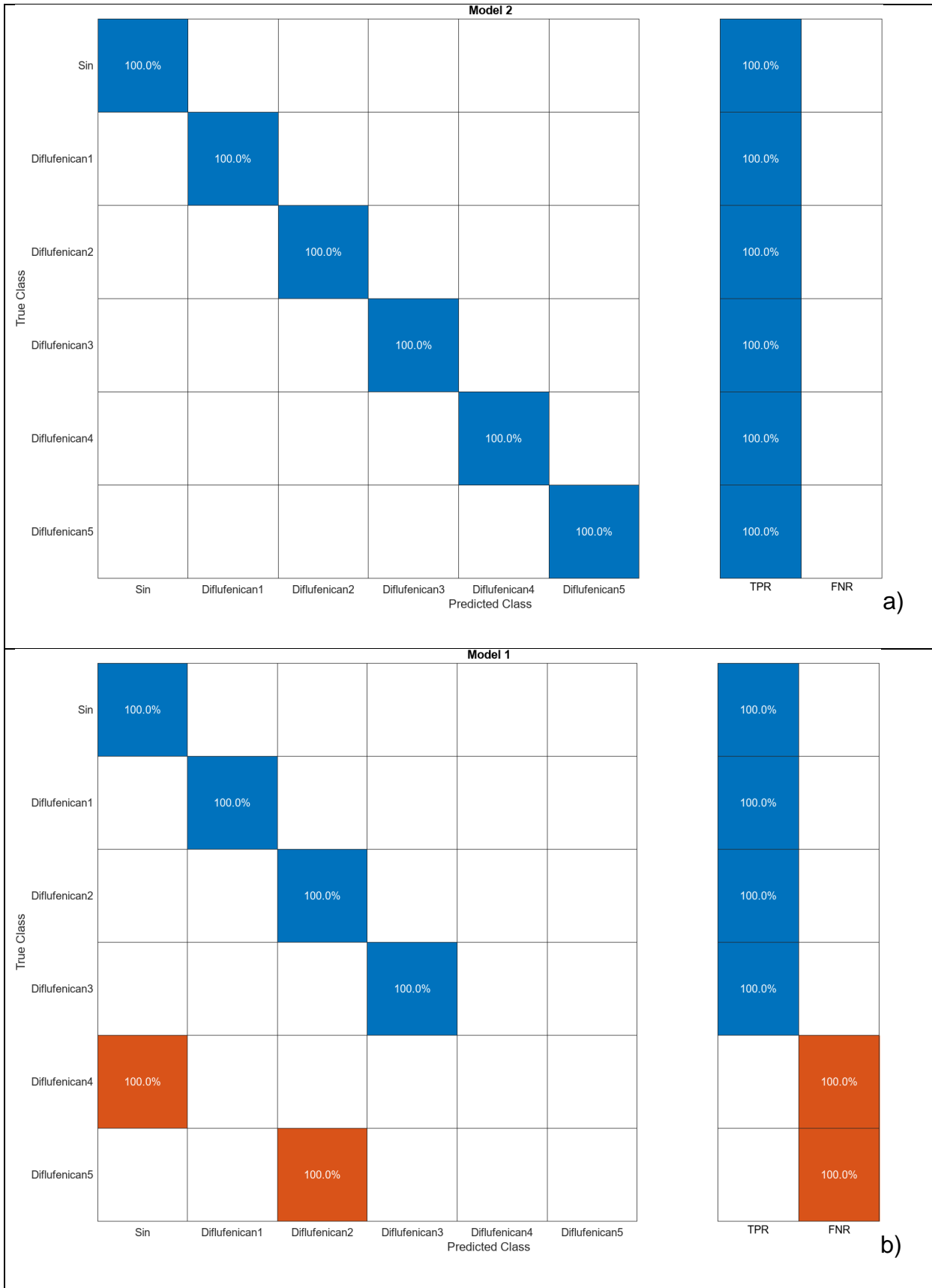


Ilustración 3-9. Matrices de confusión de los datos de test en el aprendizaje automático para los niveles de diflufenican en los espectros medios de las aceitunas: a) Análisis discriminante; b) Árboles de decisión.

El análisis del nivel de **lambda-cihalotrina** (Tabla 3-4) proporciona resultados bastante aceptables para la mayor parte de los métodos (en torno al 80%), destacando el análisis discriminante y el SVM, mientras que k-NN y los árboles de decisión muestran valores más moderados. Los tiempos de ejecución se encuentran en la mayor parte de los casos entre 10 y 15 segundos.

Por su parte, el análisis con las componentes principales, 5 CP al 99% de la varianza, supone en general una reducción de la exactitud de hasta 20 puntos en términos medios. Algo menor es la reducción que se produce con la selección de características (30), que en algunos métodos como el análisis discriminante o la red neuronal es prácticamente nula, lo que permitiría aplicar la técnica en estos casos.

Las matrices de confusión (Ilustración 3-10), aun siendo poco significativas muestran la buena diferenciación entre los niveles de este producto estando la mayor confusión entre los niveles medios a bajos del mismo (4) y las muestras sin tratar, es decir, siempre bajo el límite europeo.

Modelo	Validac.	Test	Tiempo	Caract.	Técnica	ACP
Árboles Decisión	68,52	66,67	6,75	280/280		
An. Discriminante	87,04	83,33	12,10	280/280		
SVM	87,04	83,33	9,34	280/280		
KNN	75,93	50,00	8,90	280/280		
Bagged / Bosques Al.	75,93	83,33	13,03	280/280		
Red Neuronal	74,07	83,33	4,40	280/280		
Árboles Decisión	59,26	50,00	7,16	280/280		99% Var.
An. Discriminante	61,11	83,33	12,82	280/280		99% Var.
SVM	55,56	66,67	13,10	280/280		99% Var.
KNN	51,85	66,67	13,10	280/280		99% Var.
Bagged / Bosques Al.	51,85	66,67	13,95	280/280		99% Var.
Red Neuronal	51,85	83,33	15,00	280/280		99% Var.
Árboles Decisión	72,22	66,67	10,79	30/280	ANOVA	
An. Discriminante	87,04	100,00	14,67	30/280	ANOVA	
SVM	85,19	83,33	17,14	30/280	ANOVA	
KNN	68,52	83,33	8,87	30/280	ANOVA	
Bagged / Bosques Al.	66,67	66,67	20,14	30/280	ANOVA	
Red Neuronal	72,22	66,67	10,96	30/280	ANOVA	

Tabla 3-4. Resultados de la exactitud (accuracy) en validación y test del aprendizaje automático, para los niveles de lambda-cihalotrina en los espectros medios de las aceitunas.

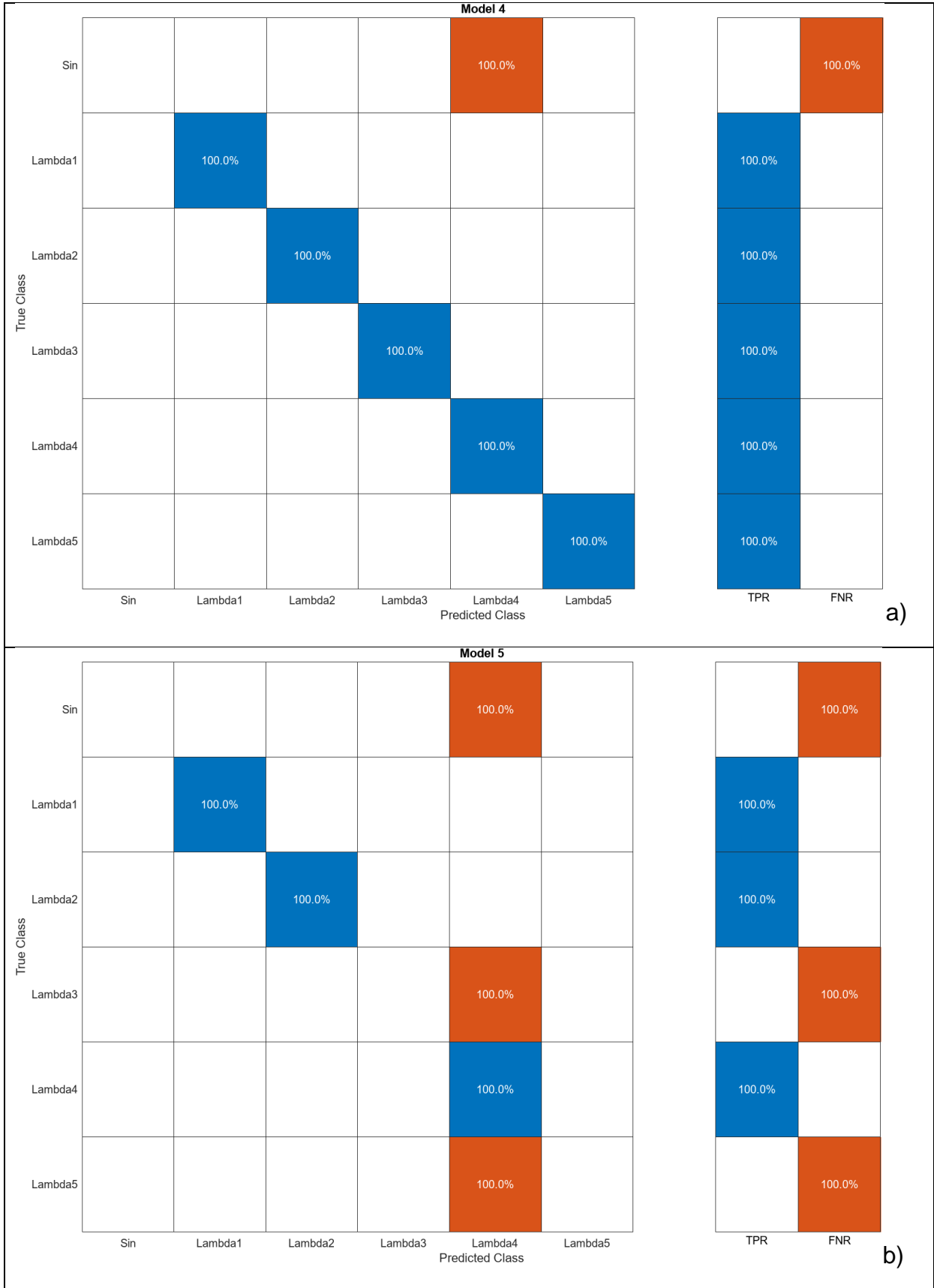


Ilustración 3-10. Matrices de confusión de los datos de test en el aprendizaje automático para los niveles de lambda-cihalotrina en los espectros medios de las aceitunas: a) SVM; b) k-NN.

El siguiente producto analizado es el **oxifluorfen** (Tabla 3-5), que muestra valores moderados en la validación, más altos en algunos clasificadores como el análisis discriminante, que en otros casos como el k-N; y algo más altos en general en el test (en algunos casos llegando al 100%). Los tiempos de ejecución son inferiores a 20 segundos en la mayor parte de los casos.

En el análisis de componentes principales, en el que se mantienen 6 componentes al 99% de la varianza, los valores de exactitud son bastante similares en la mayor parte de los casos, a diferencia de la selección de características (30), donde se produce una reducción bastante significativa en alguno de los métodos.

Las matrices de confusión (Ilustración 3-11) son muy poco significativas por el tamaño reducido de la muestra, por lo que no permiten realizar observaciones de interés, salvo la buena diferenciación en general entre los niveles.

Modelo	Validac.	Test	Tiempo	Caract.	Técnica	ACP
Árboles Decisión	65,91	100,00	7,87	280/280		
An. Discriminante	84,09	100,00	13,76	280/280		
SVM	72,73	80,00	13,55	280/280		
KNN	61,36	80,00	13,26	280/280		
Bagged / Bosques Al.	79,55	100,00	13,03	280/280		
Red Neuronal	72,73	60,00	11,73	280/280		
Árboles Decisión	61,36	60,00	11,38	6 CP		99% Var.
An. Discriminante	81,82	100,00	14,26	6 CP		99% Var.
SVM	75,00	80,00	18,87	6 CP		99% Var.
KNN	65,91	40,00	22,19	6 CP		99% Var.
Bagged / Bosques Al.	75,00	80,00	27,71	6 CP		99% Var.
Red Neuronal	72,73	80,00	27,21	6 CP		99% Var.
Árboles Decisión	38,64	80,00	14,56	30/280	ANOVA	
An. Discriminante	61,36	100,00	8,35	30/280	ANOVA	
SVM	71,00	80,00	17,07	30/280	ANOVA	
KNN	60,45	80,00	8,74	30/280	ANOVA	
Bagged / Bosques Al.	69,55	80,00	20,15	30/280	ANOVA	
Red Neuronal	77,27	80,00	10,90	30/280	ANOVA	

Tabla 3-5. Resultados de la exactitud (accuracy) en validación y test del aprendizaje automático, para los niveles de oxifluorfen en los espectros medios de las aceitunas.

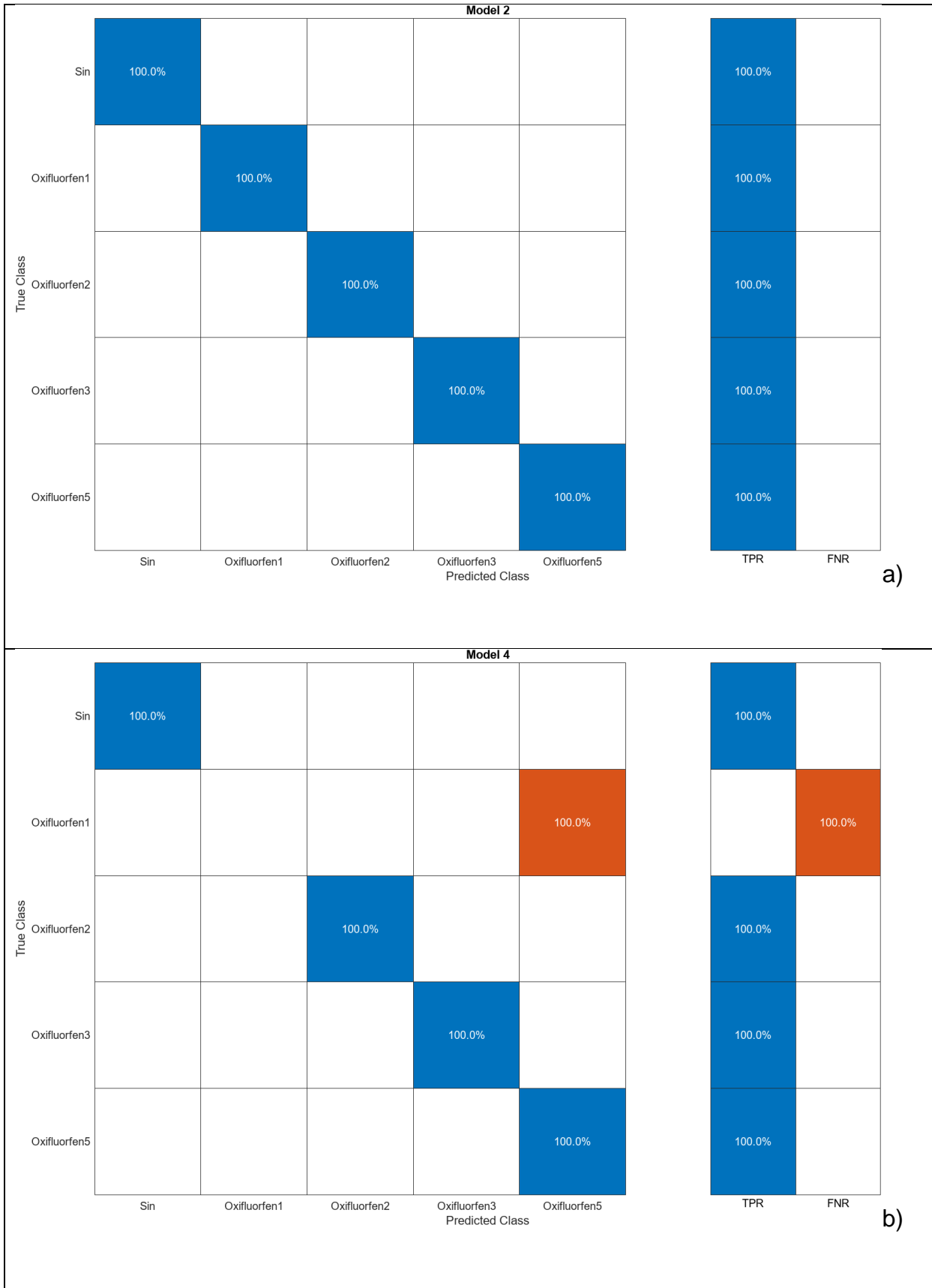


Ilustración 3-11. Matrices de confusión de los datos de test en el aprendizaje automático para los niveles de oxifluorfen en los espectros medios de las aceitunas: a: Análisis discriminante; b: k-NN.

El último producto analizado es el **tebuconazol** (Tabla 3-6), que igualmente presenta satisfactorios valores de exactitud en la validación y el test. En este caso, el mejor resultado se obtiene con SVM y el peor, aunque aceptable, en los árboles de decisión. Los tiempos de ejecución varían entre 10 y 20 segundos en la mayor parte de los casos.

El análisis de componentes principales, en el que se mantienen 5 componentes al 99% de la varianza muestra una cierta reducción en los valores de exactitud respecto al análisis con todo el espectro de longitudes de onda. En el caso de la selección de características, aun manteniendo 30, se produce una disminución generalizada de al menos 10 puntos en la exactitud en la mayor parte de los métodos.

Las matrices de confusión (Ilustración 3-12), de nuevo poco significativas, muestran en general la buena diferenciación entre los niveles, presentándose las mayores confusiones entre los valores medios (2-3) y bajos (4-5) de concentración del producto.

Modelo	Validac.	Test	Tiempo	Caract.	Técnica	ACP
Árboles Decisión	72,55	66,67	7,75	280/280		
An. Discriminante	82,35	83,33	13,12	280/280		
SVM	84,31	100,00	10,41	280/280		
KNN	74,51	83,33	9,41	280/280		
Bagged / Bosques Al.	70,59	66,67	11,84	280/280		
Red Neuronal	80,39	66,67	7,13	280/280		
Árboles Decisión	62,75	83,33	8,17	5 CP		99% Var.
An. Discriminante	76,27	83,33	10,24	5 CP		99% Var.
SVM	75,27	83,33	10,64	5 CP		99% Var.
KNN	70,39	83,33	10,18	5 CP		99% Var.
Bagged / Bosques Al.	66,47	83,33	12,44	5 CP		99% Var.
Red Neuronal	74,31	83,33	9,40	5 CP		99% Var.
Árboles Decisión	45,10	50,00	11,39	30/280	ANOVA	
An. Discriminante	80,39	100,00	17,12	30/280	ANOVA	
SVM	50,98	33,33	19,22	30/280	ANOVA	
KNN	54,90	50,00	16,22	30/280	ANOVA	
Bagged / Bosques Al.	62,75	50,00	13,80	30/280	ANOVA	
Red Neuronal	70,59	100,00	15,09	30/280	ANOVA	

Tabla 3-6. Resultados de la exactitud (accuracy) en validación y test del aprendizaje automático, para los niveles de tebuconazol en los espectros medios de las aceitunas.

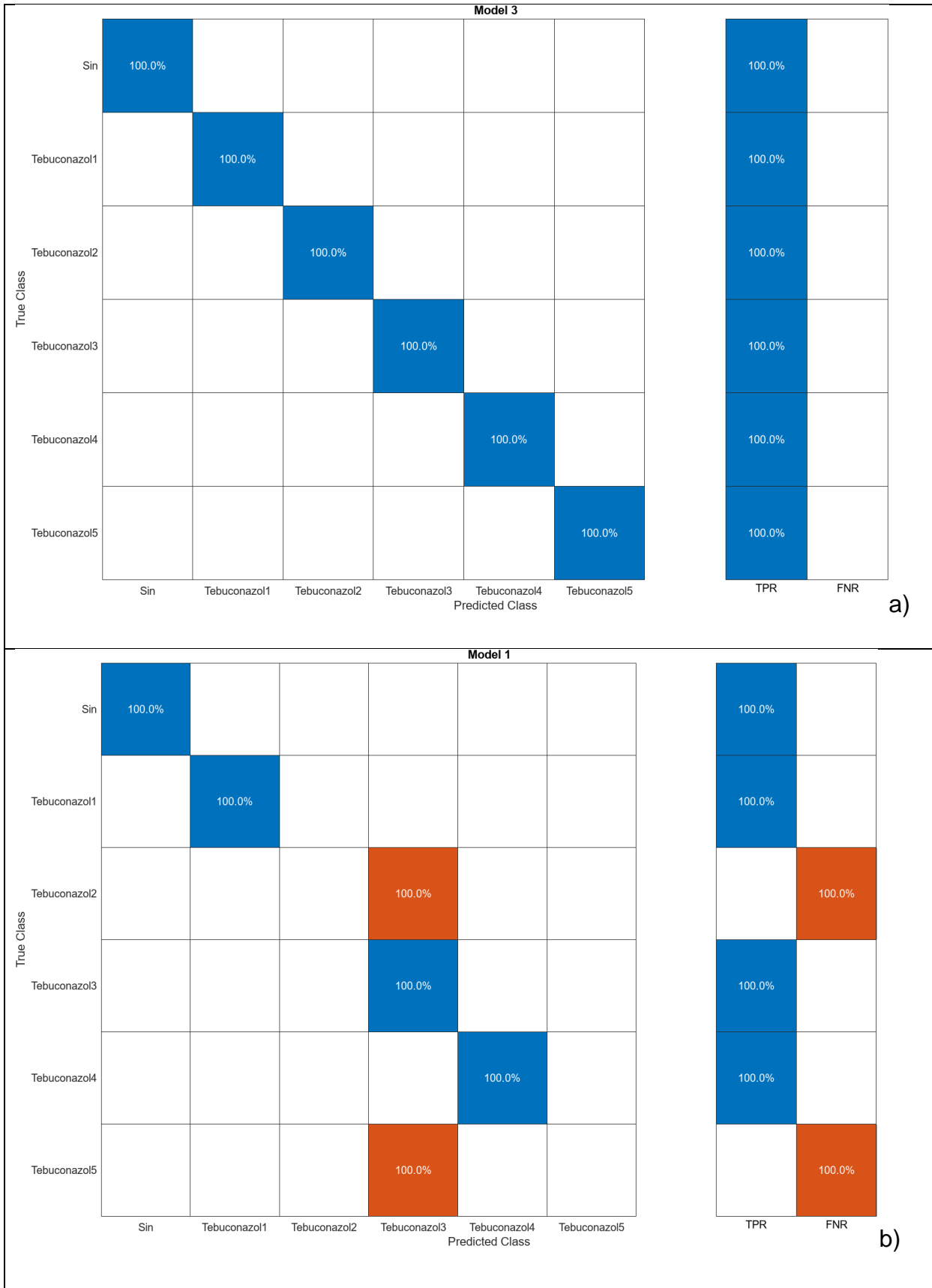


Ilustración 3-12. Matrices de confusión de los datos de test en el aprendizaje automático para los niveles de tebuconazol en los espectros medios de las aceitunas: a: SVM; b: Bosques aleatorios.

Del análisis realizado, se desprende una serie de observaciones:

- Los métodos de clasificación mediante aprendizaje automático, de uso general y disponibles en el Classification Learner de Matlab, proporcionan en general resultados bastante aceptables para la diferenciación y clasificación adecuada del tipo de fitosanitario (en torno al 80% en test y validación).
- En el nivel o concentración de producto aplicado sobre las aceitunas hay variaciones entre los distintos productos, estando los mejores resultados en diflufenican y los peores en la deltametrina.
- Los métodos son fáciles de aplicar y muy rápidos (por debajo de 10-20 segundos), en el caso de muestras tan reducidas en tamaño, como las correspondientes a las curvas medias de las muestras de aceitunas (234 muestras con sus correspondientes curvas espectrales).
- No obstante, el propio tamaño reducido de las muestras produce algunos desajustes y falta de significación en los resultados, especialmente en el caso del test (un 15%). Por ello, los resultados de la validación cruzada (5 *folds* en el 85% de la muestra total) parecen resultar más significativos y coherentes.
- En general, algunos métodos proporcionan mejores resultados, como son el análisis discriminante y SVM, mientras que otros presentan resultados más mediocres como el k-NN y especialmente los árboles de decisión.
- Esto contrasta en alguna medida con los resultados obtenidos en (Martínez-Gila, et al., 2024), donde los resultados para los métodos empleados (análisis discriminante, k-NN y distintas variantes de SVM) alcanzan valores próximos al 100% en exactitud y otros parámetros como la precisión, recuperación, área bajo la curva ROC y puntuación F1, para la fase de entrenamiento (y validación). Los valores son más parecidos a los obtenidos aquí cuando se consideran los datos de test independientes (exactitud entre 66% para y 95%), estando los mejores en el diflufenican y los peores en el oxifluorfen.
- En dicho trabajo se emplea la misma muestra de imágenes de aceitunas que en este, aunque el muestreo de los datos en las imágenes (100 píxeles con sus correspondientes curvas en cada una de los tipos y concentraciones) y otras partes del proceso (análisis de componentes principales, muestreo, partición de la muestra) son distintos.

- El análisis de componentes principales no muestra resultados demasiado prometedores y es necesario recurrir a una varianza del 99%, con aparición de entre 4 y 6 CP, para alcanzar resultados aceptables, aunque inferiores a los obtenidos con todas las bandas espectrales. Con el 95% (2-3 componentes) los resultados son bastante más deficientes.
- Lo mismo ocurre, incluso en mayor medida, con la selección de características (mediante ANOVA) donde es necesario recurrir a un mínimo de 30 características, que de todas formas no permite alcanzar los mismos resultados que con el espectro completo.

3.2.1.2. Curvas espectrales extraídas de las ventanas manualmente

Como se ha descrito en la metodología y se ha apuntado en las observaciones anteriores, a pesar de los aceptables e, incluso en algunos casos, satisfactorios resultados obtenidos con los espectros medios de las muestras de aceitunas, se optó en este trabajo por aumentar el tamaño de la muestra mediante la realización de una serie de ventanas que permiten extraer las curvas espectrales medias de las mismas. El objetivo es obtener una muestra más representativa, sobre todo a la hora de realizar tanto el entrenamiento, como sobre todo en la validación y el test externo en su caso.

Puesto que se han definido manualmente 10 ventanas por aceituna la muestra se multiplica por 10 (2340 curvas espectrales). Es una estrategia algo diferente a la seguida por (Martínez-Gila, et al., 2024), donde se extraen las curvas espectrales de todos los píxeles de una ventana centrada en las imágenes de las aceitunas. En definitiva, se ha seguido un método de aumento de la muestra para mejorar el proceso.

Los resultados del análisis se muestran en las tablas y figuras siguientes, siguiendo un proceso de explicación similar al empleado con los espectros medios. Así, en la Tabla 3-7 se muestran los resultados del **análisis por tipos de producto**, siendo el método del análisis discriminante, SVM y la red neuronal los que obtienen mejores resultados, en torno al 80%. El resto de métodos presentan menores valores de la exactitud, siendo los árboles de decisión los que presenta los resultados más bajos de hasta por debajo del 60% en test. Por lo tanto, se produce una situación muy similar a la de los espectros medios, aunque en este caso SVM presentan notablemente mejores resultados que en el caso anterior.

Los tiempos de ejecución son lógicamente superiores a los empleados con los espectros medios, aunque varían entre los distintos métodos, desde menos de 10 segundos para el análisis discriminante y árboles de decisión, hasta casi dos minutos en los bosques aleatorios.

El análisis de componentes principales, donde se mantienen 6 CP al 99% de la varianza no proporciona resultados especialmente satisfactorios reduciéndose en decenas de puntos la exactitud. Por el contrario, la selección de características solo muestra pequeñas reducciones de la exactitud, aunque siempre que se mantenga un número importante de características (30).

Las matrices de confusión (Ilustración 3-13), tomando como referencia el análisis discriminante, muestra una buena diferenciación general entre todos los productos, aunque una parte de la muestra sin tratar se confunde con la lambda-cihalotrina, que a su vez es etiquetada muchas veces como otros tipos de producto. En el método que proporciona los valores más bajos de exactitud hay muchas otras confusiones, además de la comentada, quedando el porcentaje de verdaderos positivos o sensibilidad en torno al 50-60% en prácticamente todos los casos, frente al caso anterior donde era en general superior al 80%.

Modelo	Validac.	Test	Tiempo	Caract.	Técnica	ACP
Árboles Decisión	58,44	57,14	9,51	280/280		
An. Discriminante	90,45	93,71	4,82	280/280		
SVM	77,09	81,43	19,96	280/280		
KNN	59,10	61,14	14,54	280/280		
Bagged / Bosques Al.	69,40	71,71	106,33	280/280		
Red Neuronal	89,60	91,71	35,75	280/280		
Árboles Decisión	53,22	52,86	7,08	6 CP		99% Var.
An. Discriminante	55,88	54,57	7,98	6 CP		99% Var.
SVM	59,85	59,43	13,36	6 CP		99% Var.
KNN	58,49	62,00	7,45	6 CP		99% Var.
Bagged / Bosques Al.	60,75	66,00	18,32	6 CP		99% Var.
Red Neuronal	61,81	62,57	19,69	6 CP		99% Var.
Árboles Decisión	51,91	49,71	13,05	30/280	ANOVA	
An. Discriminante	87,69	89,71	19,58	30/280	ANOVA	
SVM	76,38	79,14	21,74	30/280	ANOVA	
KNN	54,87	50,57	11,68	30/280	ANOVA	
Bagged / Bosques Al.	61,81	58,57	37,58	30/280	ANOVA	
Red Neuronal	90,95	93,43	31,45	30/280	ANOVA	

Tabla 3-7. Resultados de la exactitud (accuracy) en validación y test del aprendizaje automático por tipos de producto en los espectros medios de las ventanas definidas manualmente.

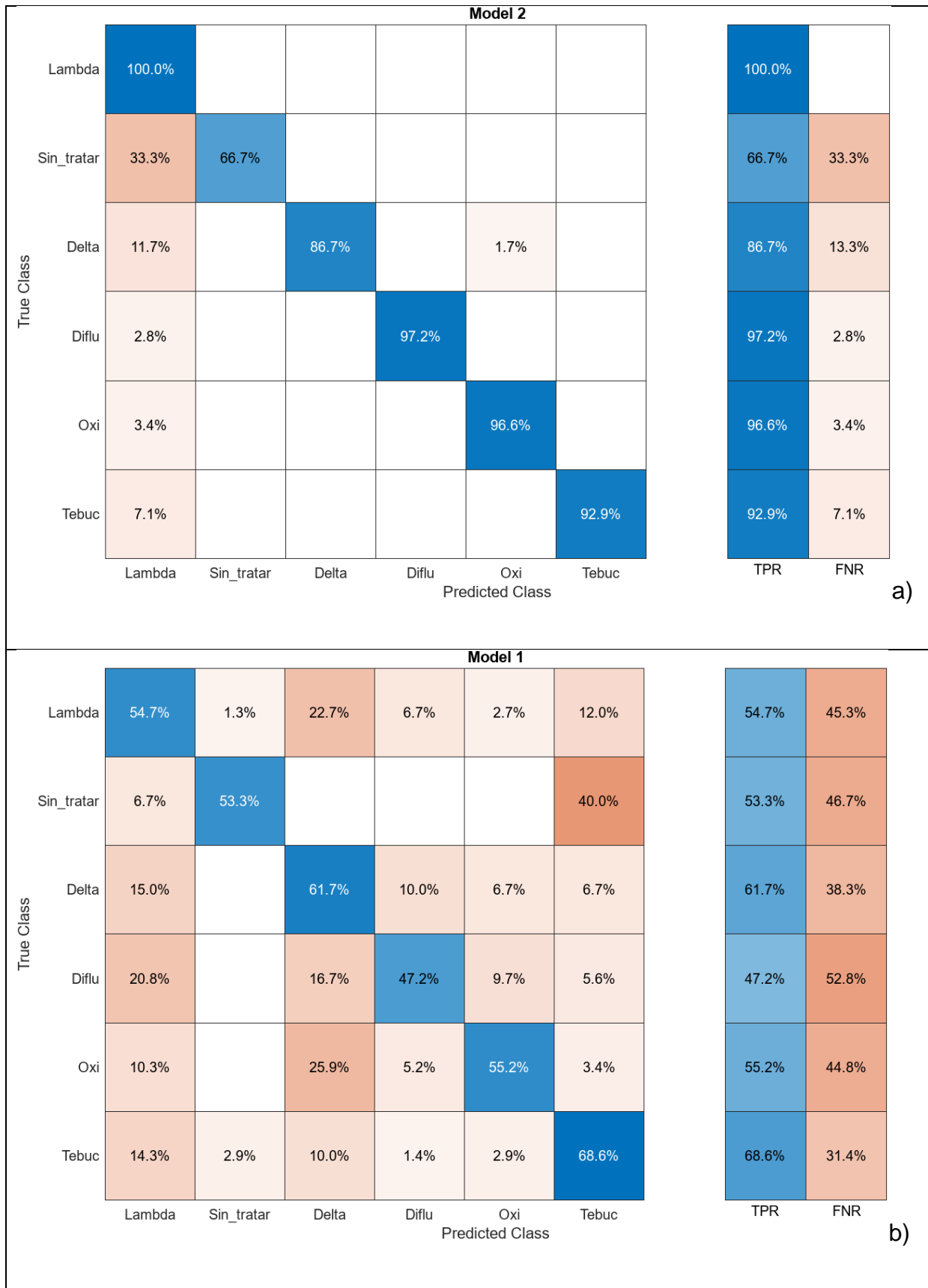


Ilustración 3-13. Matrices de confusión de los datos de test en el aprendizaje automático por tipos de producto en los espectros medios de las ventanas definidas manualmente: a: Análisis discriminante; b: Árboles de decisión.

En cuanto al análisis por niveles de producto aplicado, en el caso de la deltametrina (Tabla 3-8) se observa que los métodos que proporcionan los mejores resultados son la red neuronal y SVM, mientras que los peores se producen en los árboles de decisión y en el análisis discriminante (aunque en este caso hay cierta discrepancia entre los valores de validación y test). En general los resultados son algo mejores que en el correspondiente análisis del producto mediante los espectros medios. Los tiempos de ejecución son bajos de nuevo, ya que la muestra es más reducida que la anterior (tipos de producto) y no suelen superar los 10-15 segundos.

El análisis de componentes principales muestra en este caso resultados inferiores para los distintos métodos, reteniéndose 4 componentes. Sin embargo, la selección de factores, considerando 30, presenta valores bastante similares a los de todo el espectro, y en alguno caso, los mejora, como en el análisis discriminante.

Las matrices de confusión (Ilustración 3-14) muestran que la mayor confusión se produce entre los niveles próximos, como el nivel 2 con el 1 (no hay problema en este caso, ya que ambos están sobre el límite), el 3 con el 5 o incluso el sin tratar (en este caso el nivel 3 está sobre el límite a diferencia de los otros dos); y menos entre los niveles más alejados, aunque en general hay una cierta confusión general en este producto.

Modelo	Validac.	Test	Tiempo	Caract.	Técnica	ACP
Árboles Decisión	60,71	56,00	5,71	280/280		
An. Discriminante	54,59	86,67	4,04	280/280		
SVM	83,53	73,33	8,52	280/280		
KNN	68,24	66,67	4,91	280/280		
Bagged / Bosques Al.	69,88	69,33	22,81	280/280		
Red Neuronal	84,00	80,00	7,26	280/280		
Árboles Decisión	59,29	57,33	5,09	4 CP		99% Var.
An. Discriminante	67,29	58,67	3,30	4 CP		99% Var.
SVM	72,00	64,00	6,31	4 CP		99% Var.
KNN	66,12	70,67	3,97	4 CP		99% Var.
Bagged / Bosques Al.	66,82	65,33	10,14	4 CP		99% Var.
Red Neuronal	64,71	60,00	8,70	4 CP		99% Var.
Árboles Decisión	55,76	65,33	9,45	30/280	ANOVA	
An. Discriminante	93,41	89,33	7,45	30/280	ANOVA	
SVM	81,88	70,67	11,44	30/280	ANOVA	
KNN	67,06	66,67	6,91	30/280	ANOVA	
Bagged / Bosques Al.	64,00	62,67	25,63	30/280	ANOVA	
Red Neuronal	89,18	86,67	9,80	30/280	ANOVA	

Tabla 3-8. Resultados de la exactitud (accuracy) en validación y test del aprendizaje automático, para los niveles de deltametrina en las ventanas definidas manualmente.

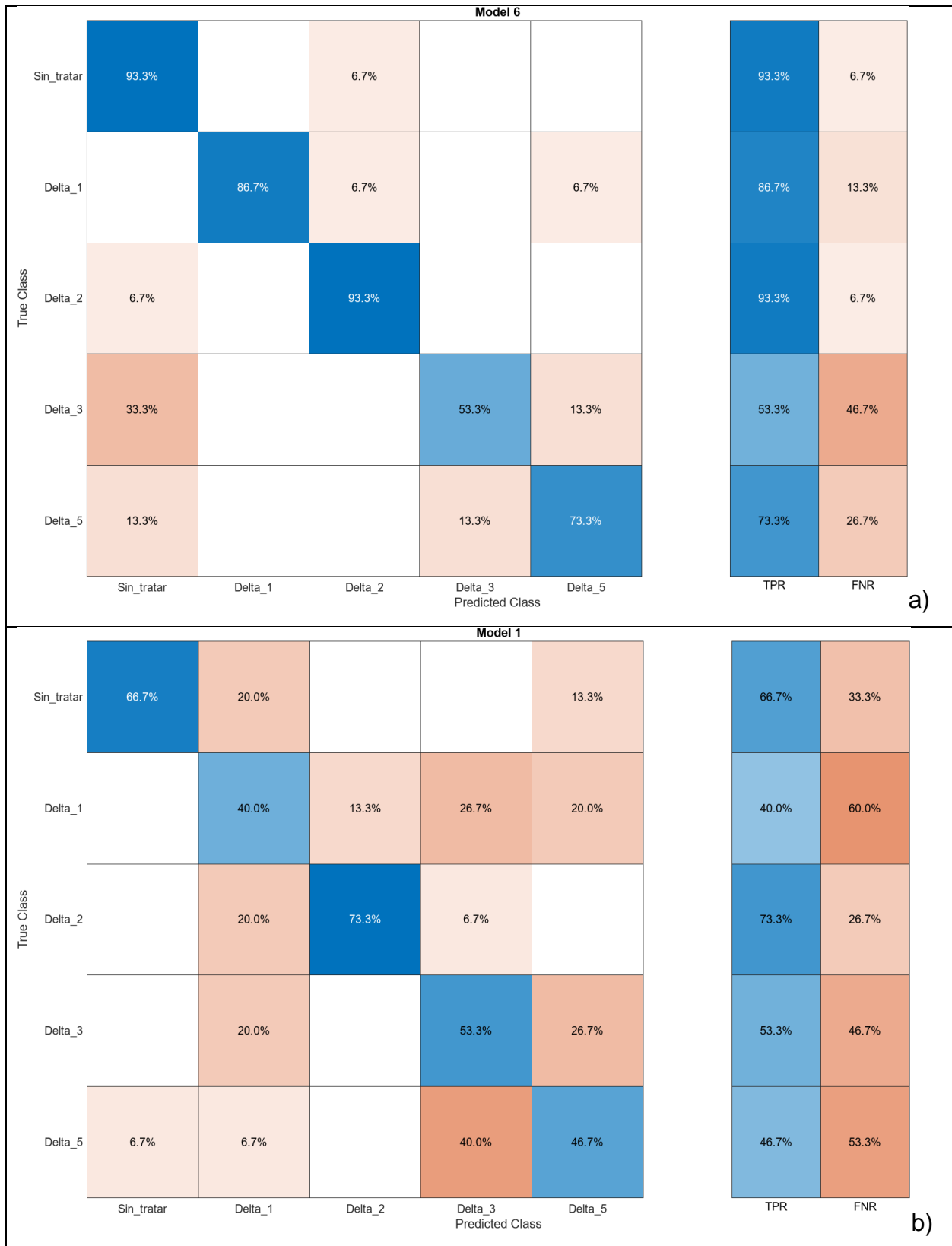


Ilustración 3-14. Matrices de confusión de los datos de test en el aprendizaje automático para los niveles de deltametrina en los espectros medios de las ventanas definidas manualmente: a: Red neuronal; b: Árboles de decisión.

Los valores de exactitud para los niveles de **diflufenican** (Tabla 3-9), son del mismo orden a los obtenidos mediante los espectros medios en prácticamente todos los métodos, no llegando a superarse casi nunca el 70-80% ni en validación ni en test. El método que muestra mejores valores ligeramente, sobre todo en test, es el análisis discriminante y el que peor los obtiene es el árbol de decisión, junto con k-NN. Los tiempos de ejecución de los algoritmos son en general inferiores a 10-20 segundos.

Al introducir el análisis de componentes principales, en este caso se obtienen valores bastante similares al operar con las 4 componentes que se mantienen al 99% de la varianza. Lo mismo sucede en la selección de factores donde los resultados son incluso más parecidos a los obtenidos con el espectro completo, empleando no obstante las 30 características más importantes.

Las matrices de confusión (Ilustración 3-15) muestra una cierta falta de diferenciación entre los niveles de concentración del producto más próximos como los niveles 2, 3 y 4, junto al 5 y las muestras sin tratar. De todos ellos, el más problemático es el nivel 2, por encima del límite europeo, mientras que el resto está por debajo.

Modelo	Validac.	Test	Tiempo	Caract.	Técnica	ACP
Árboles Decisión	60,85	52,87	6,32	280/280		
An. Discriminante	79,51	82,76	4,71	280/280		
SVM	77,69	74,71	9,10	280/280		
KNN	59,03	65,52	5,61	280/280		
Bagged / Bosques Al.	68,76	66,67	27,73	280/280		
Red Neuronal	78,30	77,01	7,23	280/280		
Árboles Decisión	69,57	68,97	5,33	4 CP		99% Var.
An. Discriminante	78,70	74,71	3,64	4 CP		99% Var.
SVM	79,92	75,86	7,70	4 CP		99% Var.
KNN	75,46	75,86	4,51	4 CP		99% Var.
Bagged / Bosques Al.	73,83	70,11	10,15	4 CP		99% Var.
Red Neuronal	76,06	71,26	9,59	4 CP		99% Var.
Árboles Decisión	48,68	50,57	10,05	30/280	ANOVA	
An. Discriminante	78,70	79,31	7,79	30/280	ANOVA	
SVM	70,18	77,01	10,93	30/280	ANOVA	
KNN	53,96	48,28	6,56	30/280	ANOVA	
Bagged / Bosques Al.	55,38	56,32	24,46	30/280	ANOVA	
Red Neuronal	83,16	83,91	16,03	30/280	ANOVA	

Tabla 3-9. Resultados de la exactitud (accuracy) en validación y test del aprendizaje automático, para los niveles de diflufenican en los espectros medios de las ventanas definidas manualmente.

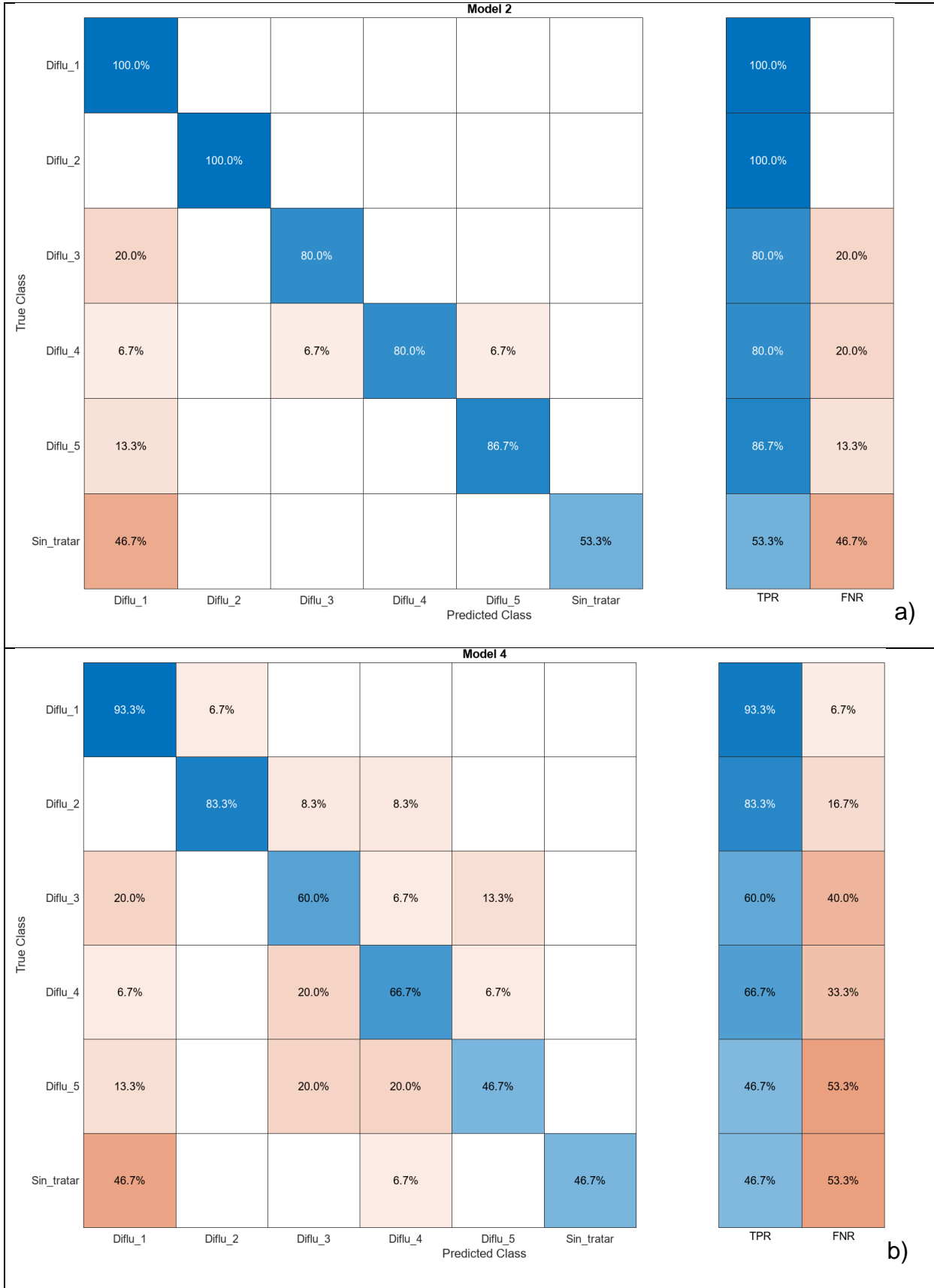


Ilustración 3-15. Matrices de confusión de los datos de test en el aprendizaje automático para los niveles de diflufenican en los espectros medios de las ventanas definidas manualmente: a: Análisis discriminante; b: k-NN.

El análisis del nivel de **lambda-cihalotrina** (Tabla 3-10) proporciona resultados mediocres (60%), incluso inferiores a los obtenidos con el espectro medio, especialmente en la validación, aunque en el test los valores sean superiores en algunos de los métodos (70-80%). Destacan el análisis discriminante y la red neuronal, mientras que k-NN y los árboles de decisión muestran valores más bajos. Los tiempos de ejecución se encuentran en la mayor parte de los casos entre 10 y 20 segundos.

Por su parte, el análisis con las componentes principales, con 5 CP al 99% de la varianza, muestra en este caso valores muy similares a los obtenidos con todo el espectro. Por su parte, al aplicar la selección de características (30), se produce una reducción ligera (5-10 puntos) de la exactitud en la mayor parte de los métodos.

Las matrices de confusión (Ilustración 3-16) muestra como en otros casos ciertas mezclas entre las concentraciones más próximas como el nivel 5 y las muestras sin tratar, o los niveles medios-bajos (3 al 5), quedando más definidos los niveles más altos (1 y 2). En este sentido hay un cierto problema con el nivel 3 que queda por encima del límite europeo, frente al 4, 5 y la muestra sin tratar, por debajo.

Modelo	Validac.	Test	Tiempo	Caract.	Técnica	ACP
Árboles Decisión	50,78	51,11	5,49	280/280		
An. Discriminante	59,80	81,11	6,18	280/280		
SVM	58,43	62,22	8,83	280/280		
KNN	44,71	56,67	5,06	280/280		
Bagged / Bosques Al.	60,20	60,00	27,59	280/280		
Red Neuronal	60,59	74,44	8,87	280/280		
Árboles Decisión	38,82	46,67	4,70	5 CP		99% Var.
An. Discriminante	48,63	51,11	4,66	5 CP		99% Var.
SVM	50,20	52,22	6,09	5 CP		99% Var.
KNN	47,06	52,22	4,32	5 CP		99% Var.
Bagged / Bosques Al.	45,29	56,67	10,24	5 CP		99% Var.
Red Neuronal	46,67	56,67	11,79	5 CP		99% Var.
Árboles Decisión	47,25	35,56	9,25	30/280	ANOVA	
An. Discriminante	74,71	68,89	7,58	30/280	ANOVA	
SVM	55,10	56,67	9,42	30/280	ANOVA	
KNN	48,82	36,67	16,05	30/280	ANOVA	
Bagged / Bosques Al.	50,78	38,89	18,93	30/280	ANOVA	
Red Neuronal	71,18	66,67	17,83	30/280	ANOVA	

Tabla 3-10. Resultados de la exactitud (accuracy) en validación y test del aprendizaje automático, para los niveles de lambda-cihalotrina en los espectros medios de las ventanas definidas manualmente.

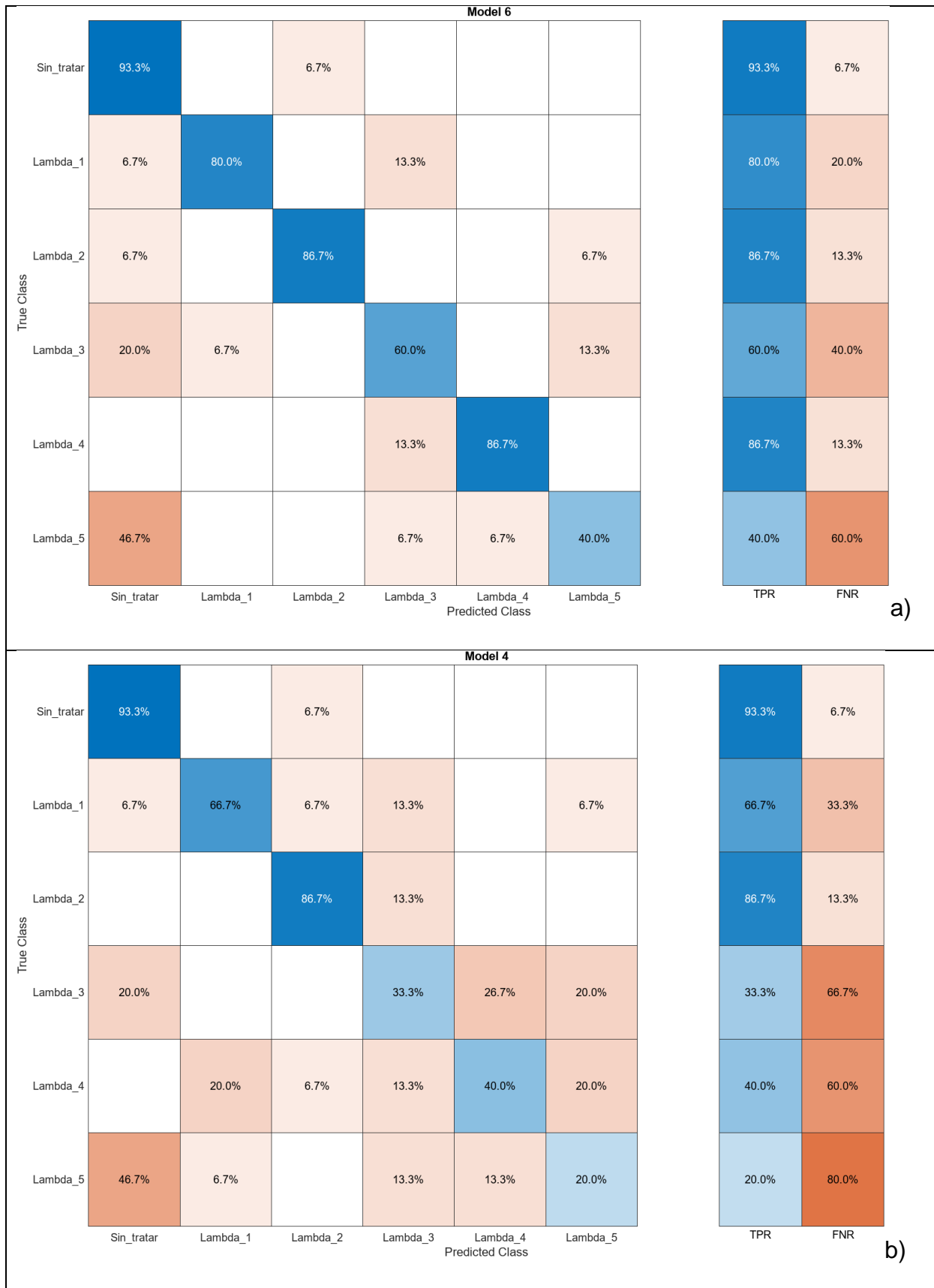


Ilustración 3-16. Matrices de confusión de los datos de test en el aprendizaje automático para los niveles de lambda-cihalotrina en los espectros medios de las ventanas definidas manualmente: a: Red neuronal; b: k-NN.

El **oxifluorfen** muestra valores moderados en la mayor parte de los clasificadores (Tabla 3-11), aunque en algunos llega a alcanzar valores bastante satisfactorios como SVM o análisis discriminante, especialmente en las muestras de test (cercanos a 90). Son del mismo orden que los obtenidos para los espectros medios. Los valores más bajos aparecen en los árboles de decisión y en k-NN. Los tiempos de ejecución son bastante reducidos, en general inferiores a 20 segundos.

En el análisis de componentes principales, en el que se mantienen 6 componentes al 99% de la varianza, los valores de exactitud son bastante similares en la mayor parte de los casos, por lo que es una técnica a aplicar en este caso. En la selección de características (30), se produce una reducción significativa en alguno de los métodos, mientras en otros como el análisis discriminante se produce incluso una ligera mejora.

Las matrices de confusión (Ilustración 3-17) muestran que la clasificación funciona bastante bien en el análisis discriminante, en prácticamente todos los niveles; mientras que, en la matriz correspondiente a los árboles de decisión, aparece confusión entre los niveles próximos (1 y 2), pero también en algún caso entre niveles de alejados (1 y 5 o 2 y 5), lo que plantea algún problema en relación al límite europeo.

Modelo	Validac.	Test	Tiempo	Caract.	Técnica	ACP
Árboles Decisión	68,59	63,01	5,90	280/280		
An. Discriminante	77,46	93,15	7,24	280/280		
SVM	88,01	84,93	7,87	280/280		
KNN	70,26	72,60	4,86	280/280		
Bagged / Bosques Al.	71,94	79,45	23,31	280/280		
Red Neuronal	81,06	86,30	6,82	280/280		
Árboles Decisión	58,27	53,42	8,79	6 CP		99% Var.
An. Discriminante	73,62	73,97	4,95	6 CP		99% Var.
SVM	75,30	72,60	11,91	6 CP		99% Var.
KNN	66,91	68,49	6,62	6 CP		99% Var.
Bagged / Bosques Al.	65,95	63,01	16,02	6 CP		99% Var.
Red Neuronal	68,82	60,27	16,52	6 CP		99% Var.
Árboles Decisión	54,68	50,68	11,46	30/280	ANOVA	
An. Discriminante	80,10	79,45	8,48	30/280	ANOVA	
SVM	71,22	73,97	8,67	30/280	ANOVA	
KNN	58,03	58,90	17,78	30/280	ANOVA	
Bagged / Bosques Al.	57,07	63,01	13,71	30/280	ANOVA	
Red Neuronal	79,14	78,08	17,69	30/280	ANOVA	

Tabla 3-11. Resultados de la exactitud (accuracy) en validación y test del aprendizaje automático, para los niveles de oxifluorfen en los espectros medios de las ventanas definidas manualmente.

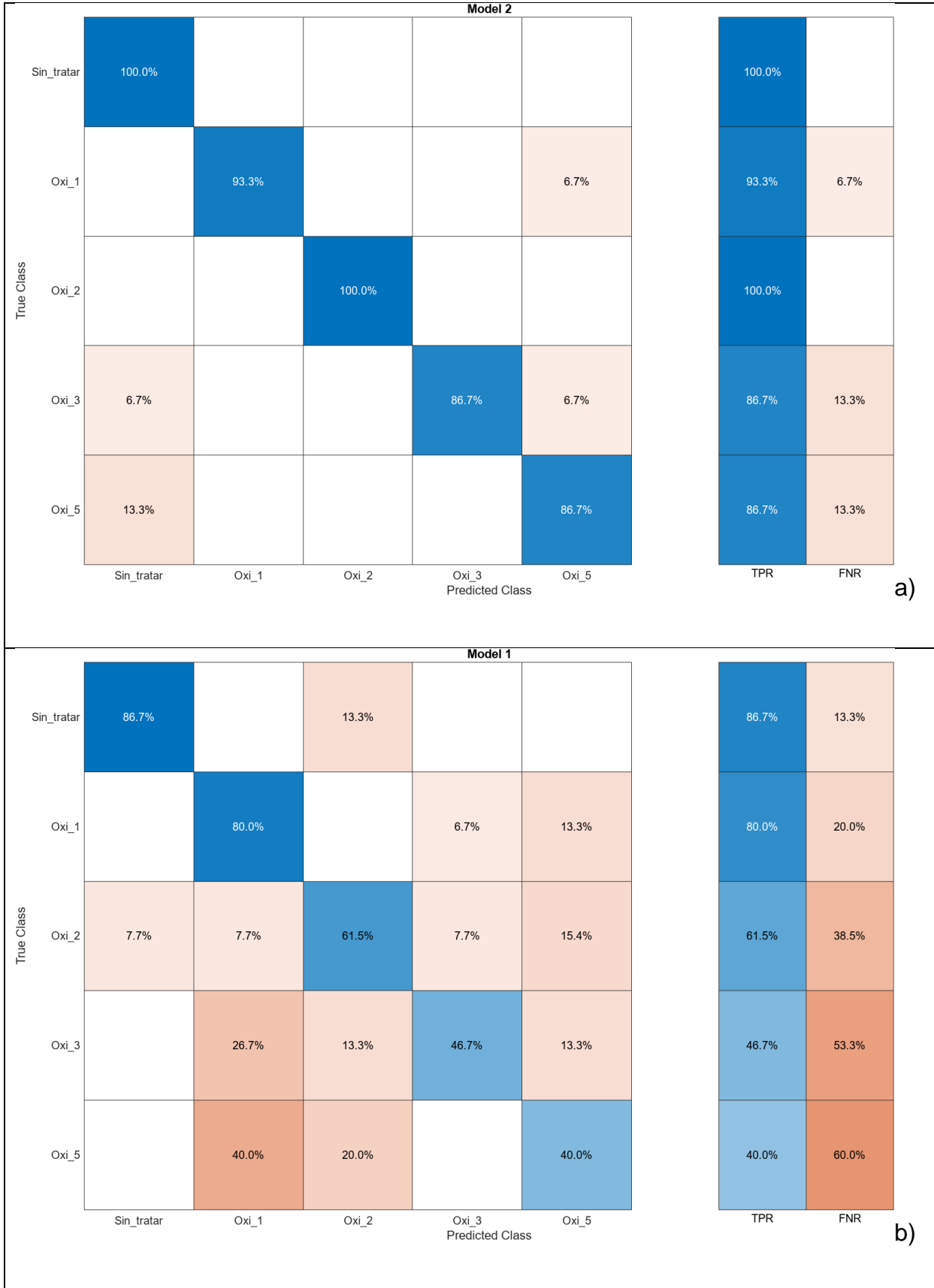


Ilustración 3-17. Matrices de confusión de los datos de test en el aprendizaje automático para los niveles de oxifluorfen en los espectros medios de las ventanas definidas manualmente: a: Análisis discriminante; b: Árboles de decisión.

En cuanto al último producto analizado, el **tebuconazol** (Tabla 3-12) presenta valores diferentes de exactitud en la validación y el test, para los distintos métodos. Así, presenta resultados aceptables (en torno al 70%) en el análisis discriminante y la red neuronal y valores bastante más discretos en k-NN y árboles de decisión. En cualquier caso, los valores son sensiblemente inferiores a los obtenidos en el espectro medio para este producto. Los tiempos de ejecución están casi siempre por debajo de 20 segundos y en algunos casos de 10 segundos.

El análisis de componentes principales, en el que se mantienen 5 componentes al 99% de la varianza, muestra unos valores de exactitud del mismo orden que los que emplean todo el espectro de bandas. En el caso de la selección de características, manteniendo 30, se obtienen valores muy similares en prácticamente todos los métodos.

Las matrices de confusión (Ilustración 3-18) muestran una cierta mezcla entre los valores más altos (1-2 y 2-3) y los más bajos (4-5) de concentración del producto, estando bastante bien diferenciadas las muestras sin tratar. A la vista de esos resultados no se plantea ningún problema ya que el límite europeo está en este caso entre 1-3 (sobre el límite) y 4-5 (bajo el límite).

Modelo	Validac.	Test	Tiempo	Caract.	Técnica	ACP
Árboles Decisión	48,77	48,81	6,61	280/280		
An. Discriminante	73,46	84,52	7,92	280/280		
SVM	67,90	72,62	9,34	280/280		
KNN	45,68	55,95	5,65	280/280		
Bagged / Bosques Al.	58,23	60,71	31,02	280/280		
Red Neuronal	70,16	75,00	9,62	280/280		
Árboles Decisión	44,24	41,67	5,01	5 CP		99% Var.
An. Discriminante	57,20	54,76	4,92	5 CP		99% Var.
SVM	54,94	54,76	6,47	5 CP		99% Var.
KNN	51,65	47,62	4,51	5 CP		99% Var.
Bagged / Bosques Al.	53,09	53,57	10,46	5 CP		99% Var.
Red Neuronal	53,91	57,14	11,43	5 CP		99% Var.
Árboles Decisión	42,18	41,67	9,78	30/280	ANOVA	
An. Discriminante	77,16	80,95	7,68	30/280	ANOVA	
SVM	58,64	57,14	15,57	30/280	ANOVA	
KNN	41,36	44,05	10,06	30/280	ANOVA	
Bagged / Bosques Al.	46,50	46,43	19,74	30/280	ANOVA	
Red Neuronal	65,64	75,00	18,00	30/280	ANOVA	

Tabla 3-12. Resultados de la exactitud (accuracy) en validación y test del aprendizaje automático, para los niveles de tebuconazol en los espectros medios de las ventanas definidas manualmente.

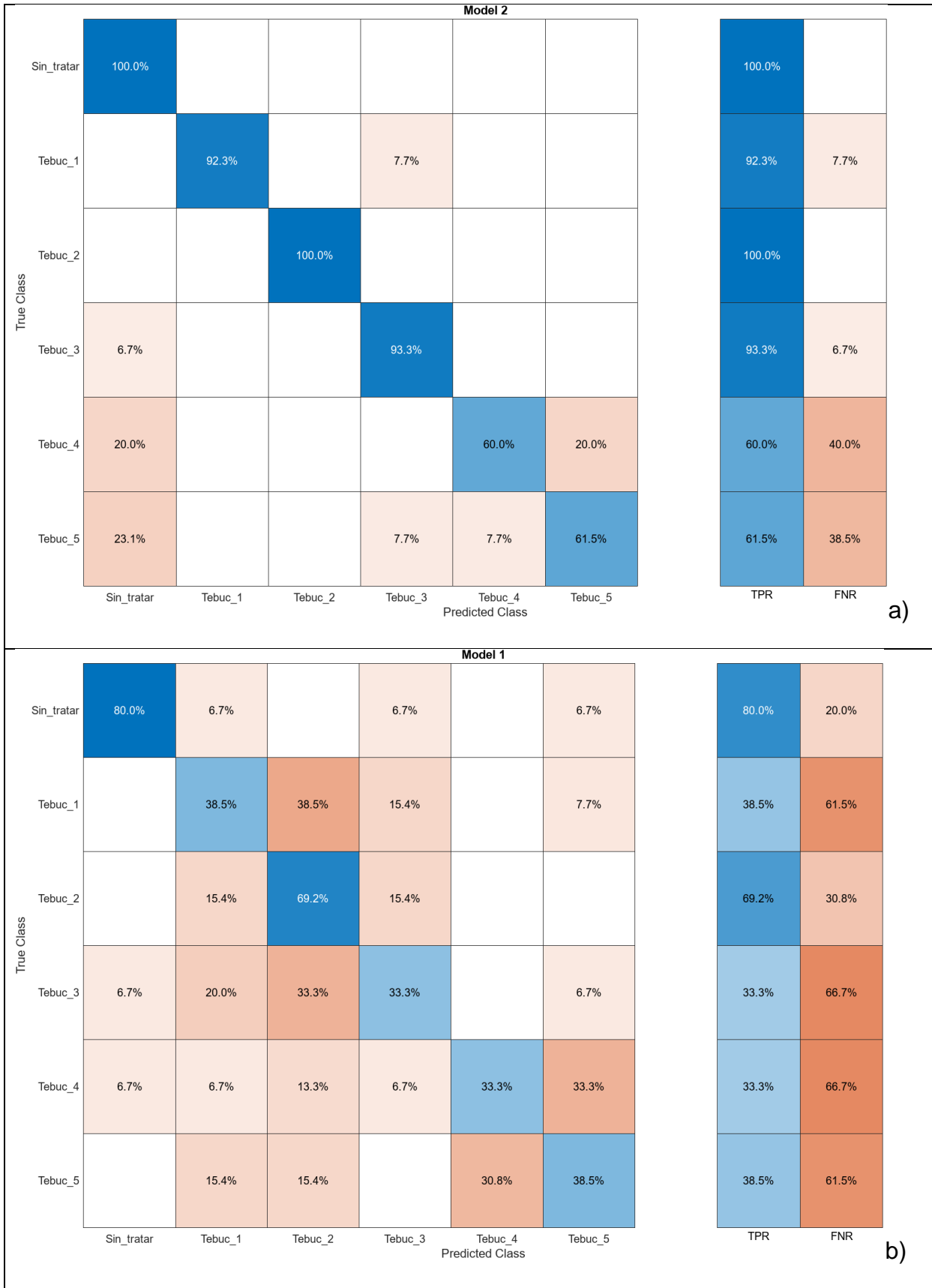


Ilustración 3-18. Matrices de confusión de los datos de test en el aprendizaje automático para los niveles de tebuconazol en los espectros medios de las ventanas definidas manualmente: a) Análisis discriminante; b) Árboles de decisión.

Del análisis realizado, se desprende una serie de observaciones:

- Los métodos de clasificación mediante aprendizaje automático empleados proporcionan de nuevo resultados bastante aceptables. En la clasificación del tipo de fitosanitario se alcanzan valores variables entre el 60 y 90%
- Lo mismo sucede entre los niveles de concentración, con aún mayor variabilidad entre los métodos y productos. En este último caso, los mejores resultados se producen en el oxifluorfen y los peores en tebuconazol.
- De nuevo los tiempos de ejecución son muy rápidos (por debajo de 20-30 segundos y menos de 10 en algunos casos), a pesar de que las muestras son de un orden de magnitud mayor que las correspondientes a los espectros medios (2340 ventanas con sus correspondientes curvas espectrales).
- El mayor tamaño de las muestras permite una mayor estabilidad y significación en los cálculos de la exactitud de las muestras, especialmente las de test que ahora tienen un número suficiente de unidades. Igualmente el entrenamiento produce resultados más fiables.
- En general, los resultados son bastante similares a los obtenidos con los espectros medios, en algunos casos mejoran ligeramente como en los tipos de producto o los niveles de deltametrina; son bastante similares como en los niveles de diflufenican y oxifluorfen; o se reducen de forma clara como en los niveles de lambda-cihalotrina o tebuconazol.
- En la comparativa entre métodos, algunos proporcionan mejores resultados, que nuevamente son el análisis discriminante, SVM y las redes neuronales mientras que otros presentan resultados peores como el k-NN y especialmente los árboles de decisión. Como anteriormente, los valores son inferiores a los obtenidos en validación (Martínez-Gila, et al., 2024), en este caso con un tamaño más equiparable de la muestra, aunque con un método de muestreo distinto. Sin embargo, son bastante similares para el test.
- El análisis de componentes presenta resultados comparativamente mejores que en el caso anterior, aunque de nuevo recurriendo a una varianza del 99%, con aparición de entre 4 y 6 CP; al igual que la selección de características (mediante ANOVA) con 30; en ambos casos, no se produce una reducción demasiado grande de los resultados, excepto en algunos casos.

3.2.1.3. Curvas espectrales extraídas de las cuadrículas automáticamente

Con objeto de probar con una muestra más grande para entrenar la red, obtenida de forma automática desde las imágenes, se procedió de esta manera generando una muestra de 63800 espectros medios de cuadrículas generadas automáticamente en las imágenes de aceitunas, descartando los elementos ajenos a las mismas y sus partes dañadas o defectos.

Los resultados del análisis se muestran en las tablas y las figuras siguientes, siguiendo un esquema de explicación similar al empleado con los espectros medios de las aceitunas y los espectros medios de las ventanas definidas manualmente. Así, en la Tabla 3-13 se muestran los resultados del **análisis por tipos de producto fitosanitario**. El análisis discriminante y la red neuronal son los métodos que obtienen mejores resultados, en torno al 98-99%, seguido por todos los demás métodos (SVM, k-NN y bosques aleatorios) con exactitud cercana al 90%; mientras que los árboles aleatorios apenas superan el 60%. Supone por lo tanto un aumento mayor o menor, según los métodos, respecto a los espectros medios y los espectros de las ventanas definidas por el usuario.

Los tiempos de ejecución son, sin embargo, bastante superiores a los empleados en los casos anteriores, desde unos pocos minutos hasta sobre pasar 1 hora como en el caso de los bosques aleatorios. Se da la circunstancia de que algunos métodos con análisis de componentes principales o selección de características no han llegado a finalizar su ejecución.

El análisis de componentes principales, donde se mantienen 6 CP al 99% de la varianza, no proporciona resultados satisfactorios reduciéndose la exactitud en la mayor parte de los casos, salvo excepciones. Lo mismo sucede con la selección de características, aun manteniendo un número importante de características (30).

Las matrices de confusión (Ilustración 3-19), tomando como referencia el análisis discriminante, muestra una buena diferenciación general entre todos los productos, prácticamente sin errores. En el método que proporciona los valores más bajos de exactitud, los árboles de decisión, hay una cierta confusión entre los distintos productos analizados.

Modelo	Validac.	Test	Tiempo	Caract.	Técnica	ACP
Árboles Decisión	59,30	59,14	463,51	280/280		
An. Discriminante	98,16	98,24	337,44	280/280		
SVM	89,72	90,80	4286,35	280/280		
KNN	90,27	91,32	2973,42	280/280		
Bagged / Bosques Al.	89,67	89,99	5662,57	280/280		
Red Neuronal	99,36	99,51	2103,26	280/280		
Árboles Decisión	69,10	67,94	984,73	6 CP		99% Var.
An. Discriminante	47,94	47,16	967,48	6 CP		99% Var.
SVM			1807,61	6 CP		99% Var.
KNN	84,86	84,93	999,69	6 CP		99% Var.
Bagged / Bosques Al.	84,46	84,48	1356,24	6 CP		99% Var.
Red Neuronal	75,26	75,15	1684,46	6 CP		99% Var.
Árboles Decisión	42,06	41,87	1627,42	30/280	ANOVA	
An. Discriminante	64,84	64,86	1611,22	30/280	ANOVA	
SVM			2505,35	30/280	ANOVA	
KNN	66,20	67,02	1738,12	30/280	ANOVA	
Bagged / Bosques Al.	69,76	70,10	2260,66	30/280	ANOVA	
Red Neuronal	78,48	78,35	2274,67	30/280	ANOVA	

Tabla 3-13. Resultados de la exactitud (accuracy) en validación y test del aprendizaje automático por tipos de producto en los espectros medios de las cuadrículas extraídas automáticamente.

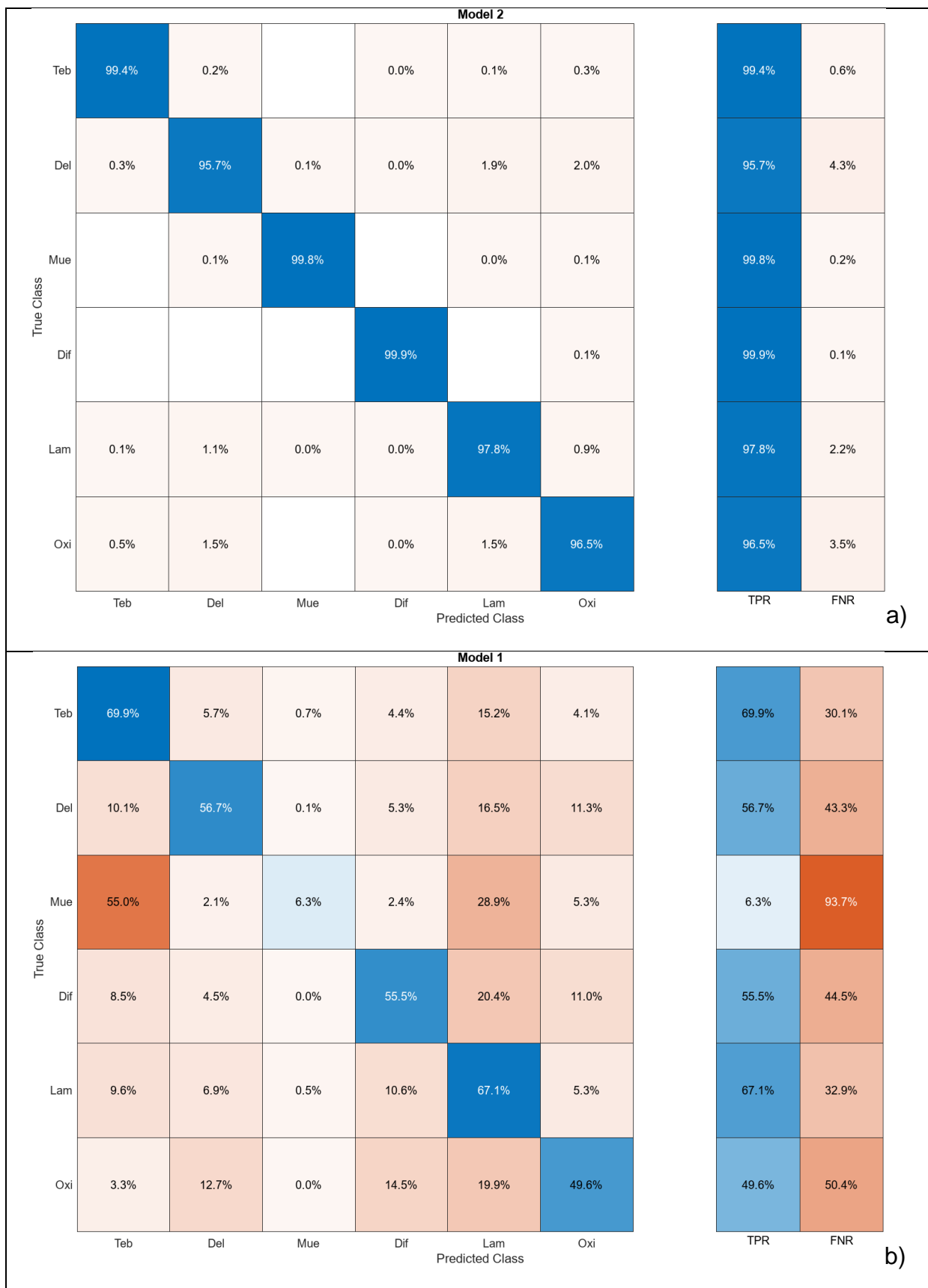


Ilustración 3-19. Matrices de confusión de los datos de test en el aprendizaje automático por tipos de producto en los espectros medios de las cuadrículas extraídas automáticamente: a) Análisis discriminante; b) Árboles de decisión.

En cuanto al análisis por niveles de producto aplicado, en el caso de la **deltametrina** (Tabla 3-14) se observa que los métodos que proporcionan los mejores resultados son la red neuronal y el análisis discriminante (98-99%), seguidos por SVM (95%), aunque en general todos los métodos tienen muy buen desempeño excepto los árboles de decisión que apenas superan el 60%. Los tiempos de ejecución son ya más bajos que en los tipos, pero significativamente más elevados que en los muestreos anteriores, entre 1 y 10 minutos, aproximadamente.

El análisis de componentes principales muestra en este caso resultados inferiores para los distintos métodos, aun con variaciones, ya que k-NN y bosque aleatorios tienen valores similares al de todo el espectro; en este caso, se retienen 4 componentes. La selección de factores, considerando 30, presenta un comportamiento bastante similar, ya que descendiendo con distinta intensidad respecto al análisis del espectro completo en los distintos métodos.

Las matrices de confusión (Ilustración 3-20) muestran que la buena diferenciación general entre los niveles del producto en los métodos que mejor desempeño tienen. En los árboles de decisión se aprecia una mayor confusión entre los niveles próximos, como el nivel 1 con el 2, o el 3 con el 5, lo que origina una cierta discrepancia ya que el nivel 3 está por encima del límite y el 5 por debajo.

Modelo	Validac.	Test	Tiempo	Caract.	Técnica	ACP
Árboles Decisión	62,67	63,83	82,88	280/280		
An. Discriminante	98,73	98,31	64,38	280/280		
SVM	94,85	94,61	183,62	280/280		
KNN	83,87	86,27	159,44	280/280		
Bagged / Bosques Al.	85,14	85,68	625,29	280/280		
Red Neuronal	99,05	99,18	428,87	280/280		
Árboles Decisión	66,58	68,74	22,06	280/280	4 CP	99% Var.
An. Discriminante	62,99	65,09	15,04	280/280	4 CP	99% Var.
SVM	67,01	67,97	85,24	280/280	4 CP	99% Var.
KNN	80,56	83,33	17,94	280/280	4 CP	99% Var.
Bagged / Bosques Al.	79,40	82,30	38,63	280/280	4 CP	99% Var.
Red Neuronal	72,21	73,86	77,84	280/280	4 CP	99% Var.
Árboles Decisión	46,61	45,97	27,37	30/280	ANOVA	
An. Discriminante	83,84	83,28	18,68	30/280	ANOVA	
SVM	75,69	77,02	342,63	30/280	ANOVA	
KNN	61,70	62,64	36,66	30/280	ANOVA	
Bagged / Bosques Al.	63,06	63,51	107,72	30/280	ANOVA	
Red Neuronal	86,32	86,55	93,40	30/280	ANOVA	

Tabla 3-14. Resultados de la exactitud (accuracy) en validación y test del aprendizaje automático, para los niveles de deltametrina de las cuadrículas extraídas automáticamente.

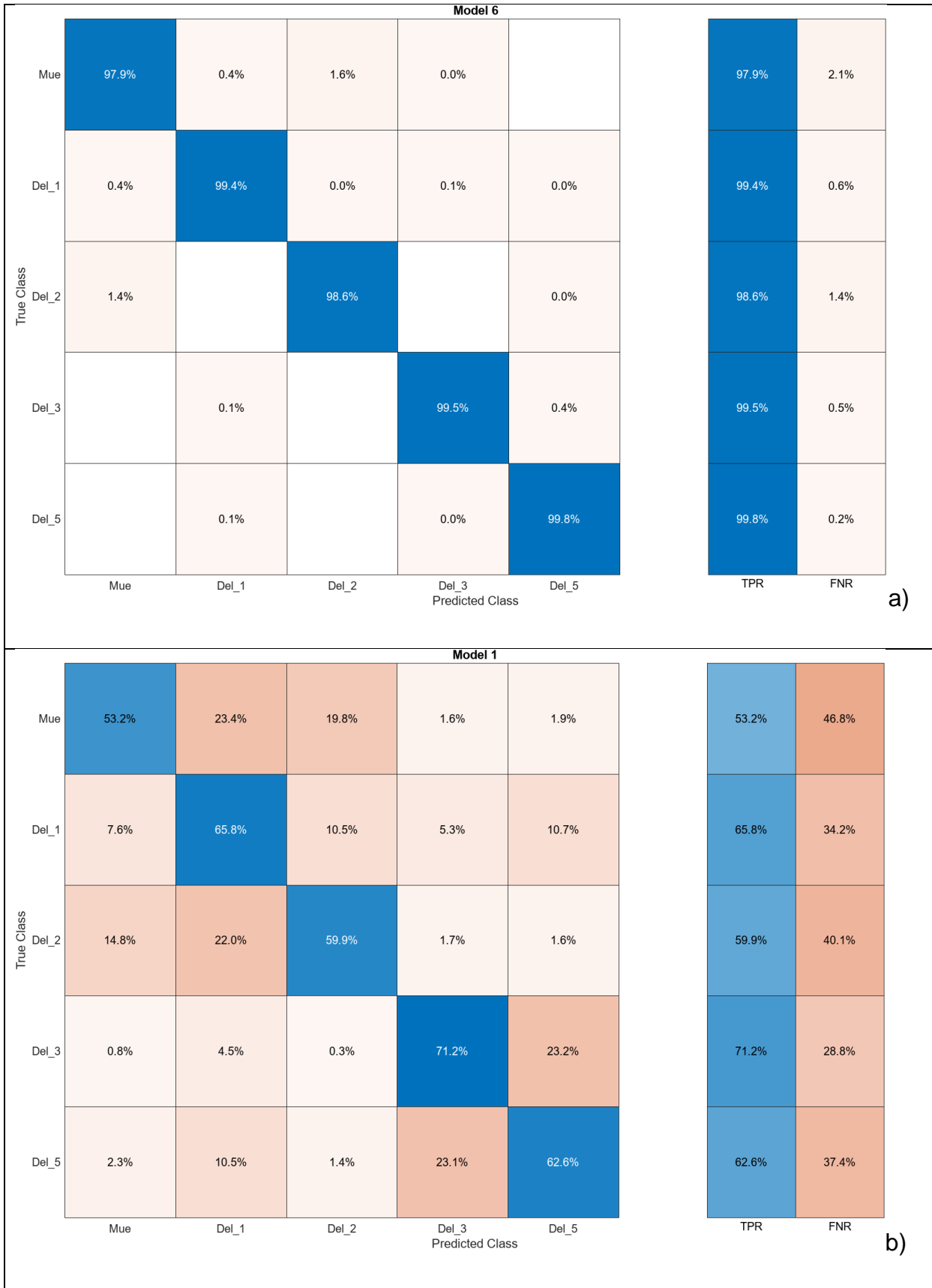


Ilustración 3-20. Matrices de confusión de los datos de test en el aprendizaje automático para los niveles de deltametrina en los espectros medios de las cuadrículas extraídas automáticamente: a: Red neuronal; b: Árboles de decisión.

Los valores de exactitud para los niveles de **diflufenican** (Tabla 3-15), son como en el caso anterior muy elevados en algunos métodos como el análisis discriminante y la red neuronal (97-98%), seguidos de SVM (91%); por el contrario, el método con menor exactitud son nuevamente los árboles de decisión (60%). Los tiempos de ejecución de los algoritmos son en general de uno a unos diez minutos.

Al introducir el análisis de componentes principales (se mantienen 4 componentes al 95%), se produce un descenso en los valores de exactitud de diferente manera según los métodos, aunque en algunos casos, como el análisis discriminante, es bastante acusado. Lo mismo sucede con la selección de características, aunque en este caso algunos de los métodos, como las redes neuronales, mantienen valores de exactitud bastante satisfactorios (90%).

Las matrices de confusión (Ilustración 3-21) muestran la buena diferenciación entre los niveles de concentración de los distintos productos, aunque en el método de los árboles aleatorios se aprecia cierta confusión entre los niveles de concentración contiguos (3 y 4), lo que no representa ningún problema ya que ambos están bajo el límite europeo.

Modelo	Validac.	Test	Tiempo	Caract.	Técnica	ACP
Árboles Decisión	60,23	59,65	81,54	280/280		
An. Discriminante	97,01	96,82	61,19	280/280		
SVM	91,14	91,32	174,96	280/280		
KNN	80,18	81,80	188,63	280/280		
Bagged / Bosques Al.	80,65	81,26	715,87	280/280		
Red Neuronal	98,58	98,80	528,57	280/280		
Árboles Decisión	55,10	55,18	189,55	4 CP		99% Var.
An. Discriminante	51,42	50,46	195,92	4 CP		99% Var.
SVM	55,19	55,01	363,51	4 CP		99% Var.
KNN	60,71	59,94	208,72	4 CP		99% Var.
Bagged / Bosques Al.	63,73	62,64	300,34	4 CP		99% Var.
Red Neuronal	59,73	58,78	454,96	4 CP		99% Var.
Árboles Decisión	58,95	57,31	270,40	30/280	ANOVA	
An. Discriminante	86,03	86,18	294,79	30/280	ANOVA	
SVM	82,27	82,79	476,62	30/280	ANOVA	
KNN	68,80	69,48	324,59	30/280	ANOVA	
Bagged / Bosques Al.	72,51	72,77	474,79	30/280	ANOVA	
Red Neuronal	91,07	91,32	670,85	30/280	ANOVA	

Tabla 3-15. Resultados de la exactitud (accuracy) en validación y test del aprendizaje automático, para los niveles de diflufenican en los espectros medios de las cuadrículas extraídas automáticamente.

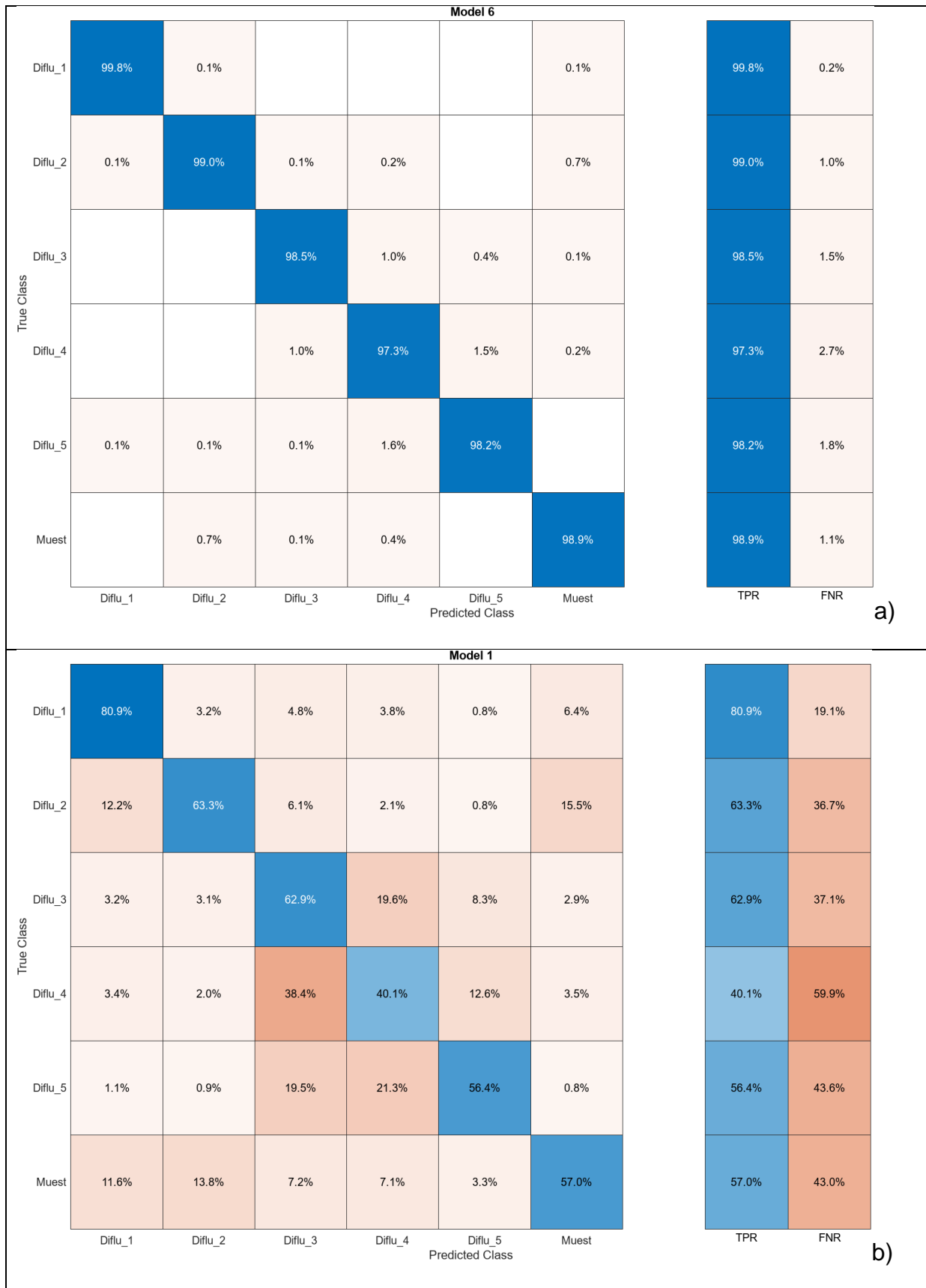


Ilustración 3-21. Matrices de confusión de los datos de test en el aprendizaje automático para los niveles de diflufenican en los espectros medios de las cuadrículas extraídas automáticamente: a: Red neuronal; b: Árboles de decisión.

El análisis del nivel de **lambda-cihalotrina** (Tabla 3-16) proporciona resultados muy elevados en dos métodos, análisis discriminante y red neuronal (por encima del 90%) y valores más moderados en el resto de los métodos, hasta llegar a los árboles aleatorios que se queda por debajo del 55%. Los tiempos de ejecución son de varios minutos e incluso algún método no ha llegado a finalizar el entrenamiento.

Por su parte, el análisis con las componentes principales, con 5 CP al 99% de la varianza, muestra en este caso valores bastante más bajos que los obtenidos con todo el espectro, no superando prácticamente el 50%. Por su parte, al aplicar la selección de características (30), se produce también una reducción clara, aunque los valores no son tan bajos como en los componentes principales.

Las matrices de confusión (Ilustración 3-22) muestran la buena diferenciación entre los niveles de concentración de los distintos productos; sin embargo, en el método de los árboles aleatorios se aprecia confusión fundamentalmente entre los niveles de concentración más alto (1 y 2, 2 y 3) o más bajos (4 y 5), lo que no representa tampoco ningún problema ya que los niveles 1-3 están sobre el límite europeo y los niveles 4-5 bajo el mismo.

Modelo	Validac.	Test	Tiempo	Caract.	Técnica	ACP
Árboles Decisión	53,71	53,75	143,03	280/280		
An. Discriminante	94,81	94,47	113,44	280/280		
SVM	79,46	81,99	325,45	280/280		
KNN	74,61	76,82	279,69	280/280		
Bagged / Bosques Al.	74,73	77,28	941,94	280/280		
Red Neuronal	95,34	96,69	710,22	280/280		
Árboles Decisión	47,98	47,58	415,20	5 CP		99% Var.
An. Discriminante	42,18	41,75	430,98	5 CP		99% Var.
SVM	44,42	43,23	744,33	5 CP		99% Var.
KNN	51,45	52,30	442,68	5 CP		99% Var.
Bagged / Bosques Al.	55,18	57,19	589,42	5 CP		99% Var.
Red Neuronal	51,52	50,60	809,00	5 CP		99% Var.
Árboles Decisión	40,04	42,07	524,19	30/280	ANOVA	
An. Discriminante	61,55	62,54	536,16	30/280	ANOVA	
SVM			1712,43	30/280	ANOVA	
KNN	51,01	53,63	555,57	30/280	ANOVA	
Bagged / Bosques Al.	53,59	55,77	790,79	30/280	ANOVA	
Red Neuronal	65,83	67,03	1004,09	30/280	ANOVA	

Tabla 3-16. Resultados de la exactitud (accuracy) en validación y test del aprendizaje automático, para los niveles de lambda-cihalotrina en los espectros medios de las cuadrículas extraídas automáticamente.9

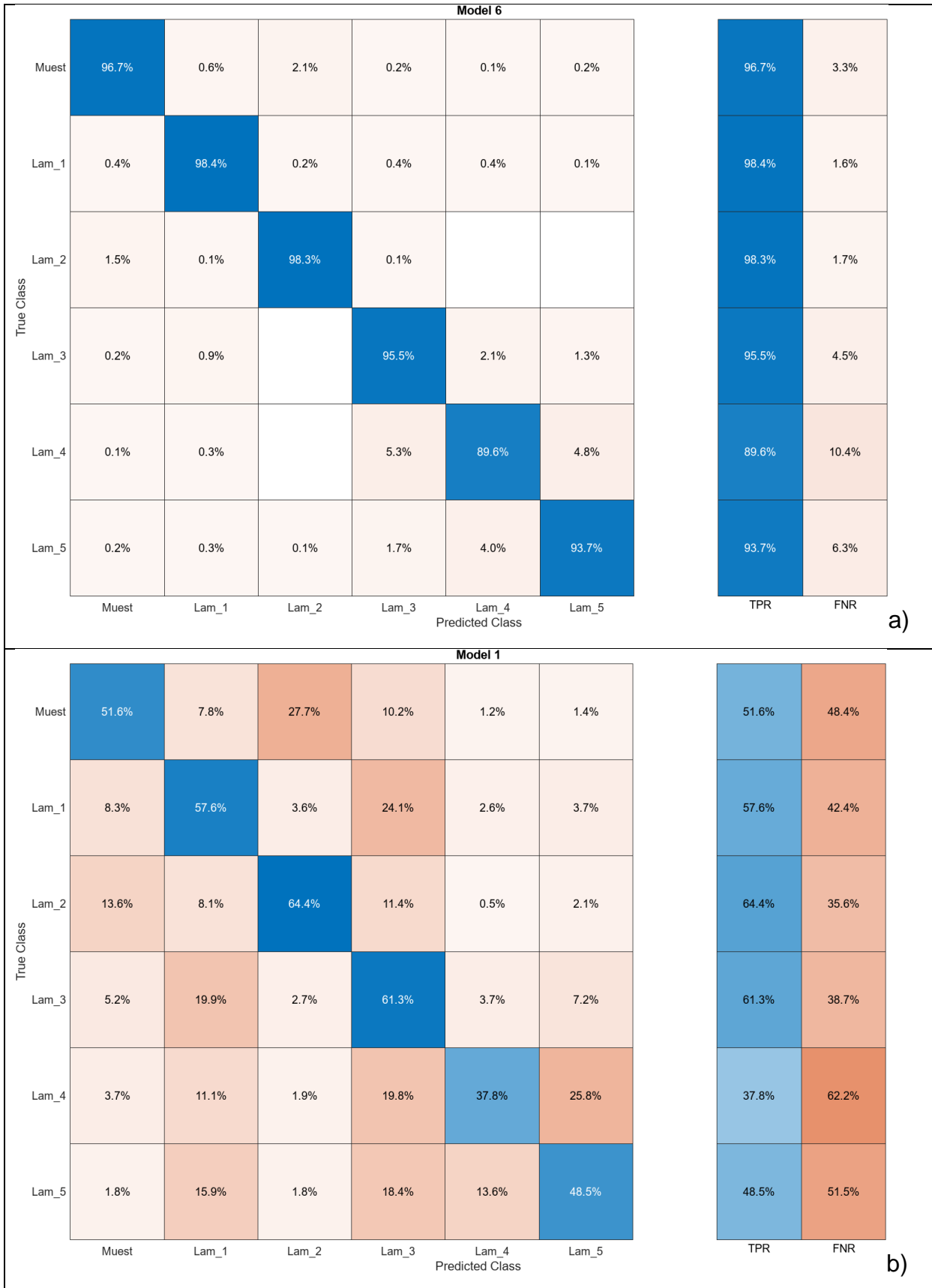


Ilustración 3-22. Matrices de confusión de los datos de test en el aprendizaje automático para los niveles de lambda-cihalotrina en los espectros medios de las cuadrículas extraídas automáticamente: a: Red neuronal; b: k-NN.

El **oxifluorfen** (Tabla 3-17) muestra valores igualmente elevados en la mayor parte de los clasificadores (más del 80%), especialmente en el análisis discriminante y la red neuronal (más del 95%), siendo los árboles de decisión el método que proporciona los resultados más deficientes (60%). Los tiempos de ejecución varían desde 1 a unos 30 minutos.

De forma muy similar a lo que se produce en otros productos en este tipo de muestreo, los valores de exactitud en el análisis de componentes principales se reducen notablemente, al igual que con la selección de características, a pesar de mantener 30 características, aunque de forma no tan acusada como en otros productos.

Las matrices de confusión (Ilustración 3-23) muestran que la clasificación funciona bastante bien en la red neuronal, en prácticamente todos los niveles; mientras que, en la matriz correspondiente a los árboles de decisión, aparece confusión entre los niveles próximos (1 y 2), pero también en algún caso entre niveles alejados (1 y 5 o 2 y 5). En general, salvo en estos casos extremos no representa un problema ya que el nivel 1 y 2 está sobre el nivel, frente al resto que están por debajo.

Modelo	Validac.	Test	Tiempo	Caract.	Técnica	ACP
Árboles Decisión	61,65	59,90	175,62	280/280		
An. Discriminante	96,61	96,04	194,11	280/280		
SVM	89,70	91,17	313,98	280/280		
KNN	81,12	81,53	311,51	280/280		
Bagged / Bosques Al.	82,10	84,21	813,81	280/280		
Red Neuronal	97,70	98,02	800,60	280/280		
Árboles Decisión	48,60	50,16	540,76	280/280		99% Var.
An. Discriminante	40,64	40,10	551,43	280/280		99% Var.
SVM	44,88	44,43	781,44	280/280		99% Var.
KNN	54,93	53,69	544,11	280/280		99% Var.
Bagged / Bosques Al.	57,73	59,42	700,89	280/280		99% Var.
Red Neuronal	53,33	53,80	956,85	280/280		99% Var.
Árboles Decisión	43,83	45,13	623,96	30/280	ANOVA	
An. Discriminante	66,81	65,47	633,34	30/280	ANOVA	
SVM	64,03	64,19	1098,67	30/280	ANOVA	
KNN	53,86	55,30	638,35	30/280	ANOVA	
Bagged / Bosques Al.	57,08	58,51	815,40	30/280	ANOVA	
Red Neuronal	75,06	74,95	1032,92	30/280	ANOVA	

Tabla 3-17. Resultados de la exactitud (accuracy) en validación y test del aprendizaje automático, para los niveles de oxifluorfen en los espectros medios de las cuadrículas extraídas automáticamente.

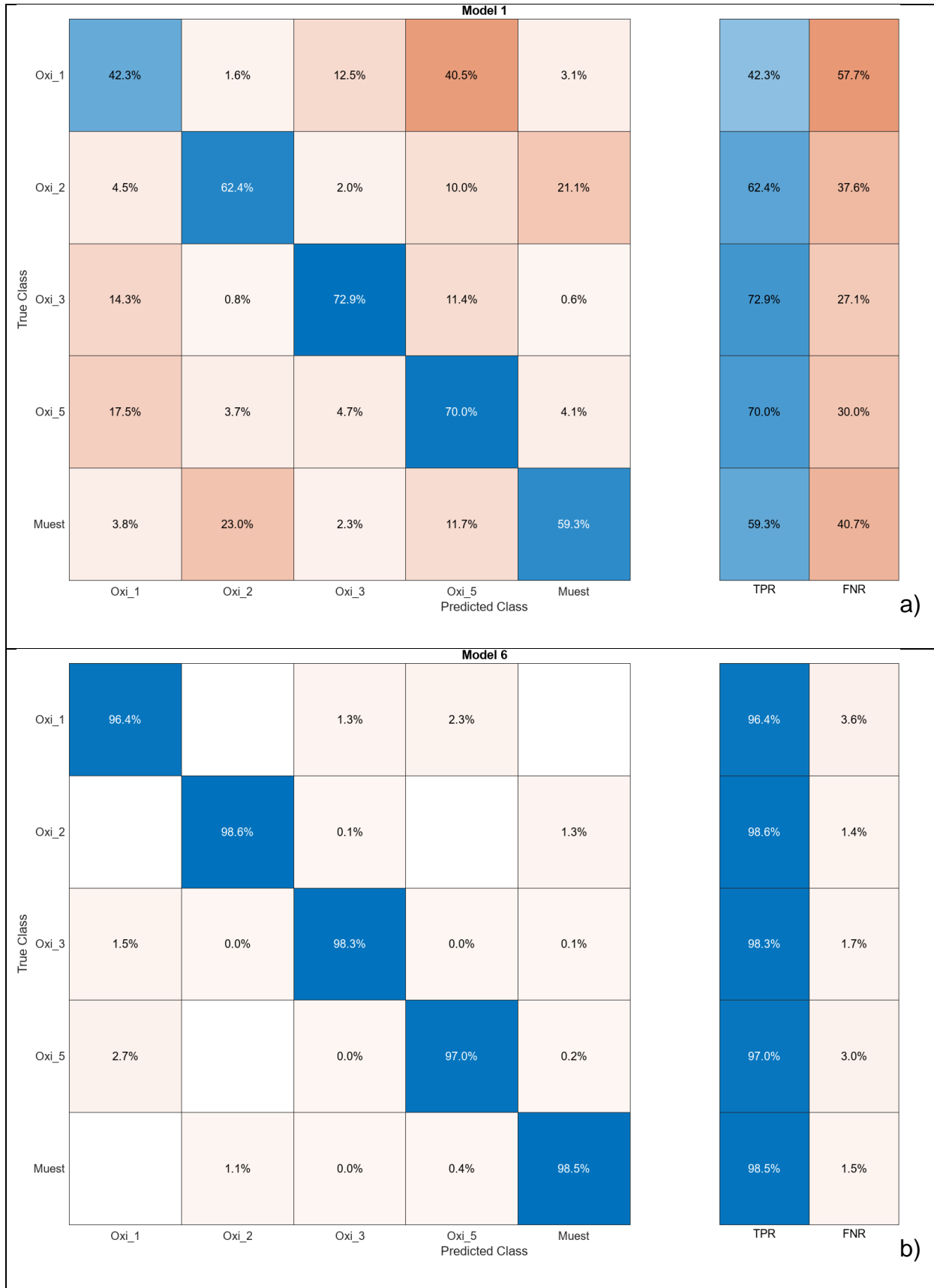


Ilustración 3-23. Matrices de confusión de los datos de test en el aprendizaje automático para los niveles de oxifluorfen en los espectros medios de las cuadrículas extraídas automáticamente: a: Red neuronal; b: Árboles de decisión.

En cuanto al último producto analizado, el **tebuconazol** (Tabla 3-18) presenta valores elevados de exactitud en la validación y el test (por encima del 90%), para el análisis discriminante y la red neuronal; y valores bastante más discretos en el resto, especialmente en los árboles de decisión, que no superan el 50. Los tiempos de ejecución esaután casi siempre entre 1 y 15 minutos.

El análisis de componentes principales, en el que se mantienen 5 componentes al 99% de la varianza, muestra unos valores de exactitud bastante reducidos respecto a los obtenidos con el total del espectro. En el caso de la selección de características (30), se obtienen igualmente valores más bajos aunque no tanto como en el caso anterior.

El análisis de las matrices de confusión (Ilustración 3-24) es bastante simple como en ellos casos anteriores: en la matriz correspondiente a la red neuronal se aprecia la casi perfecta diferenciación entre los niveles de concentración; mientras que la correspondiente a los árboles de decisión muestra una cierta confusión entre los niveles más próximos, pero en general no afectando a la confusión a los niveles por debajo (1-3) y por encima (4-5) del límite.

Modelo	Validac.	Test	Tiempo	Caract.	Técnica	ACP
Árboles Decisión	45,44	44,58	50,74	280/280		
An. Discriminante	93,28	93,50	17,65	280/280		
SVM	79,93	80,55	285,97	280/280		
KNN	69,53	71,61	173,73	280/280		
Bagged / Bosques Al.	70,28	72,14	981,32	280/280		
Red Neuronal	94,40	94,95	333,41	280/280		
Árboles Decisión	43,18	43,05	302,65	280/280		99% Var.
An. Discriminante	32,88	32,45	317,42	280/280		99% Var.
SVM	36,44	36,75	857,73	280/280		99% Var.
KNN	47,97	47,97	352,33	280/280		99% Var.
Bagged / Bosques Al.	51,66	52,98	524,62	280/280		99% Var.
Red Neuronal	44,66	44,16	1003,69	280/280		99% Var.
Árboles Decisión	39,03	39,36	350,55	30/280	ANOVA	
An. Discriminante	75,48	75,54	367,94	30/280	ANOVA	
SVM	69,80	70,61	951,67	30/280	ANOVA	
KNN	49,36	49,25	368,24	30/280	ANOVA	
Bagged / Bosques Al.	55,64	59,31	621,67	30/280	ANOVA	
Red Neuronal	77,14	76,95	823,42	30/280	ANOVA	

Tabla 3-18. Resultados de la exactitud (accuracy) en validación y test del aprendizaje automático, para los niveles de tebuconazol en los espectros medios de las cuadrículas extraídas automáticamente.

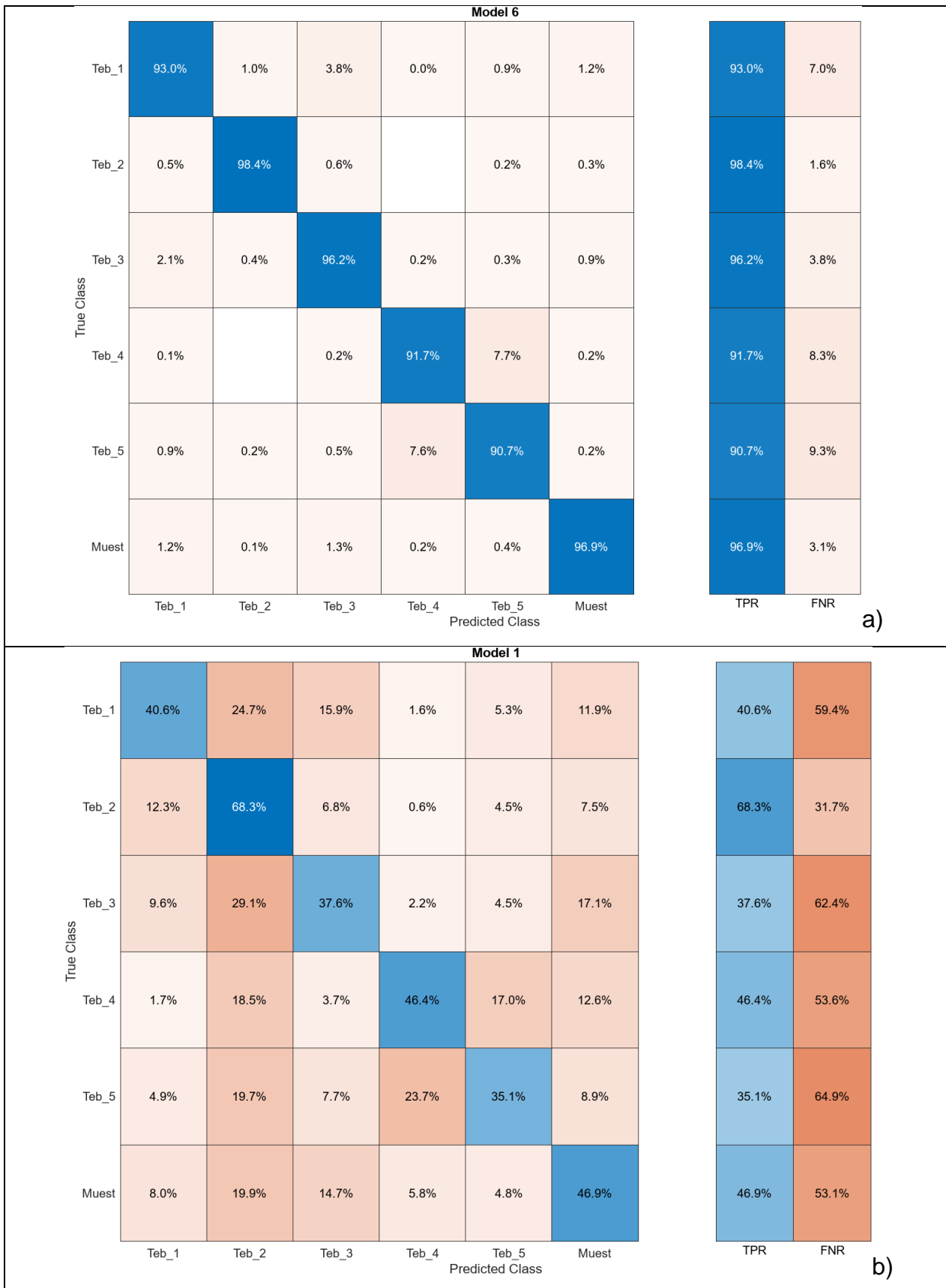


Ilustración 3-24. Matrices de confusión de los datos de test en el aprendizaje automático para los niveles de tebuconazol en los espectros medios de las cuadrículas extraídas automáticamente: a: Red neuronal; b: Árboles de decisión.

Del análisis realizado, se desprenden una serie de observaciones:

- Los métodos de clasificación mediante aprendizaje automático proporcionan en este caso valores satisfactorios y en algunos casos muy elevados. En la clasificación del tipo de fitosanitario se alcanzan valores superiores al 90% en la mayor parte de los factores. Esto se repite en algunos de los niveles de los productos, estando los mejores resultados en la deltametrina y el diflufenican, y los peores en la lambda-cihalotrina y el tebuconazol.
- En este caso, los tiempos de ejecución son más bastante más lentos (por encima del minuto y en algunos casos cercanos a los 10 minutos e incluso a 1 hora). Las muestras son de más de un orden de magnitud mayor que las correspondientes a los espectros medios de las ventanas definidas manualmente y más de dos órdenes respecto a los espectros medios (63800 cuadrículas con sus correspondientes curvas espectrales).
- El mayor tamaño de las muestras en principio permite una mayor estabilidad y significación aún en los cálculos de la exactitud de las muestras, aunque esto no se percibe tan claramente como en el paso anterior, desde el nivel de los valores medios de las aceitunas hasta los valores de las ventanas definidas manualmente.
- En general, los resultados son bastante coherentes con los obtenidos con los espectros medios y los de las ventanas, normalmente mejorando tanto para los tipos de pesticida como para los niveles de los distintos productos.
- En la comparativa entre métodos, algunos proporcionan excelentes resultados (entre el 90 y 100%), como el análisis discriminante, las redes neuronales y en menor medida SVM; mientras que otros presentan resultados peores como los árboles de decisión. Los resultados son similares a los obtenidos por (Martínez-Gila, et al., 2024) para validación y test.
- El análisis de componentes presenta resultados que no mejoran e incluso empeoran sensiblemente la exactitud, a pesar de recurrir a una varianza del 99%, con aparición de entre 4 y 6 CP. Lo mismo sucede, aunque en menor medida, en la selección de características (mediante ANOVA y mantenimiento de 30 características).

3.2.2. Técnicas de aprendizaje profundo

Al igual que en la sección anterior, con vistas a la comparación, se describen primero los resultados obtenidos a partir de las curvas espectrales medias de las aceitunas; en segundo lugar, los correspondientes a las curvas espectrales medias de las ventanas definidas manualmente; y, en tercer lugar, los obtenidos con las cuadrículas de extracción automática.

3.2.2.1. Curvas espectrales medias

En primer lugar, se presentan los resultados globales de la aplicación de la red convolucional sobre los espectros medios de las aceitunas (Tabla 3-19) y posteriormente se pasa a describir en detalle cada uno de los análisis.

Modelo	Entren.	Validac.	Test	Tiempo	Optimiz.
Tipos de producto	98,29	71,19	81,03	34m 46s	-
Niveles deltametrina	96,00	69,23	66,66	10m 15s	-
Niveles diflufenican	100,00	73,33	78,57	6m 30s	-
Niveles lambda-cihal.	100,00	86,67	80,00	6m 14s	-
Niveles oxifluorfen	100,00	83,33	91,67	6m 03s	-
Niveles tebuconazol	100,00	92,86	71,43	10m 06s	-

Tabla 3-19. Resultados de la exactitud (accuracy) en entrenamiento, validación y test de la red convolucional para los tipos y niveles de producto en los espectros medios de las aceitunas.

Analizando tanto los datos de validación como los de test, la aplicación de la red convolucional sobre los espectros medios produce en general unos resultados del mismo orden que los mejores métodos en el aprendizaje automático (análisis discriminante y red neuronal poco profunda). En todos los casos se aprecia además una exactitud para la validación, y especialmente el test, inferior a la obtenida para el de entrenamiento, lo que supone un cierto sobreajuste del modelo. En este caso se ha optado por una partición 50-25-25, por lo que la muestra de test es algo mayor y proporciona resultados más fiables, dentro de que el número de casos es aún bajo.

En el caso de los **tipos de pesticida**, el valor de exactitud en test (81%) está a nivel de los valores de los métodos que producen mejor resultado como el análisis discriminante y la red neuronal. El tiempo de ejecución es de casi 35 minutos, mucho mayor que en el aprendizaje automático. La evolución de las curvas de rendimiento en el entrenamiento y la validación muestra la primera ascendiendo rápidamente y alcanzando pronto un nivel elevado, mientras que la de la validación no llega a valores tan altos y se muestra irregular (Ilustración 3-25). Todo ello es indicativo de sobreajuste.

Las matrices de confusión, a pesar del buen resultado general, muestran cierta confusión del diflufenican que a veces se clasifica como lambda-dihalotrina, aunque el bajo número de casos no permite llegar a conclusiones más claras (Ilustración 3-25).

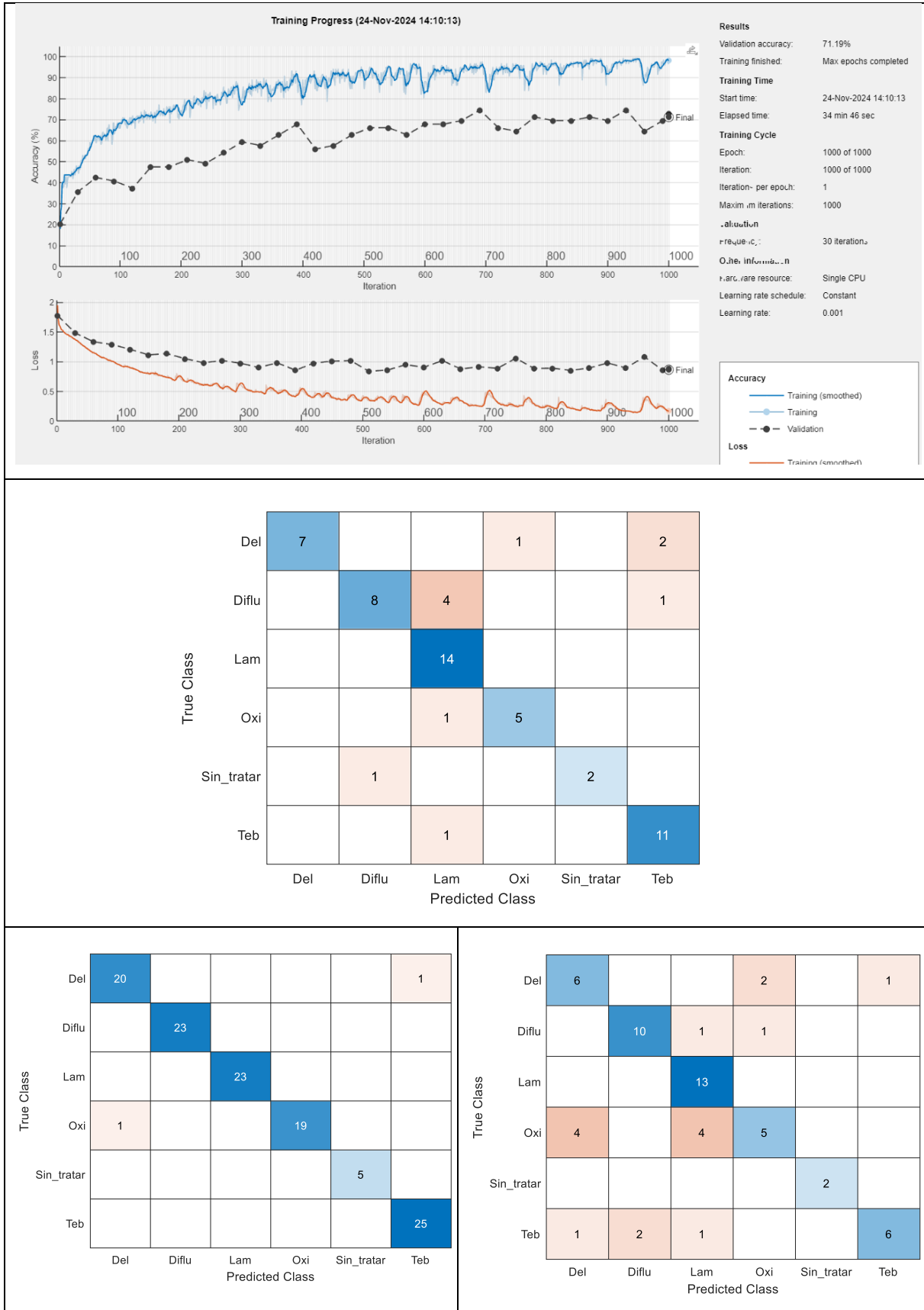


Ilustración 3-25. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para los tipos de producto en los espectros medios de las aceitunas.

Por lo que respecta al análisis por niveles de cada uno de los productos, la **deltametrina** produce unos resultados menos satisfactorios que otros productos, especialmente en el test donde la exactitud está en el 67%. No obstante, este valor es más elevado que la mayor parte de los métodos de aprendizaje automático.

La gráfica de evolución de la exactitud muestra un ascenso aún más rápido que en el caso anterior de los valores para el entrenamiento y una rápida estabilización (Ilustración 3-26); mientras que la validación sube igualmente al principio, aunque en menor medida (30 puntos menos) y luego sigue una trayectoria irregular, pero sin tendencia a aumentar. En este caso, el sobreajuste parece claro, lo que es confirmado por el valor más bajo de la exactitud en el test. El tiempo de ejecución es de unos 10 minutos, sensiblemente superior a los métodos de aprendizaje automático.

Las matrices muestran una cierta confusión entre los niveles medios (2 y 3, ambos por encima del límite europeo), mientras que el resto de los niveles se diferencian más o menos correctamente (Ilustración 3-26). De todas formas, el bajo número de casos, tanto en test como en validación no permite realizar observaciones más concluyentes.

El **diflufenican** presenta unos valores de exactitud en la validación equiparables a los obtenidos en los métodos de aprendizaje automático y algo inferiores en el test, en los que aquellos métodos presentan valores de hasta el 100%, aunque la baja proporción de casos en la muestra de test en ellos conduce a una baja significación de este análisis.

La evolución de las gráficas de exactitud es muy típica de sobreajuste a los datos de entrenamiento, ya que en ellos la curva aumenta muy rápidamente hasta alcanzar el 100% lo que dificulta el proceso de entrenamiento y hace que la curva de validación no evolucione al alza (Ilustración 3-27). El tiempo es algo inferior al de otros productos, en torno a 6 minutos.

Las matrices de confusión muestran que en general los niveles son bien clasificados, aunque existe una cierta mezcla entre los niveles 2 y 3 (Ilustración 3-27), que en este caso está por encima y por debajo del límite europeo. Las muestras sin clasificar aparecen en este caso como bien clasificadas.

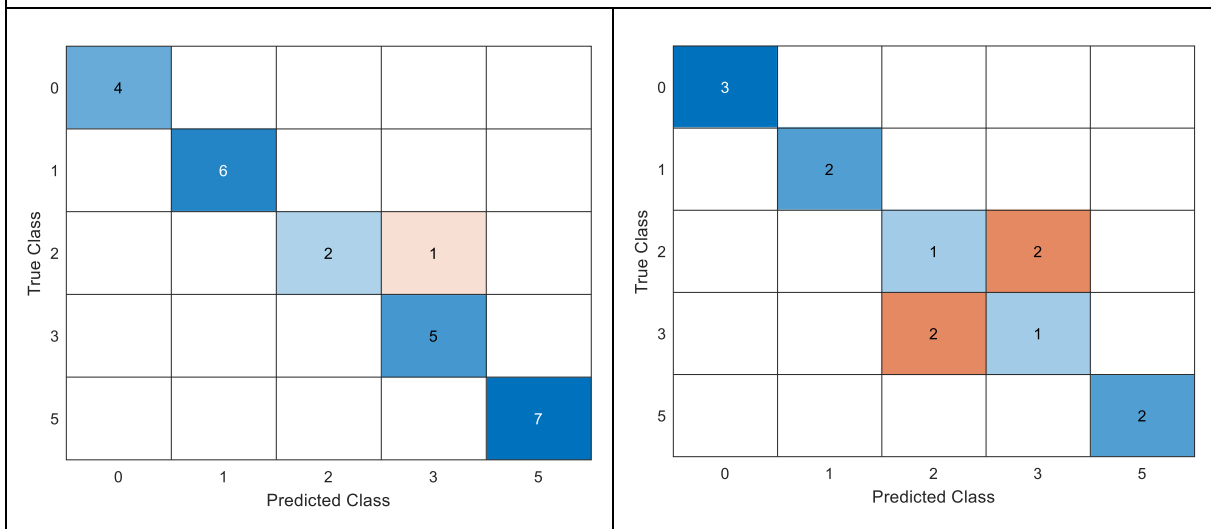
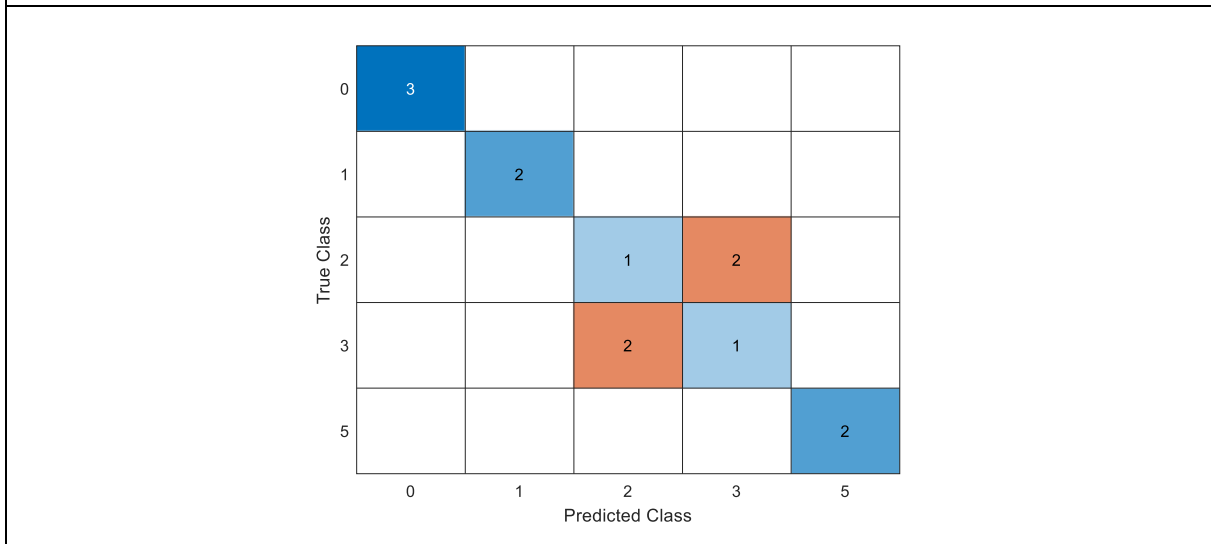
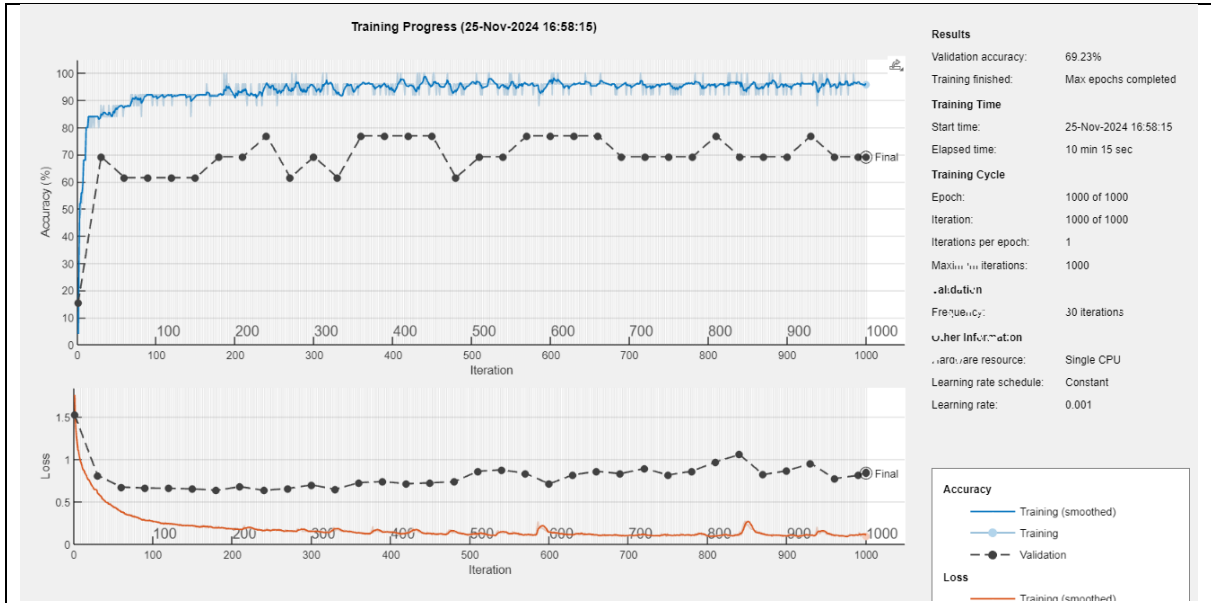


Ilustración 3-26. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para los niveles de deltametrina en los espectros medios de las aceitunas.

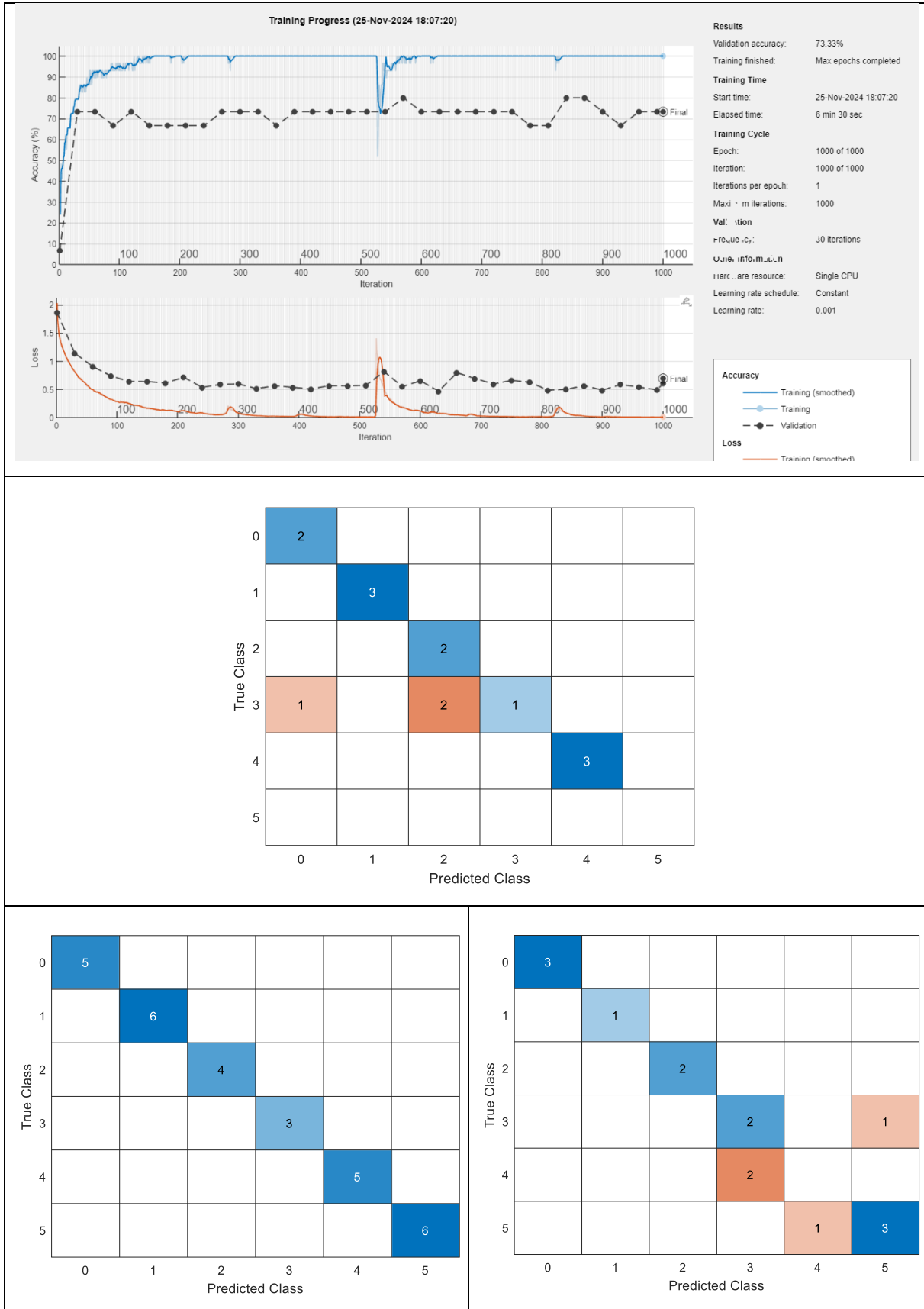


Ilustración 3-27. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para los niveles de diflufenican en los espectros medios de las aceitunas.

Por lo que respecta a la **lambda-cihalotrina**, los valores de exactitud en la validación y el test se encuentran al mismo nivel que los mejores métodos de aprendizaje automático (análisis discriminante y SVM), por encima del 80%.

La evolución de las gráficas de exactitud es igualmente muy típica de sobreajuste a los datos de entrenamiento ya que en ellos la curva aumentan muy rápidamente hasta alcanzar el 100% mientras la curva de validación no evoluciona al aumento (Ilustración 3-28) y se estanca a partir de un momento determinado. El tiempo de ejecución está también en torno a 6 minutos, de los más bajos en este análisis.

Las matrices de confusión muestran que en general los niveles son bien clasificados, aunque existe una cierta mezcla entre los niveles 2 y 3 (Ilustración 3-28), ambos por encima del límite europeo y entre los niveles 4 y 5, ambos por debajo. Las muestras sin tratar aparecen también bien clasificadas.

El **oxifluorfen** muestran valores de exactitud en la validación y el test muy elevados (90%), incluso por encima de los calculados en los métodos de aprendizaje automático, a pesar de la escasa significación de la muestra de test.

La evolución de las gráficas de exactitud vuelve a ser muy típica de sobreajuste a los datos de entrenamiento con un aumento muy rápido de la curva correspondiente hasta alcanzar el 100%, mientras la curva de validación en este caso sí se mantiene alta, pero muy plana en el 90% (Ilustración 3-29). El tiempo de ejecución está también en torno a 6 minutos, de los más bajos en este análisis.

Las matrices de confusión muestran que en general los niveles son bien clasificados, con pequeñas confusiones poco significativas (Ilustración 3-29). Las muestras sin tratar están también bien clasificadas.

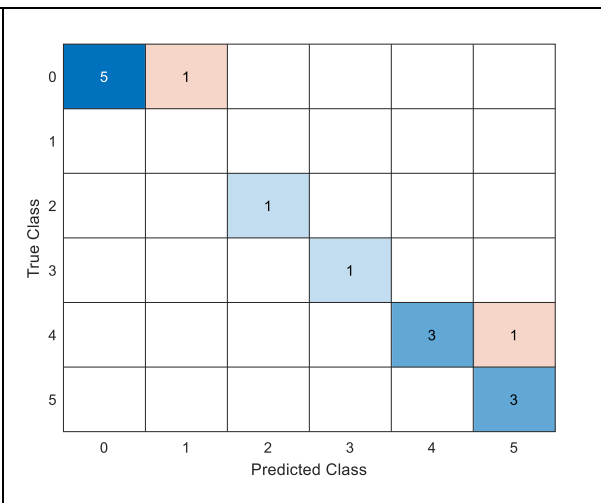
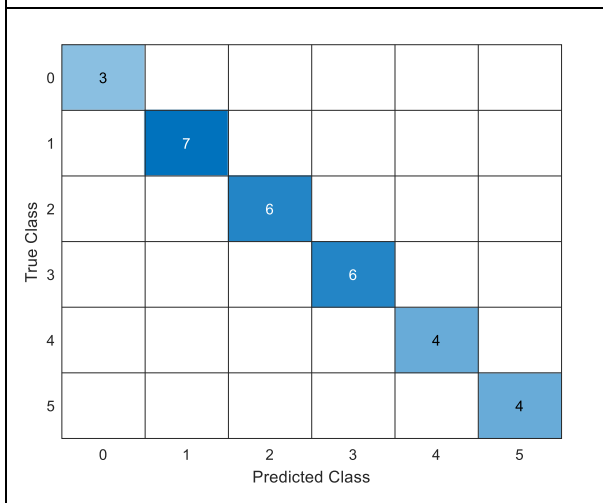
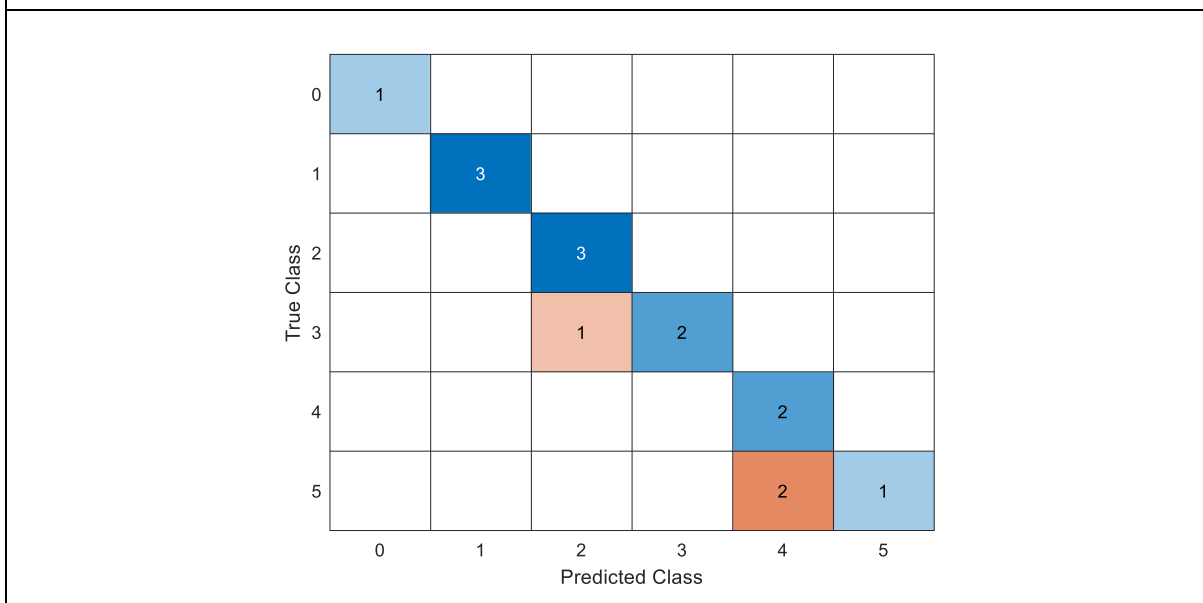
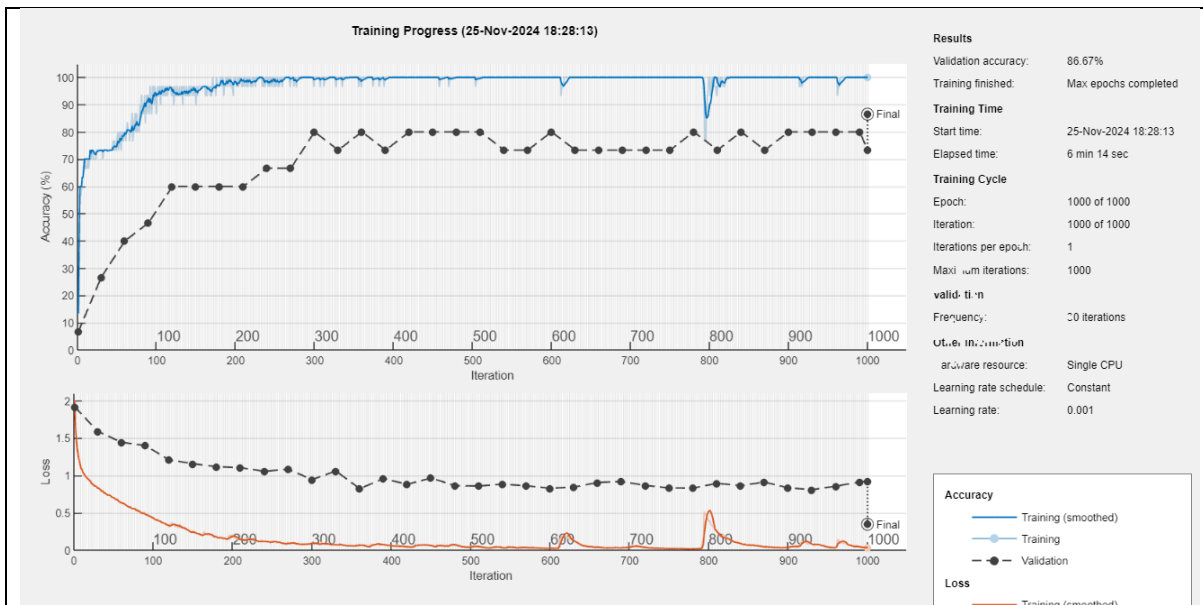


Ilustración 3-28. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para los niveles de lambda-cihalotrina en los espectros medios de las aceitunas.

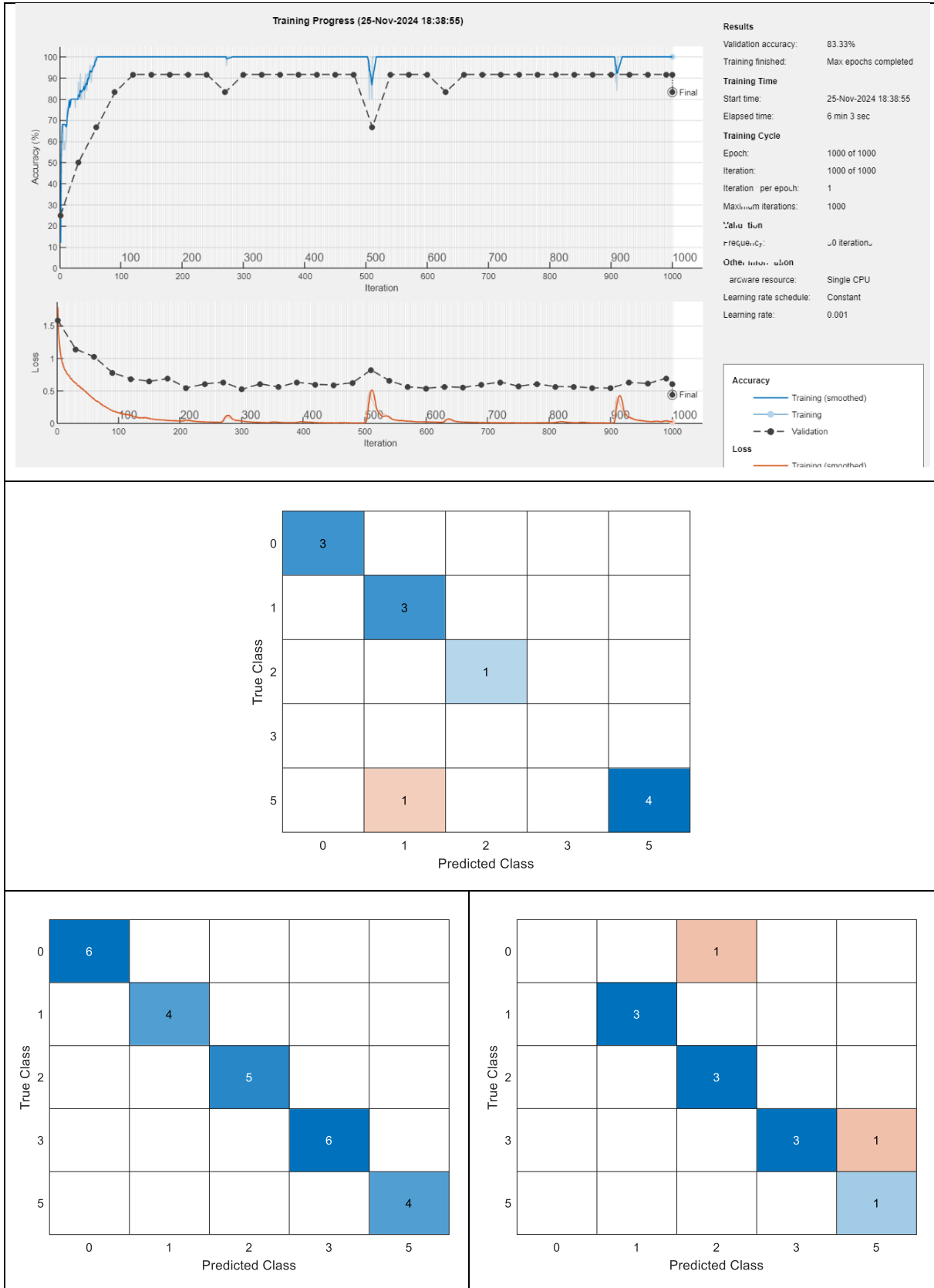


Ilustración 3-29. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para los niveles de oxifluorfen en los espectros medios de las aceitunas.

Por último, el **tebuconazol** muestra valores de exactitud equiparables o incluso superiores a los métodos de aprendizaje automático más satisfactorios en validación (93%), no así en test donde desciende al 71%, por debajo de lo que alcanzan dichos métodos. Esto es indicativo de sobreajuste.

La gráfica de evolución de la exactitud (Ilustración 3-30) es igualmente indicativa de un proceso de sobreajuste, corroborando lo anteriormente dicho con aumento rápido de las curvas, en este caso también la de validación, y estancamiento hasta el final. El tiempo de ejecución es algo mayor que algunos de los anteriores, en torno a 10 minutos.

Las matrices muestran alguna confusión entre la muestra sin tratar y valores bajos y medios (3 y 5) de concentración, casi siempre por debajo del límite europeo, aun dentro de una aceptable clasificación general (Ilustración 3-30 Ilustración 3-29).

De todo el análisis de los resultados de las curvas espectrales medias mediante la red convolucional, se pueden extraer las siguientes observaciones:

- El desarrollo de una red neuronal convolucional de 1D y su aplicación sobre los espectros medios de las imágenes de las aceitunas proporciona en general buenos resultados, con valores de exactitud equiparables y en algunos casos superiores a los obtenidos con los métodos de aprendizaje automático, incluso con los que muestran los mejores rendimientos.
- No obstante, los resultados son mejores con los datos de validación que con los de test, lo que se debe por un lado a la escasez de casos en la muestra de test, pero también puede ser indicativo de un sobreajuste de la red.
- Los mejores resultados se producen en el oxifluorfen y los peores en la deltametrina y el tebuconazol.
- La evolución de la gráfica de exactitud en entrenamiento y test corrobora la sospecha del sobreajuste, en especial en algunos de los análisis por niveles de pesticida, por cuanto la curva de entrenamiento asciende muy rápidamente desde el principio hasta casi el 100% y luego se estanca, lo que dificulta el entrenamiento. Esto hace que la curva de validación, que también muestra un rápido ascenso, luego no evolucione adecuadamente y tenga en

algunos casos un comportamiento irregular; y sobre todo que los valores de test no alcancen un valor elevado.

- El tiempo de ejecución es del orden de los minutos, mientras que en los métodos de aprendizaje era de segundos, aunque eso no es un problema demasiado importante, ya que una vez entrenada una red su aplicación es mucho más rápida.
- Las matrices de confusión permiten en líneas generales comprobar la buena clasificación de las muestras, tanto en el caso de los tipos de producto como sobre todo en los niveles de los mismos; aunque nuevamente hay que decir que las pequeñas dimensiones de las muestras hacen que el proceso de entrenamiento y aprendizaje no sea el más de adecuado y produzca problemas de inestabilidad en la red de tal forma que el distintos procesos de entrenamiento sobre la misma muestra den resultados demasiado divergentes, lo que resta significación al método en este caso.
- Todo ello conduce a la necesidad de aumentar las muestras de entrenamiento, validación y test, con el objetivo de realizar un mejor proceso de entrenamiento, lo que se va a analizar en el apartado siguiente. Por otra parte, se han aplicado técnicas de optimización de los parámetros de la red, que en este caso no producen ninguna mejora.

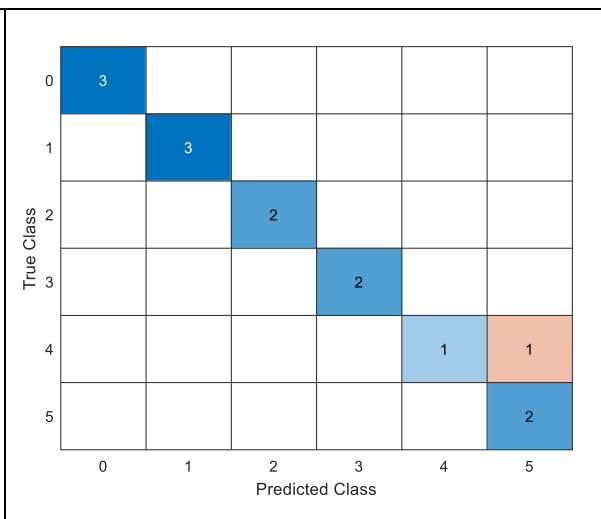
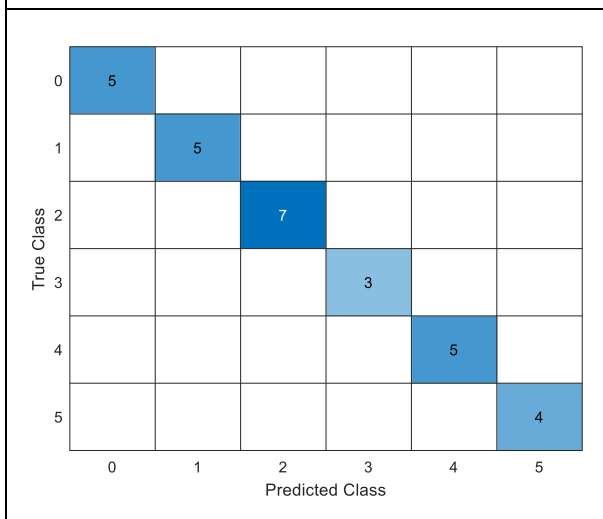
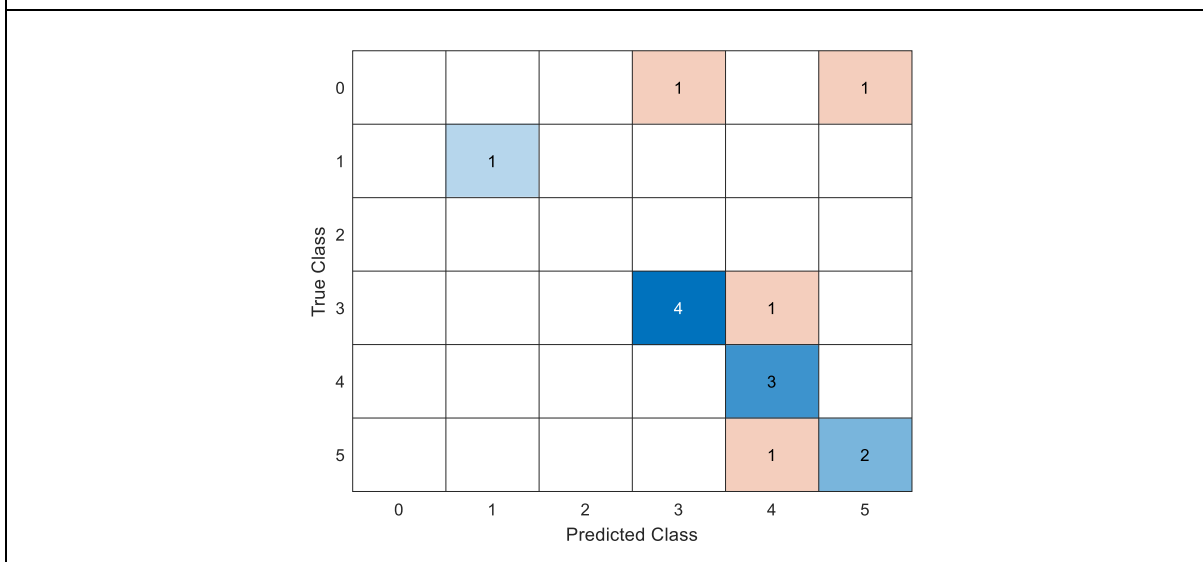
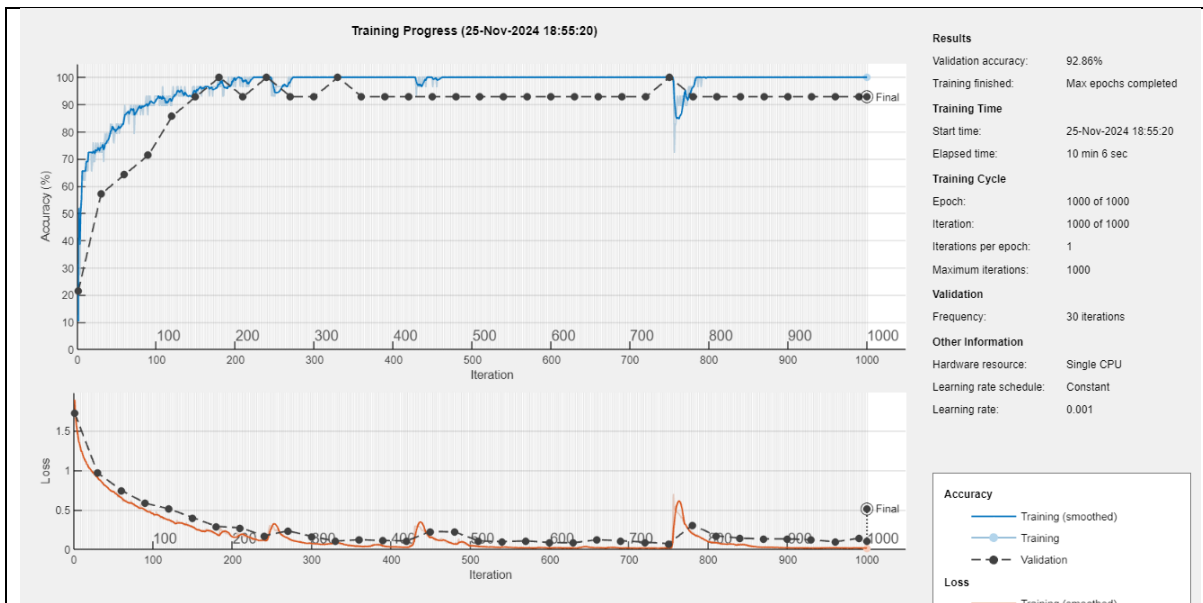


Ilustración 3-30. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para los niveles de tebuconazol en los espectros medios de las aceitunas.

3.2.2.2. Curvas espectrales extraídas de las ventanas manualmente

Siguiendo el mismo esquema que anteriormente, en primer lugar, se presentan los resultados globales de la aplicación de la red convolucional sobre los espectros medios (Tabla 3-20) y posteriormente se pasa a describir en detalle cada uno de los análisis, incluyendo el rendimiento de la red. En este caso, además del análisis por los tipos de pesticidas y por los niveles de concentración de los productos, se ha analizado el efecto de introducir una optimización bayesiana, que se ha descrito en la metodología, en aquellos casos en que el ajuste manual de las redes convolucionales no haya surtido los efectos deseados.

Modelo	Entren.	Validac.	Test	Tiempo	Optimiz.
Tipos de producto	65,32	62,56	60,01	46m 05s	-
Tipos de producto	87,48	86,32	85,18	34m 04s	Bayesiana
Niveles deltametrina	98,14	92,00	90,67	50m 14s	-
Niveles diflufenican	98,14	92,00	90,67	27m 51s	-
Niveles lambda-cihal.	80,71	72,09	70,20	-	-
Niveles lambda-cihal.	95,47	90,00	92,22	41m 35s	Bayesiana
Niveles oxifluorfen	96,94	93,24	87,67	39m 11s	-
Niveles tebuconazol	88,97	83,72	67,05	-	-
Niveles tebuconazol	86,59	80,23	80,64	31m 39s	Bay (10)
Niveles tebuconazol	79,69	81,39	71,76	-	Bay (30)

Tabla 3-20. Resultados de la exactitud (accuracy) en entrenamiento, validación y test de la red convolucional para los tipos y niveles de producto en los espectros medios de .

Analizando tanto los resultados, la aplicación de la red convolucional sobre los espectros medios de las ventanas definidas manualmente produce en general unos resultados del mismo orden que los mejores métodos en el aprendizaje automático (análisis discriminante y red neuronal poco profunda) e incluso mejorando en algunos casos. A diferencia de lo que sucedía en los espectros medios de las aceitunas, los valores de exactitud de validación y test son similares o ligeramente inferiores a los de entrenamiento, lo que indica que no se produce sobreajuste de acuerdo a este criterio. En este caso, puesto que el tamaño de la muestra se considera suficiente, se ha trabajado con una partición 70/15/15, que suele ser más común en este tipo de estudios (Torres, 2020).

Entrando ya en el detalle, el primer análisis considerado es el de los **tipos de pesticida**, que había proporcionado resultados aceptables en alguno de los métodos de aprendizaje automático, particularmente el análisis discriminante y la red neuronal. Sin embargo, el resultado no fue positivo, obteniéndose una exactitud inferior al 60% de exactitud en la clasificación para los datos de test, a pesar de haber aplicado una optimización manual de los hiperparámetros basada en las redes utilizadas previamente. Esto indicaba que este enfoque no era adecuado para abordar la complejidad del problema, como se ha explicado en la metodología. Así, ante la baja precisión lograda y la dificultad de ajustar los hiperparámetros se optó por la optimización bayesiana, que ofrecía un enfoque más sistemático y efectivo para maximizar el rendimiento del modelo. Finalmente, se ha obtenido una red 1D con 4 capas convolucionales, 343 filtros por capa y un tamaño de filtro de 13 (ver apartado de metodología).

Tras la optimización bayesiana se alcanza un resultado bastante satisfactorio (cercano al 86% en validación y test), prácticamente al nivel del análisis discriminante y la red neuronal, y superior a todos los demás métodos considerados. El tiempo de ejecución es de 34 minutos, mucho mayor que en el aprendizaje automático. La evolución de las curvas de rendimiento en el entrenamiento y la validación muestra en este caso un comportamiento adecuado, ya que el ascenso no es tan rápido (Ilustración 3-31), lo que permite que las dos curvas vayan evolucionando progresivamente y a la par, dando lugar a valores de exactitud final de la validación y el test bastante satisfactorios.

Por su parte, las matrices de confusión, a pesar del buen resultado general, muestra una cierta confusión de la deltametrina, que se clasifica en muchas ocasiones como otros productos, especialmente como lambda-dihalotrina (Ilustración 3-31). En este caso, el adecuado volumen de las muestras permite establecer este tipo de conclusiones.

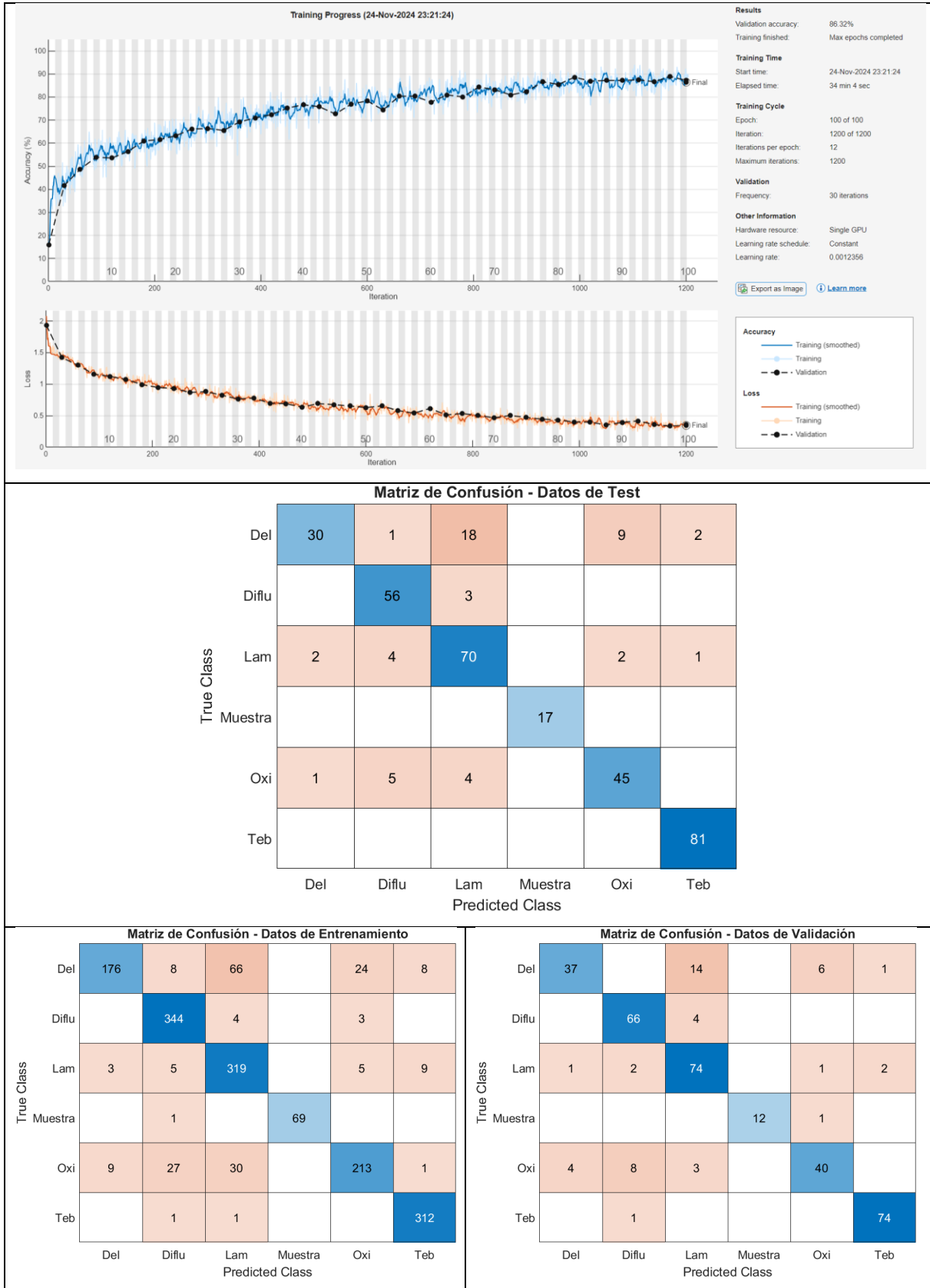


Ilustración 3-31. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para el tipo de producto en los espectros medios de las ventanas definidas manualmente.

En la **deltametrina** se observan valores de exactitud muy elevados (91%) superiores en unos 10 puntos a los obtenidos con los mejores métodos de aprendizaje automático. En general los resultados están muy equilibrados para los distintos tipos de muestras, lo que indica un buen proceso de entrenamiento.

Así se constata en las gráficas de evolución de la exactitud, donde tras el crecimiento inicial de los valores, se observa una tendencia positiva en ambas que permiten alcanzar altos valores de exactitud al final del proceso (Ilustración 3-32). El tiempo de ejecución es de unos 50 minutos.

Todo ello, se refleja en las matrices de confusión, donde pese a los pequeños errores, se puede determinar que la clasificación que realiza la red es prácticamente perfecta, por lo que se da por válida. La principal confusión se produce entre las clases 3 y 5, las de concentración más baja de la deltametrina y relativamente contiguas, aunque en este caso el nivel 5 está bajo el límite europeo y la 3 por encima del mismo (Ilustración 3-32).

Por su parte, en el **diflufenican** se observa nuevamente que el porcentaje de acierto en la clasificación es alto para los tres conjuntos de datos, superando el 90% para los datos de test, superior en unos 10 puntos a los valores alcanzados en los métodos más eficaces del aprendizaje automático.

La evolución de la exactitud con el entrenamiento igualmente es muy positiva y sostenida a medida que avanza el proceso, estando las dos curvas muy próximas (Ilustración 3-33). Todo ello es indicativo de un buen proceso de entrenamiento sin que se observe sobreajuste. El tiempo de ejecución es de unos 28 minutos, de los más bajos de los obtenidos con este conjunto de datos.

A partir de las matrices, se puede determinar que la principal confusión se produce entre la concentración 4 y 5 (Ilustración 3-33), lo que en principio no presenta un problema puesto que ambos están por debajo del límite europeo; además como se observa en la matriz la concentración 4 es mínima y la 5 casi inexistente. Por lo que se puede determinar que la clasificación es casi perfecta para las concentraciones en este pesticida.

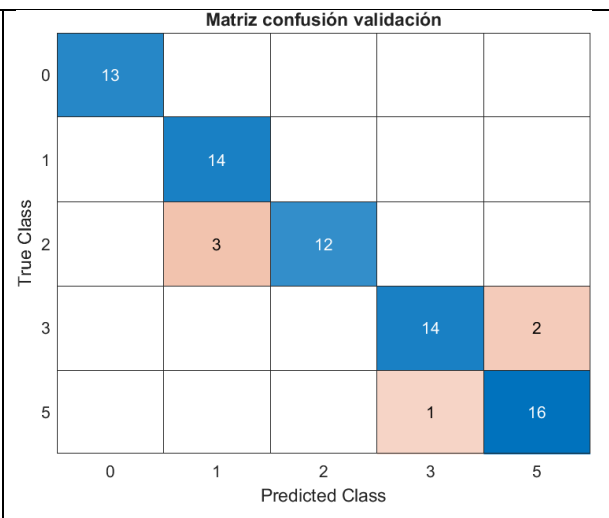
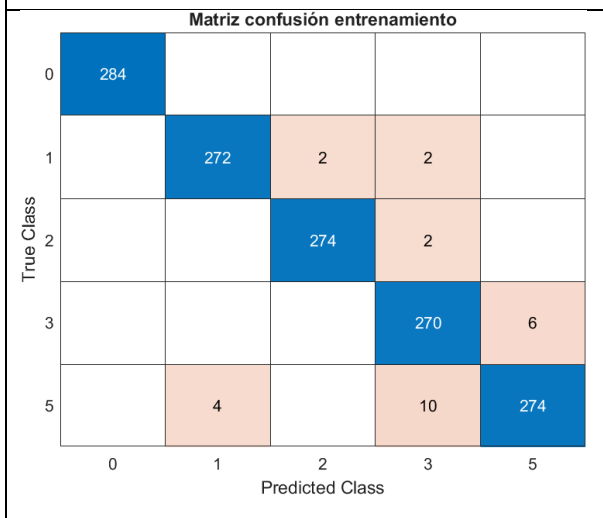
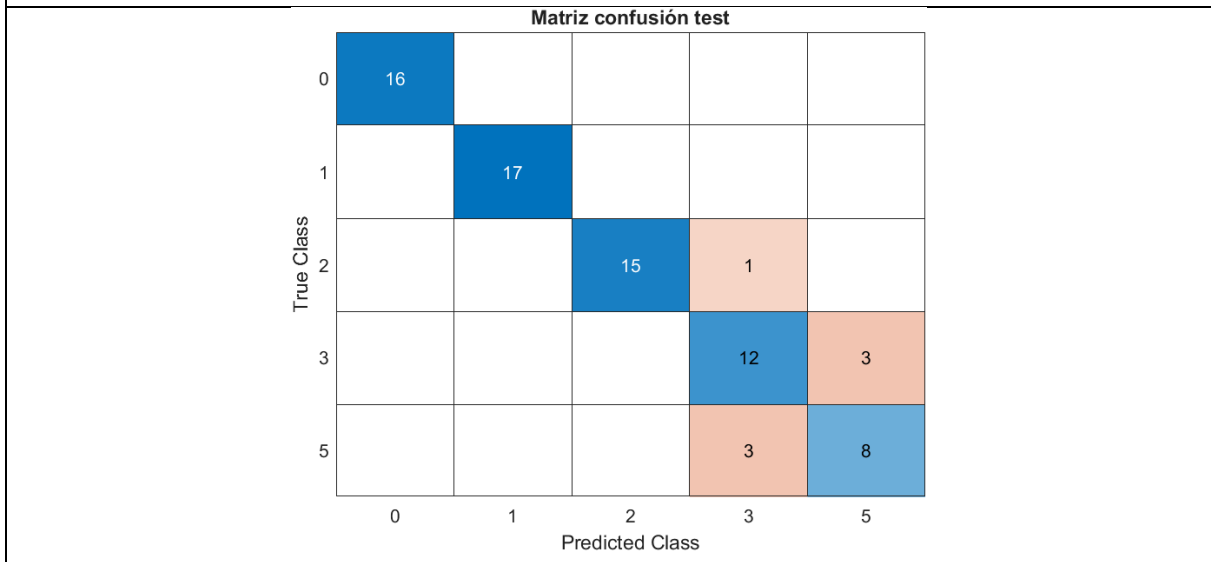
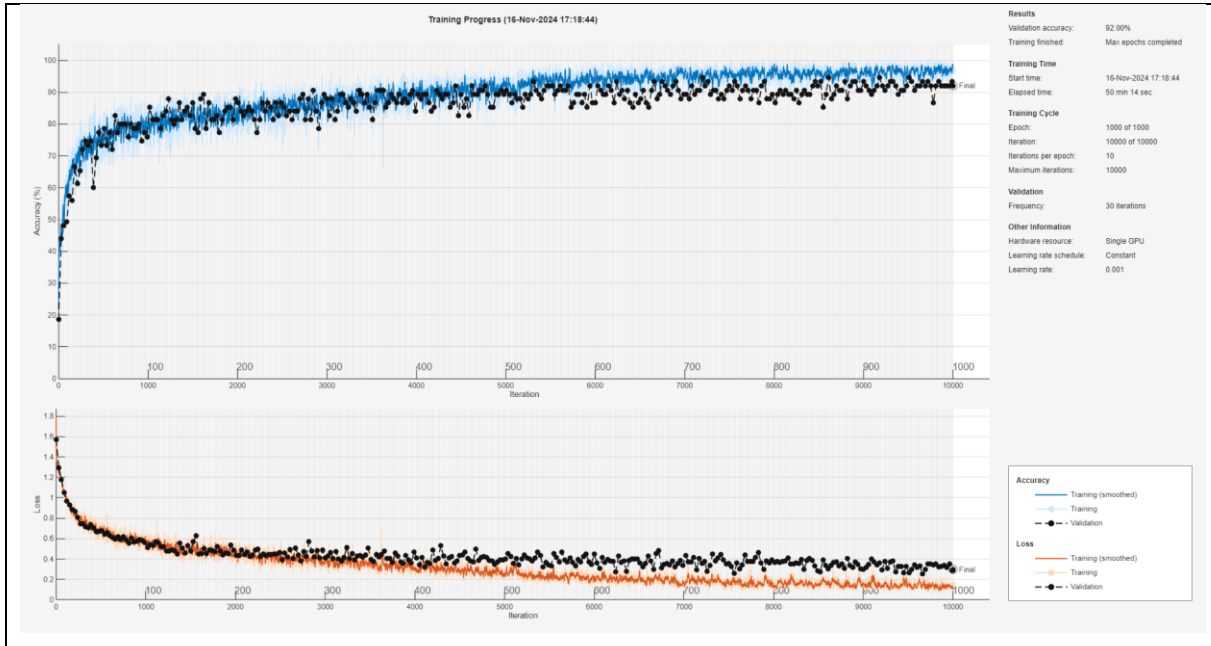


Ilustración 3-32. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para el nivel de deltametrina en los espectros medios de las ventanas definidas manualmente.

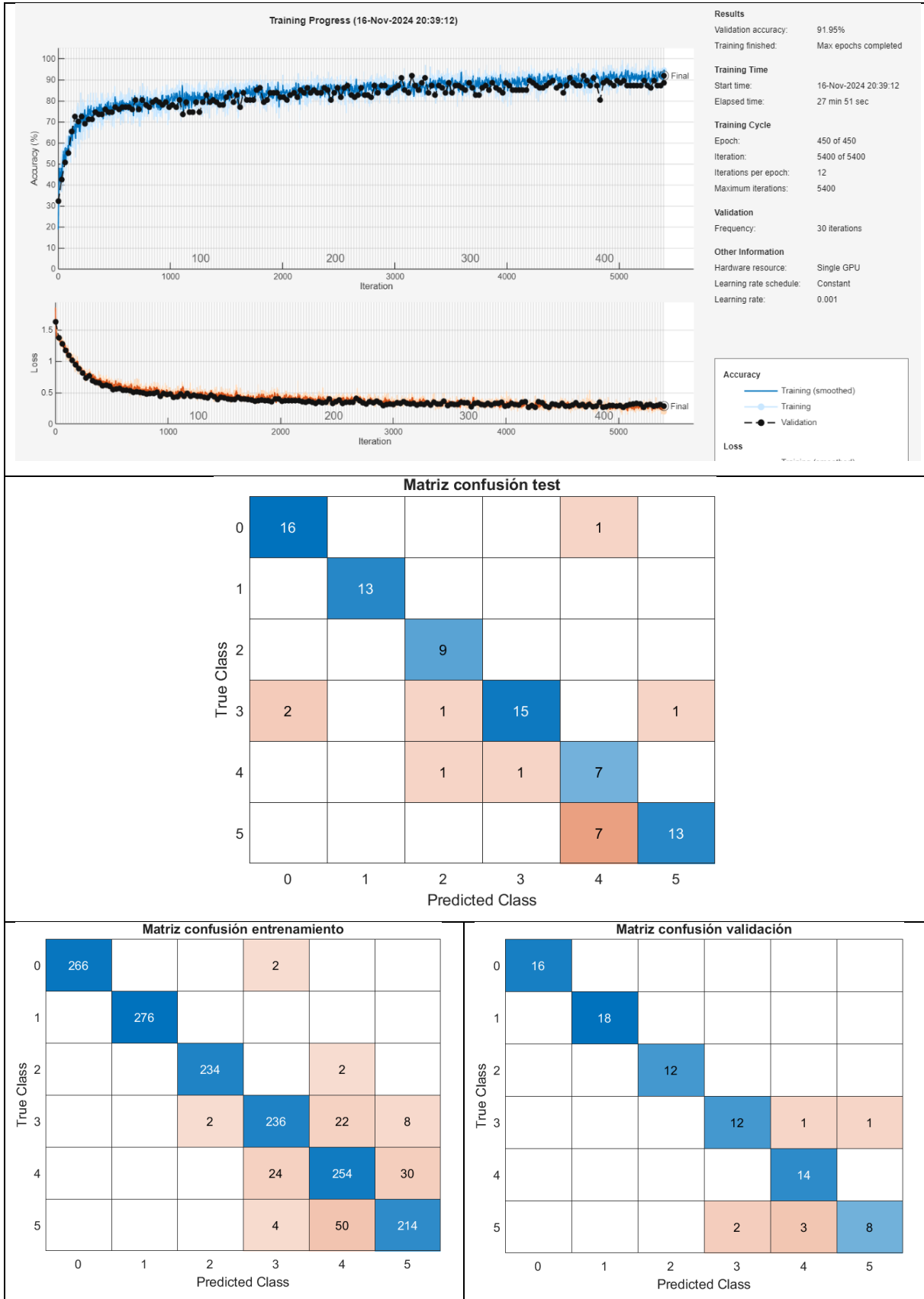


Ilustración 3-33. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para el nivel de diflufenican en los espectros medios de las ventanas definidas manualmente.

La **lambda-cihalotrina** proporciona resultados inicialmente peores que los productos anteriores, en torno al 70%, aunque aún superiores a los obtenidos con los métodos de aprendizaje automático. Para intentar mejorar este resultado se hicieron distintas pruebas cambiando los hiperparámetros que, si bien mejoraban los resultados en entrenamiento, producían una red muy sobreajustada, por lo que se desecharon. Sin embargo, aplicando la optimización bayesiana según se describe en el apartado de metodología, se obtienen unos resultados muy positivos. Así, se observa que tanto los datos de validación como de test aumentan la exactitud entre un 15 y un 20%, llegando a una exactitud superior al 90% en validación y test, bastante mejor que en los otros métodos.

Las gráficas de exactitud muestran una buena evolución, pero presentan cierta irregularidad, así como una pronunciada oscilación en la curva de validación en determinados tramos, aunque no tanto como en algunas de las anteriormente descritas, así como separación entre ambas curvas (Ilustración 3-34).

Observando la matriz de confusión de test, se puede concluir que la mayor confusión ocurre entre las concentraciones 4 y 5, que corresponden a las concentraciones menores, por debajo del límite (Ilustración 3-34). También existe algo de confusión entre la clase 3 y la 4 y la 5, estando la 3 por encima del límite europeo, lo que podría generar algún problema.

En el caso del **oxifluorfen**, igualmente se obtienen excelentes resultados, siendo la exactitud de validación y entrenamiento superior al 93% y el test superior al 87%, superior a los métodos que mejores resultados dan en el análisis discriminante (análisis discriminante y SVM).

La evolución de las gráficas de exactitud es muy correcta con un ascenso progresivo y coordinado de ambas, tras el rápido aumento inicial, lo que permite alcanzar los buenos resultados indicados. El tiempo de ejecución es algo mayor en este caso, de unos 39 minutos (Ilustración 3-35).

Se observa que existe una pequeña confusión entre las clases, en general de la 5 con el resto, pero esta es mínima, estando la mayor parte de las muestras clasificadas correctamente (Ilustración 3-35).

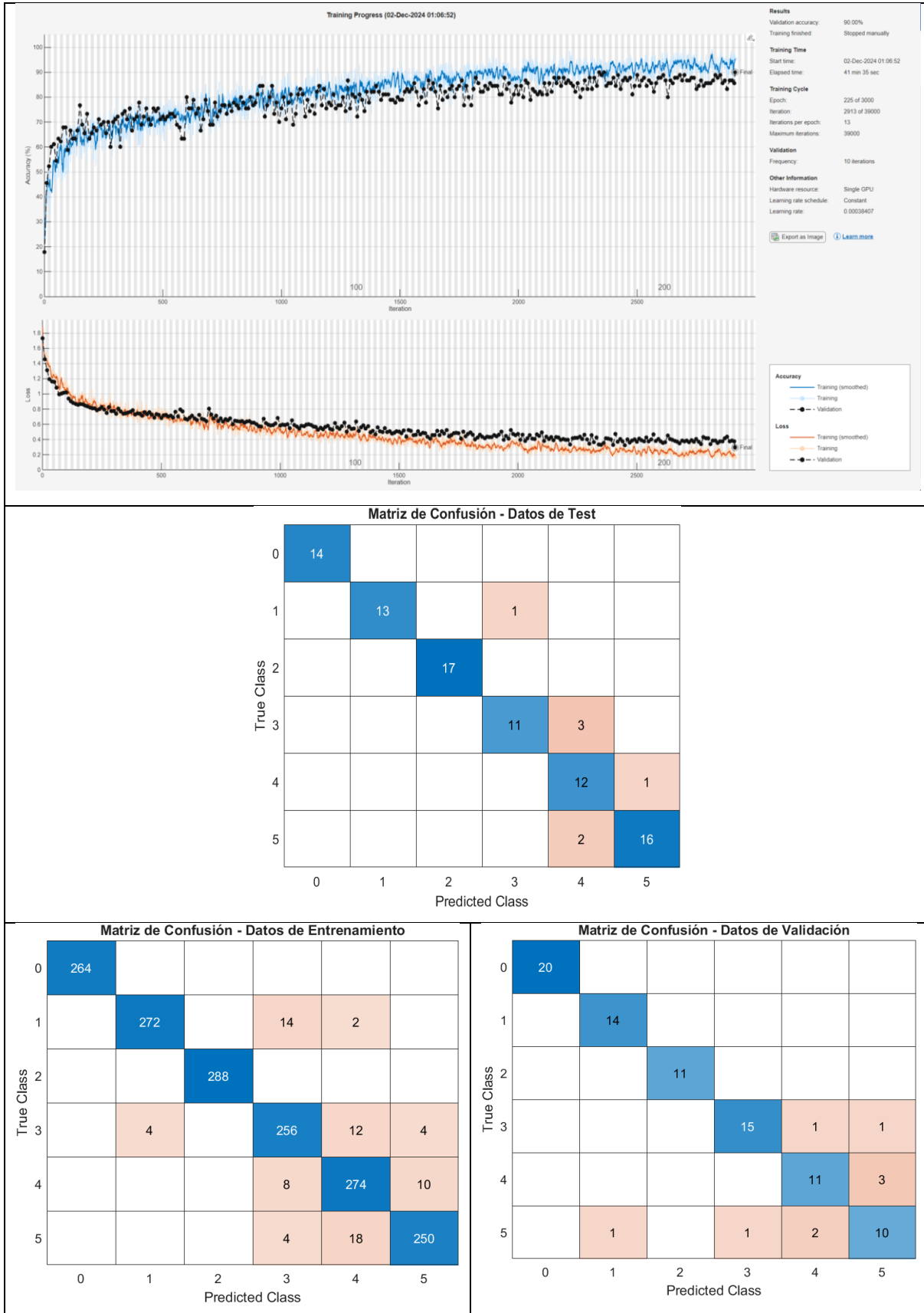


Ilustración 3-34. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para el nivel de lambda-cihalotrina en los espectros medios de las ventanas definidas manualmente.

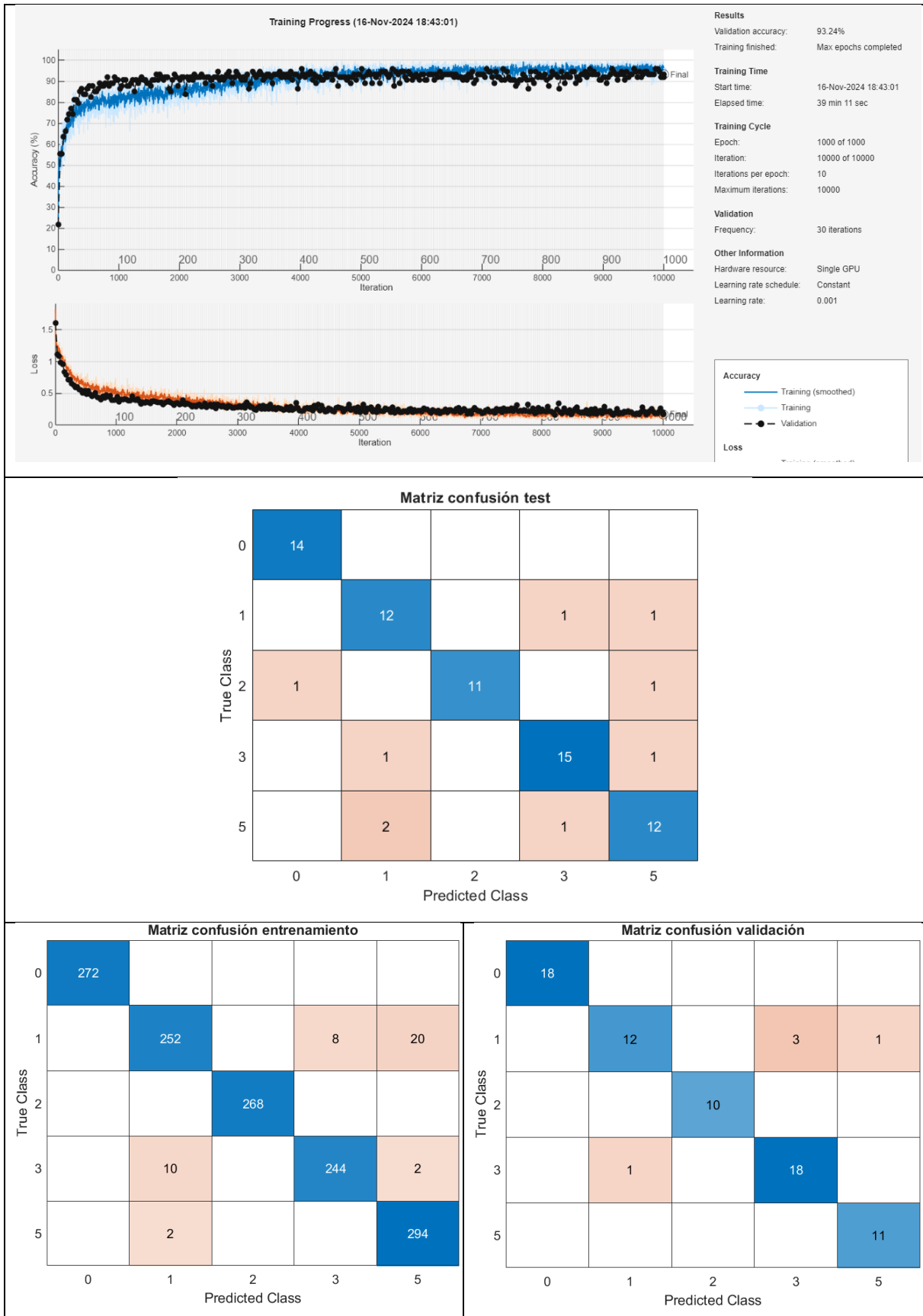


Ilustración 3-35. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para el nivel de oxifluorfen en los espectros medios de las ventanas definidas manualmente.

En el caso de tebuconazol, inicialmente los resultados no eran satisfactorios con una exactitud en el test del 67%, incluso algo inferiores a los obtenidos con algunos de los métodos de aprendizaje automático como el análisis discriminante o la red neuronal. Sin embargo, tras aplicar la optimización bayesiana, los valores aumentan hasta en 10 puntos, situándose en el mismo rango que los métodos comentados. Se ha probado la optimización con un número mayor de redes, pero no proporciona mejores resultados, así que se deja la anterior (ver apartado de metodología).

La gráfica de progresión de entrenamiento exactitudes presenta un aspecto aceptable, con un cierto estancamiento en torno al 80% después de un crecimiento exponencial, pero de forma coherente entre ambas curvas, lo que indica que no hay problema importante de sobreajuste. El proceso de entrenamiento no es demasiado satisfactorio ya que las curvas tienen una diferencia en torno al 6% (Ilustración 3-36).

A la vista de las matrices, la confusión se produce en este caso principalmente entre los niveles bajos de concentración (4 y 5) o entre los más altos (1 a 3), lo que no supone demasiado problema, ya que el límite europeo está entre 3 y 4, por lo que cada grupo (bajo o sobre el límite) queda bien diferenciado del otro (Ilustración 3-36).

De la aplicación del aprendizaje profundo (red convolucional 1D) llevado a cabo con los espectros medios de las ventanas definidas manualmente, se pueden derivar las siguientes observaciones:

- Los resultados obtenidos para la exactitud en validación y test son en general muy buenos y en algunos casos excelentes, con valores que en la mayor parte de los casos son superiores a los obtenidos con los mejores métodos de aprendizaje automático (análisis discriminante, SVM y red neuronal). Se alcanzan valores superiores al 90% en los niveles de deltametrina, diflufenican, oxifluorfen y lambda-cihalotrina (esta, tras la optimización); del 85% para los tipos; y cercanos al 80% en el caso más desfavorable (tebuconazol), y tras aplicar métodos de optimización bayesiana de las redes. En general y con matices, los resultados son mejores en los herbicidas (diflufenican y oxifluorfen) que en los insecticidas (deltametrina y lambda-cihalotrina) y sobre todo que en el fungicida (tebuconazol).

- A diferencia de lo que ocurría con los espectros medios, los resultados de exactitud con entrenamiento y con validación y test son bastante coherentes, lo que se debe al aumento de las muestras que permiten un mejor entrenamiento y testeo de la misma. Esto previene del sobreajuste.
- La evolución de la gráfica de exactitud en entrenamiento y test corrobora lo dicho, observándose un crecimiento sostenido y armónico de las curvas, sin estancamientos, ni grandes irregularidades. Este crecimiento se refleja en los valores obtenidos finalmente en los valores de exactitud final, aunque no es igual para todos los productos analizados.
- El tiempo de ejecución es del orden de los minutos, entre 20 y 50, superior a los obtenidos en los espectros medios y mucho más que en los métodos de aprendizaje. Este aparente problema no es demasiado grande, ya que una vez entrenada una red su aplicación es mucho más rápida.
- Las matrices de confusión permiten en líneas generales comprobar la buena clasificación de las muestras, tanto en el caso de los tipos de pesticida como sobre todo en los niveles de los mismos. En general, permite diferenciar bien las muestras con niveles superiores e inferiores al límite europeo de concentración de cada uno de estos productos, salvo algunas excepciones.
- En este caso, el tamaño de las muestras se considera adecuado, lo que permite un adecuado proceso de entrenamiento, así como una buena validación y testeo. Además, la estabilidad de la red permite que la repetición del proceso de entrenamiento produzca resultados fiables, e incluso que vaya mejorando a medida que la red va siendo entrenada con más datos.
- Por esta razón y por el interés de disponer de una metodología lo más automatizada posible se ha realizado un tercer tipo de muestreo a través de cuadrículas definidas automáticamente, con la extracción de una gran muestra de datos.

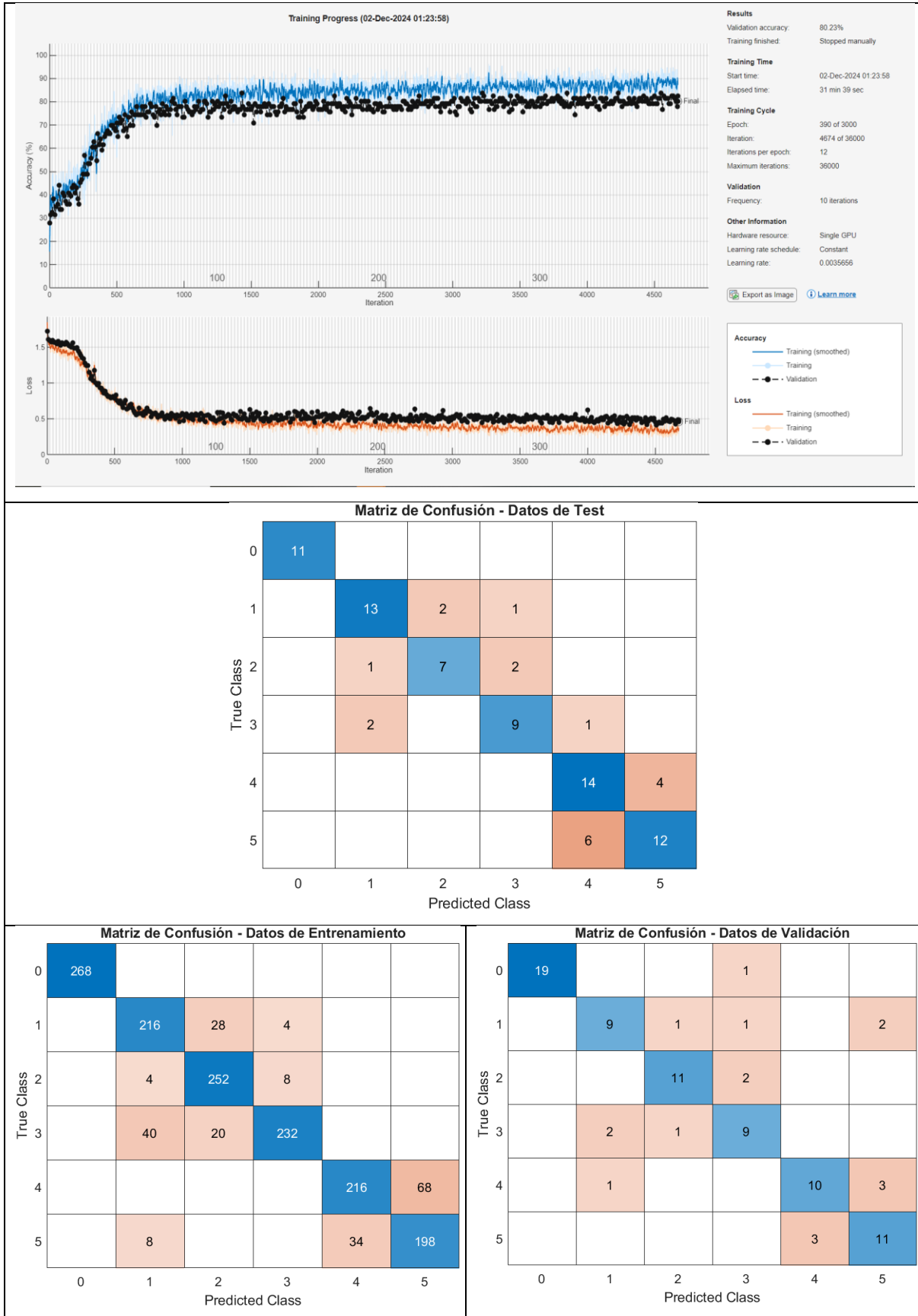


Ilustración 3-36. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para el nivel de tebuconazol en los espectros medios de las ventanas definidas manualmente.

3.2.2.3. Curvas espectrales extraídas de las cuadrículas automáticamente

Siguiendo el mismo esquema que en los dos casos anteriores, en primer lugar, se presentan los resultados globales de la aplicación de la red convolucional sobre los espectros medios (Tabla 3-21) y posteriormente se pasa a describir en detalle cada uno de los análisis, incluyendo el rendimiento de la red.

Modelo	Entren.	Validac.	Test	Tiempo	Optimiz.
Tipos de producto	91,19	91,32	90,60	622m 03s	Bayesiana
Niveles deltametrina	92,79	93,04	92,71	136m 02s	-
Niveles diflufenican	90,48	91,20	90,39	117m 00s	-
Niveles lambda-cihal.	96,25	95,43	95,23	144m 21s	Bayesiana
Niveles oxifluorfen	88,10	88,98	89,10	54m 48s	-
Niveles tebuconazol	80,93	80,53	79,94	54m 28s	Bayesiana

Tabla 3-21. Resultados de la exactitud (accuracy) en entrenamiento, validación y test de la red convolucional para los tipos y niveles de producto en los espectros medios de las cuadrículas extraídas automáticamente.

Analizando los resultados, la aplicación de la red convolucional sobre los espectros medios de las cuadrículas que permiten la extracción automática de los espectros produce en general unos buenos resultados de la exactitud, siempre por encima del 90%; aunque en este muestreo masivo, los métodos de aprendizaje llegan a alcanzar valores incluso superiores (más del 95% en métodos como el análisis discriminante, la red neuronal e incluso SVM). En este caso, los valores de exactitud de validación y test son similares a los de entrenamiento, lo que indica que no se produce sobreajuste. La partición ha sido como en el caso anterior de 70/30/30.

Entrando ya en el detalle, el primer análisis considerado es el de los **tipos de pesticida**, para lo que se empleó ya directamente la optimización bayesiana, como ya se hizo en el muestreo de las ventanas definidas manualmente. Así, se alcanza una exactitud satisfactoria (91%), algo por debajo de lo que se produce el análisis discriminante y la red neuronal que alcanzan el 98 y 99% respectivamente.

La forma del gráfico de exactitud es indicativa de un buen entrenamiento ya que, tras el impulso inicial, ambas curvas aumentan de forma progresiva y coherente (Ilustración 3-37). El tiempo de ejecución es de varias horas, aunque este dato puede no resultar significativo, ya que el ordenador trabajó en este caso en multitarea.

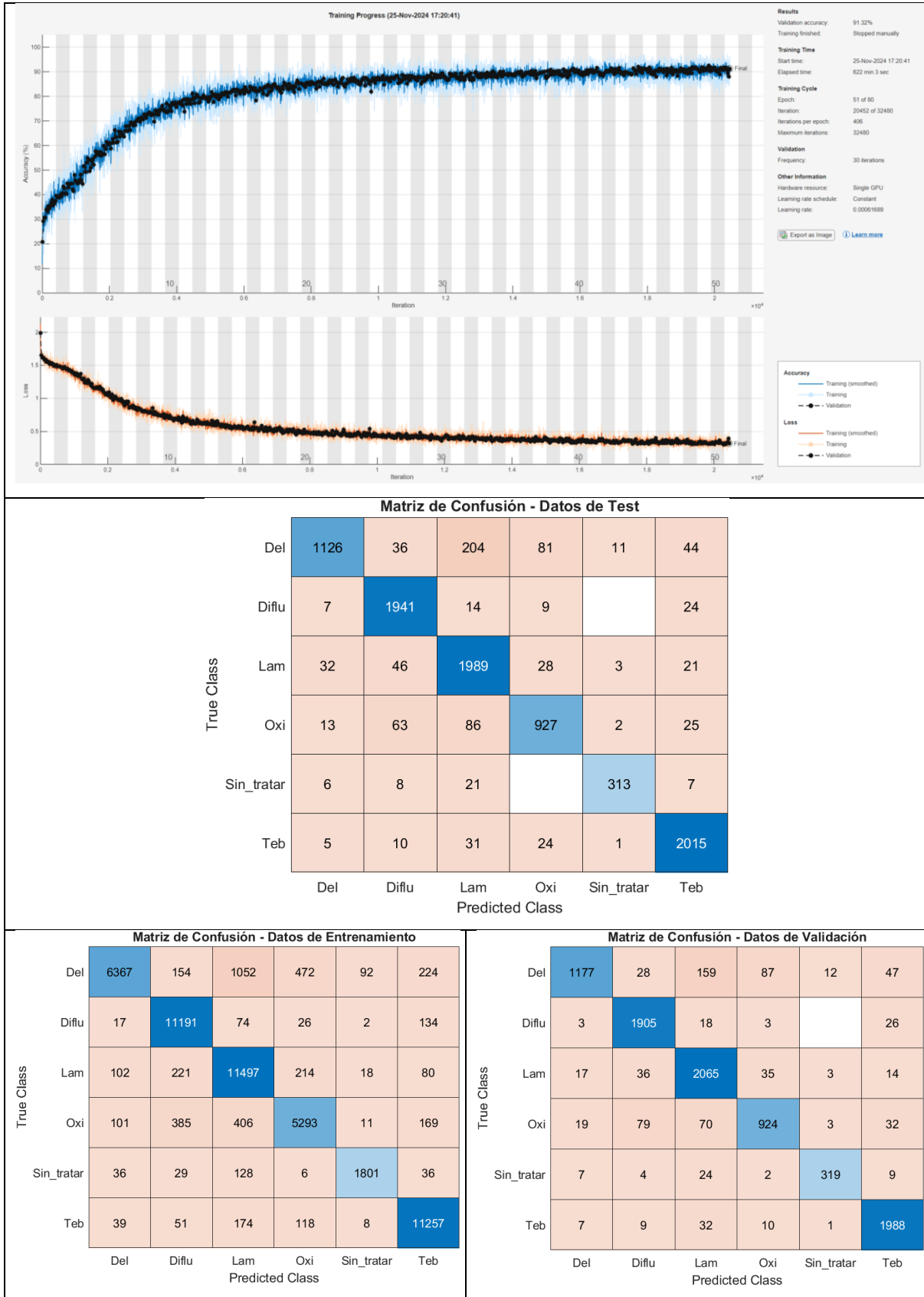


Ilustración 3-37. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para el tipo de producto en los espectros medios de las cuadrículas extraídas automáticamente.

Por su parte, las matrices de confusión, a pesar del buen resultado general, muestra una cierta confusión de la deltametrina, que se clasifica en algunas ocasiones como otros productos, especialmente como lambda-dihalotrina (Ilustración 3-37), algo que ya se había observado en los otros muestreos.

En la **deltametrina** se observan valores de exactitud muy elevados (93%), ligeramente inferiores a los que se obtienen en algunos métodos de aprendizaje automático. En general los resultados están muy equilibrados para los distintos tipos de muestras, lo que indica un buen proceso de entrenamiento.

Las gráficas de evolución de la exactitud muestran un buen aspecto con una tendencia creciente en ambas, tras el incremento inicial, lo que permiten alcanzar altos valores de exactitud al final del proceso (Ilustración 3-38). El tiempo de ejecución es de poco más de 2 horas.

Las matrices de confusión muestran una buena clasificación general, aunque se observa una cierta confusión en las muestras sin tratar (Ilustración 3-38), que aparecen en ocasiones clasificadas como de nivel de concentración 2, lo que puede ocasionar algún problema ya que las clases 1, 2 y 3 están sobre el límite europeo.

Por su parte, en el **diflufenican** se observa nuevamente que la exactitud en la clasificación es alta para los tres conjuntos de datos, superando el 90% para los datos de test. Como en el producto anterior, siendo un resultado muy bueno, es inferior a los obtenidos con algunos de los métodos de aprendizaje profundo como el análisis discriminante y la red neuronal, y similar a SVM.

La evolución de la exactitud con el entrenamiento igualmente es muy positiva y sostenida a medida que avanza el proceso, estando las dos curvas muy próximas (Ilustración 3-39), sin que por tanto se observe sobreajuste. El tiempo de ejecución es de algo menos de 2 horas.

Las matrices de confusión muestran la buena clasificación general de las muestras (Ilustración 3-39), y la confusión, aunque limitada, se produce entre las clases 3 a 5, lo cual no origina ningún problema ya que todas estas clases están por debajo del límite europeo de concentración de los niveles de este pesticida.

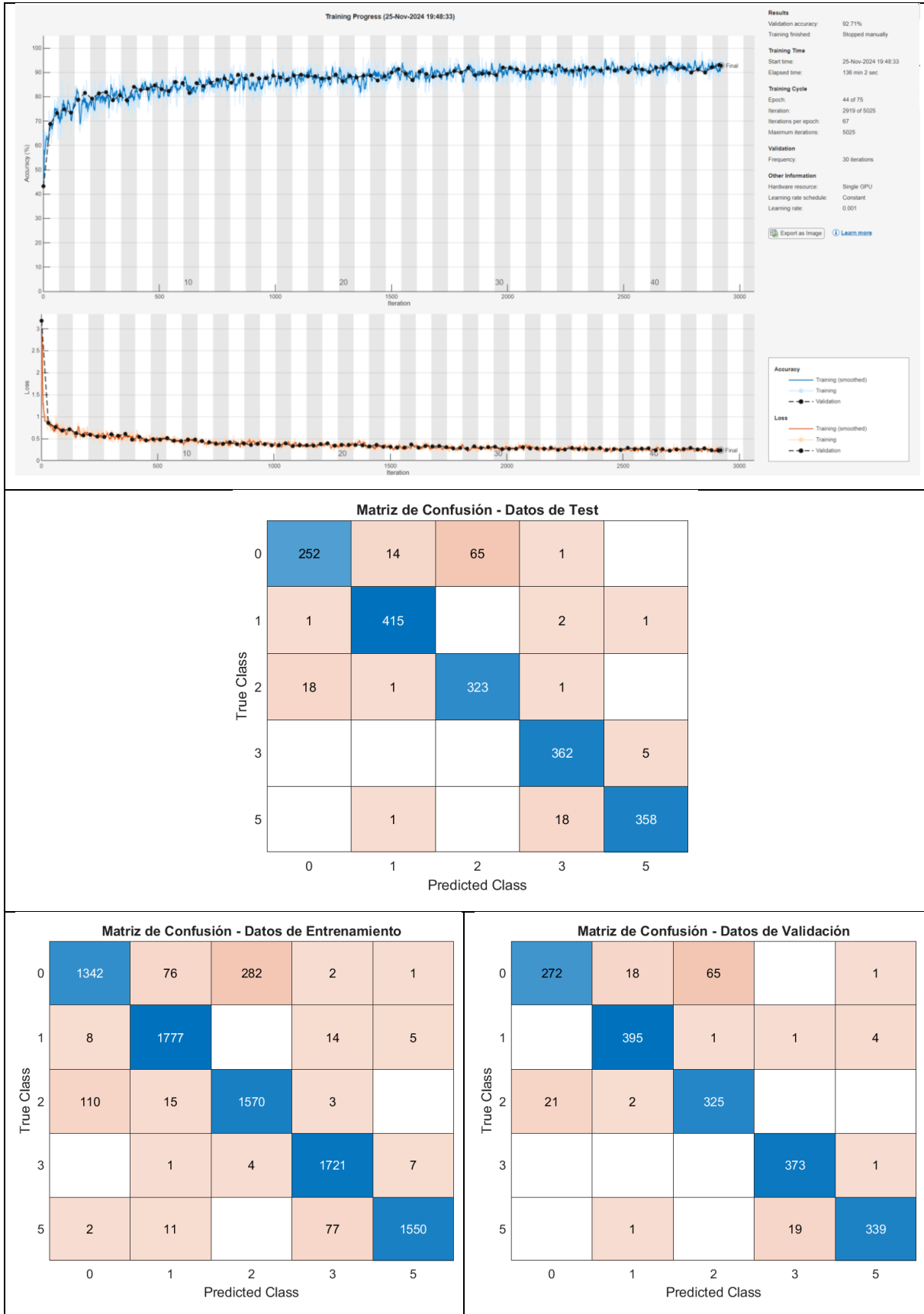


Ilustración 3-38. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para el nivel de deltametrina en los espectros medios de las cuadrículas extraídas automáticamente.

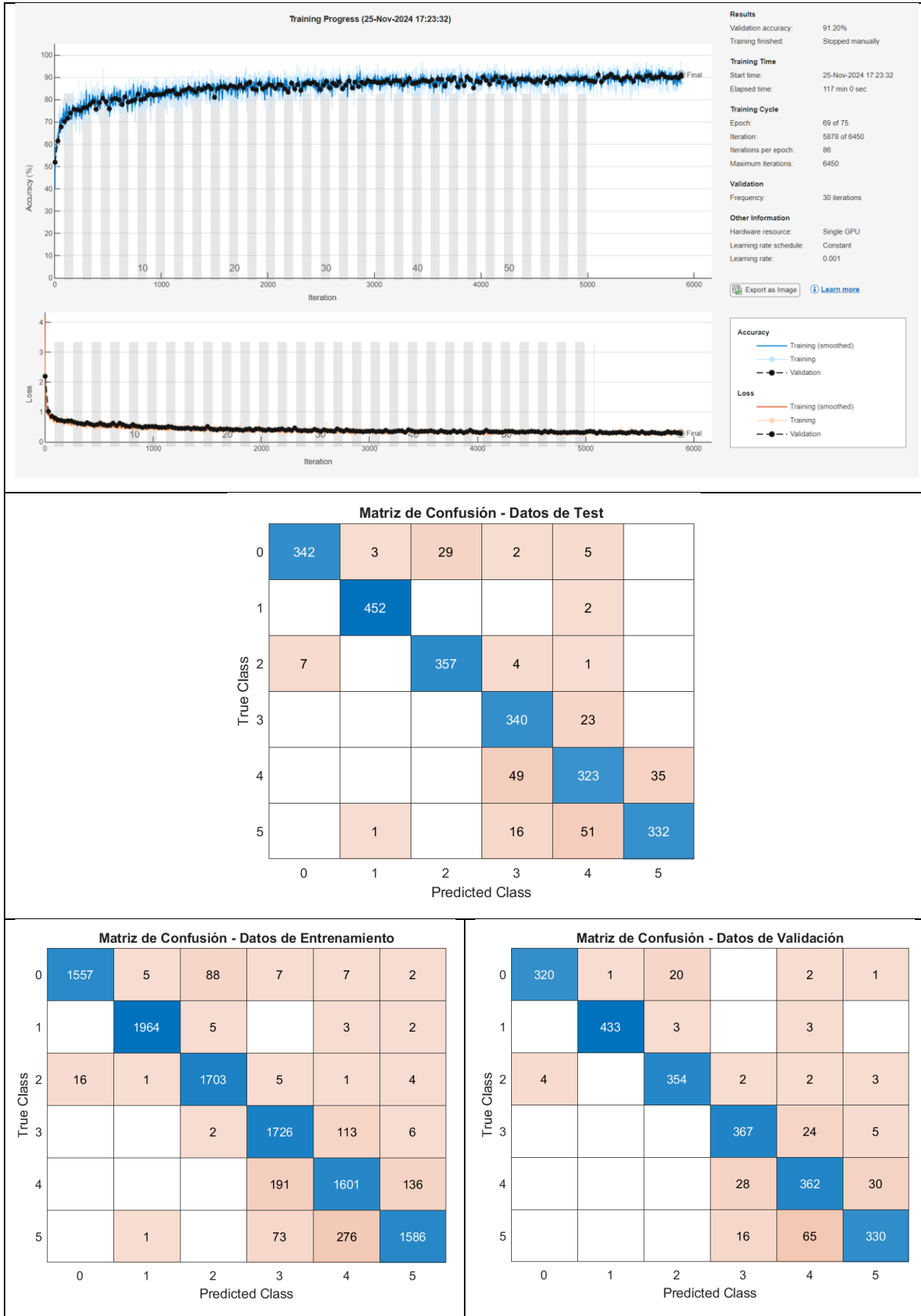


Ilustración 3-39. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para el nivel de diflufenican en los espectros medios de las cuadrículas extraídas automáticamente.

La **lambda-cihalotrina** proporciona unos resultados excelentes con una exactitud que supera el 95% en los datos de test, que incluso llegan a superar los resultados de los métodos de aprendizaje automático. Este valor tan elevado llega, no obstante, después de aplicar una optimización bayesiana.

Las gráficas de exactitud presentan una evolución muy clara con un crecimiento positivo y separación entre ambas curvas (Ilustración 3-40). El tiempo de ejecución es de poco más de 2 horas.

Las matrices indican una buena clasificación con pocos casos identificados de forma incorrecta (Ilustración 3-40). La poca confusión se centra entre las clases de baja concentración (4 y 5), ambas por debajo del límite europeo.

En el caso del **oxifluorfen**, igualmente se obtienen muy buenos resultados, próximos al 90%, aunque de nuevo inferiores a los obtenidos con algunos métodos de aprendizaje automático como el análisis discriminante y la red neuronal.

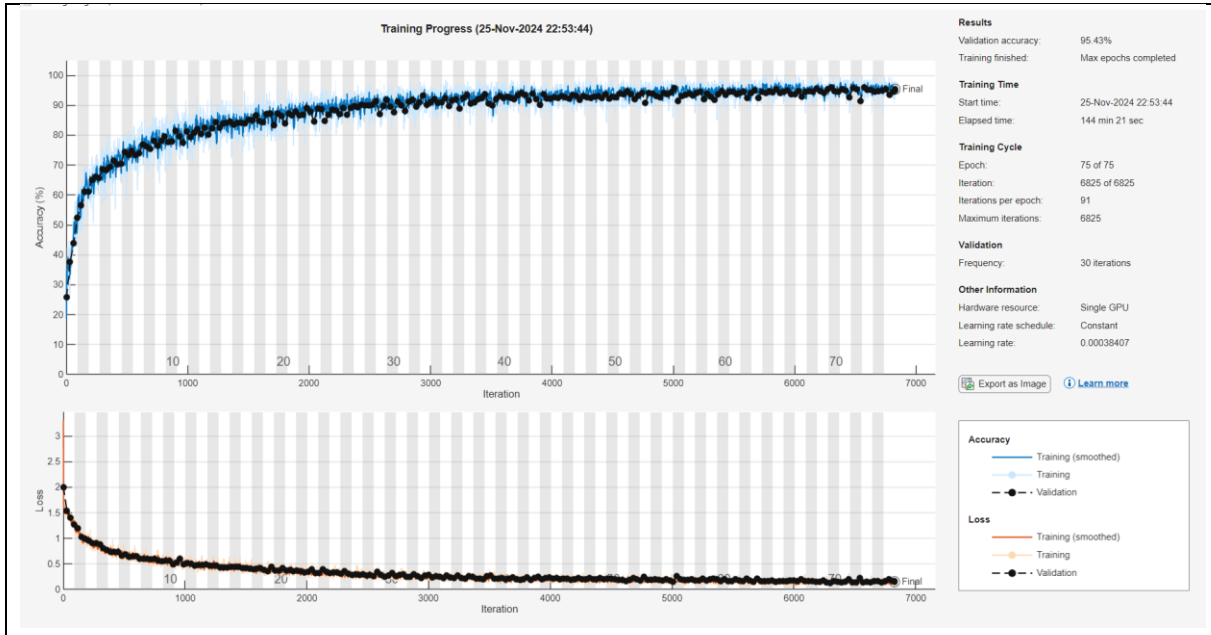
La evolución de las gráficas de exactitud es también muy correcta con un ascenso progresivo y coordinado de ambas curvas. El tiempo de ejecución es algo menor en este caso, de algo menos de 1 hora (Ilustración 3-41).

Las matrices de confusión permiten observar la calidad de la clasificación aunque se observa una cierta confusión entre las concentraciones 2 y 5 (Ilustración 3-41), lo que no es muy habitual y en este caso supone un cierto problema de catalogación.

Por último, el **tebuconazol** es el producto que produce los peores resultados con una exactitud en test del 80%, inferior en unos 10 puntos a la que aparece en los métodos de aprendizaje automático, y después de haber aplicado la optimización.

La gráfica de exactitudes presenta un aspecto bastante correcto con un crecimiento progresivo y coherente de ambas curvas (Ilustración 3-42), indicativo de un proceso de entrenamiento correcto a pesar de los resultados.

Las matrices presentan solo una limitada confusión entre las clases de media-alta concentración (1 y 3) y las de baja concentración (4 y 5), lo que no supone ningún problema que el límite europeo está entre las clases 3 y 4 (Ilustración 3-42).



Matriz de Confusión - Datos de Test

0	360	1	10			
1	2	403	1	5		6
2	6		422			1
3		3		435	12	1
4				15	416	22
5	1			8	25	342
	0	1	2	3	4	5

True Class

Predicted Class

Matriz de Confusión - Datos de Entrenamiento

0	1602	2	32			
1	6	1993	2	29	6	15
2	20		1928			7
3	2	5		2021	35	24
4	1	4	1	85	1843	44
5	1	3	1	29	82	1826
	0	1	2	3	4	5

True Class

Predicted Class

Matriz de Confusión - Datos de Validación

0	367	3	14			
1	2	390		9	2	7
2	4		438			3
3	1	4		434	7	4
4	1	2	1	11	390	17
5	1			5	16	363
	0	1	2	3	4	5

True Class

Predicted Class

Ilustración 3-40. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para el nivel de lambda-cihalotrina en los espectros medios de las cuadrículas extraídas automáticamente.

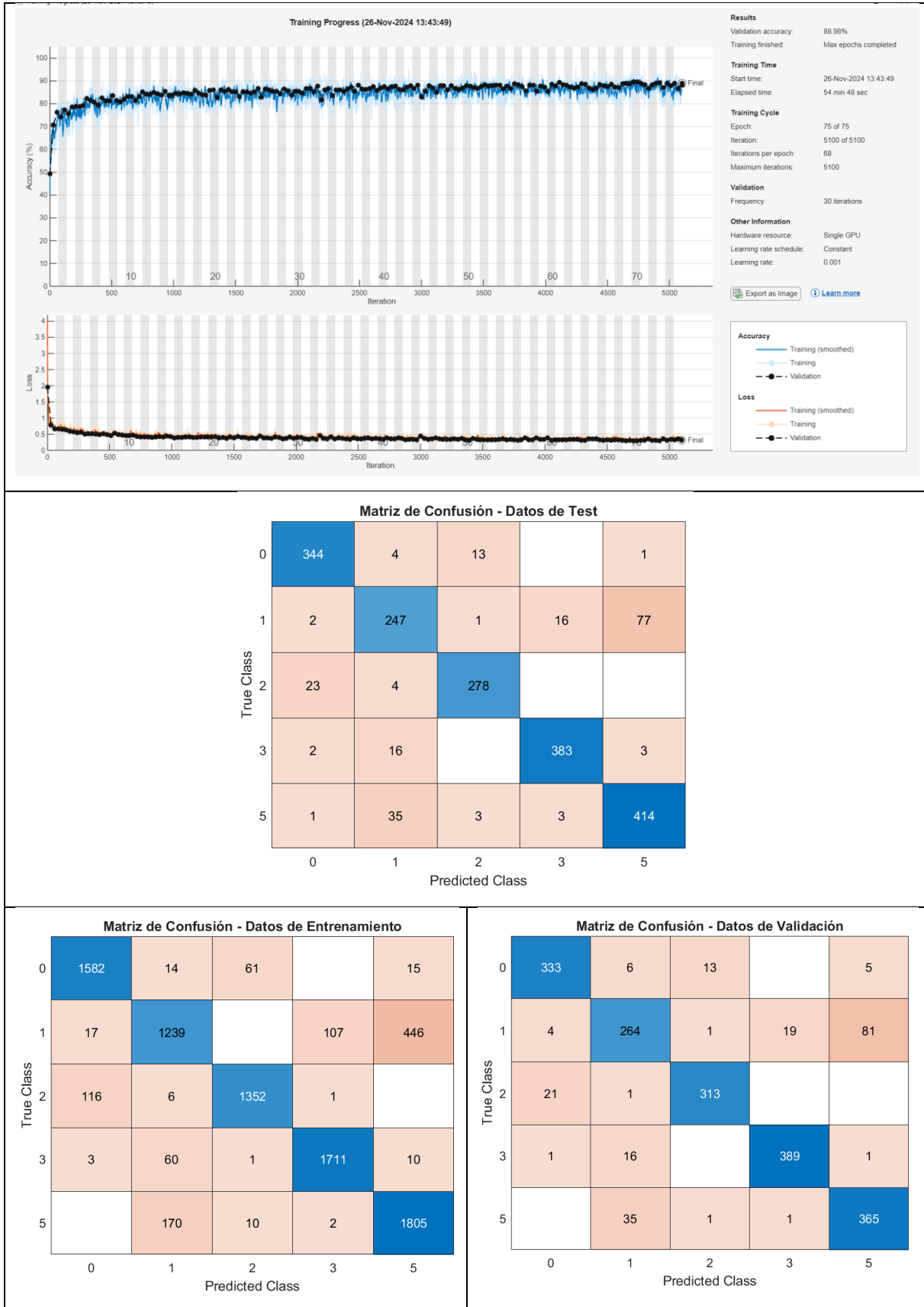
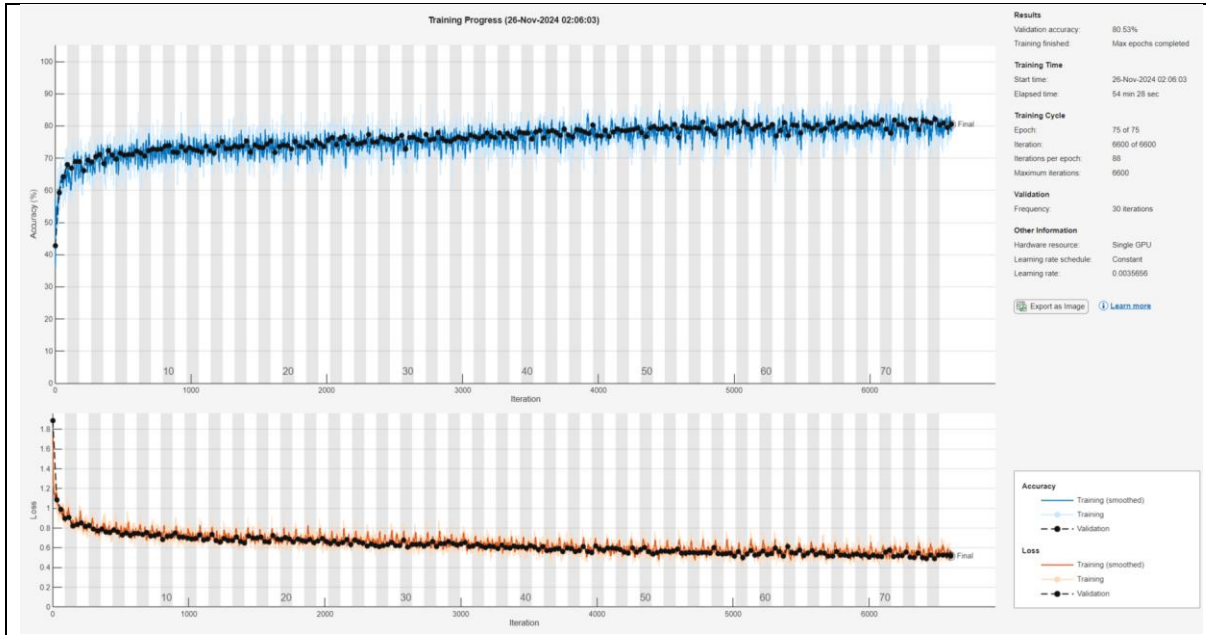


Ilustración 3-41. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para el nivel de oxifluorfen en los espectros medios de las cuadrículas extraídas automáticamente.



Matriz de Confusión - Datos de Test

	0	1	2	3	4	5
0	321	10	5	22	2	3
1	9	290	9	81	5	9
2	1	14	350	35	2	10
3	32	22	11	366	2	2
4	7	2		1	352	27
5	3	12	5	8	134	254
	0	1	2	3	4	5

Predicted Class

Matriz de Confusión - Datos de Entrenamiento

	0	1	2	3	4	5
0	1469	61	14	94	13	19
1	24	1341	53	320	13	29
2	6	40	1526	154	1	30
3	165	82	33	1794	5	16
4	37	4		12	1810	176
5	14	80	28	48	581	1195
	0	1	2	3	4	5

Predicted Class

Matriz de Confusión - Datos de Validación

	0	1	2	3	4	5
0	312	9	4	25	4	4
1	4	289	6	61	1	6
2	3	7	336	38		4
3	36	22	3	386	2	1
4	11	2			379	29
5	4	17	7	7	154	246
	0	1	2	3	4	5

Predicted Class

Ilustración 3-42. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para el nivel de tebuconazol en los espectros medios de las cuadrículas extraídas automáticamente.

De la aplicación del aprendizaje profundo (red convolucional 1D) llevado a cabo con los espectros medios de las ventanas creadas automáticamente, se pueden derivar las siguientes observaciones:

- Los resultados obtenidos para la exactitud en validación y test son en general muy buenos, en la mayor parte de los casos por encima del 90%, tanto para los tipos como para los niveles de producto en cada tipo, con la excepción del tebuconazol.
- No obstante, en este caso los valores de exactitud son algo inferiores en términos generales a los obtenidos con algunos métodos de aprendizaje automático como el análisis discriminante, SVM y la red neuronal, debido principalmente en los niveles tan altos que proporcionan estos últimos. Esto resulta en cierto modo sorprendente, ya que son métodos en principio más sencillos que las redes convolucionales, especialmente el análisis discriminante lineal, un método clásico que muestra limitaciones en problemas complejos como las situaciones que se presentan en la naturaleza. No obstante, en el estudio de (Martínez-Gila, et al., 2024) también se produce este mismo efecto (con valores cercanos al 100% en validación y algo menores en test), por lo cual los resultados son consistentes, a pesar de que se pueda pensar en algún tipo de sobreajuste de estos métodos en este muestreo masivo de datos.
- Dentro de los buenos niveles generales destacan los obtenidos para el diflufenican (90%), la deltametrina (93%), lambda-cihalotrina (95%) y en menor medida el oxifluorfen (89%), que ya proporcionaban valores de exactitud muy satisfactorios en el anterior muestreo, aunque algunos solo después de aplicar la optimización. En la otra parte se encuentra el tebuconazol (fungicida) que a pesar de aplicar la optimización no sube del 80%, coincidiendo con lo que ocurre en los anteriores muestreos. La diferenciación por tipos también resulta satisfactoria (91%).
- Como ya sucedía en el muestreo por ventanas definidas por el usuario, pero a diferencia de lo que ocurría con los espectros medios, los resultados de exactitud con entrenamiento y con validación y test son bastante coherentes, lo que se debe al aumento de las muestras que permiten un mejor

entrenamiento y testeo de la misma. No obstante, salvo en el caso del nivel de lambda-cihalotrina no se observa una mejora tan significativa al aumentar de forma considerable la muestra, como se observa en el paso anterior de los espectros medios al muestreo de las ventanas.

- La evolución de la gráfica de exactitud en entrenamiento y test corrobora lo dicho, observándose un crecimiento sostenido y armónico de las curvas, sin estancamientos, ni grandes irregularidades. Este crecimiento se refleja en los valores obtenidos finalmente en los valores de exactitud final, aunque no es igual para todos los productos analizados.
- La estabilidad de la red permite que la repetición del proceso de entrenamiento produzca resultados fiables, e incluso que vaya mejorando a medida que la red va siendo entrenada con más datos.
- El tiempo de ejecución es del orden de las decenas de minutos, e incluso de horas, lo que representa un aumento considerable respecto al muestreo de ventanas (pocos minutos) y por supuesto a los espectros medios (segundos). Aunque como ya se ha comentado, esto no supone un problema muy grande, ya que una vez entrenada una red su aplicación es mucho más rápida.
- Las matrices de confusión permiten en líneas generales comprobar la buena clasificación de las muestras, tanto en el caso de los tipos de pesticida como sobre todo en los niveles de los mismos. En general, permite diferenciar bien las muestras con niveles superiores e inferiores al límite europeo de concentración de cada uno de estos productos, salvo algunas excepciones.
- En este caso, el tamaño de las muestras se considera satisfactorio ya que permite un adecuado proceso de entrenamiento, así como una buena validación y testeo, aunque puede resultar demasiado grande ya que dificulta el manejo de los datos y aumenta los tiempos de computación, sin ofrecer una mejora realmente significativa en el rendimiento de las redes. Un tamaño como el anterior o quizá algo más grande, pero no tanto como en este caso, quizá sea el que ofrezca un mejor rendimiento.

3.3. Selección de características de las aceitunas

En esta sección, se presentan los resultados de la selección de características por distintos métodos, como se ha descrito en la metodología: basado en el análisis de la varianza (ANOVA), la eliminación recursiva de características, los bosques aleatorios (*random forest*) y el mapa de importancia basado en la red convolucional. Como se ha venido haciendo a lo largo de todo el trabajo, se han examinado los resultados por el tipo de producto; y dentro de cada uno de ellos, en función del nivel de concentración del mismo.

	Exactitud	Entrenamiento (%)	Validación (%)	Test (%)
Deltametrina	280 bandas	98,14	92,00	90,67
	Mapa de importancia	96,57	89,33	88,00
	Recursiva	97,48	96,00	92,00
	<i>Random Forest</i>	96,43	90,67	86,00
Diflufenican	280 bandas	91,13	91,95	84,40
	Mapa de importancia	93,10	86,90	88,51
	Recursiva	97,04	94,25	87,40
	<i>Random Forest</i>	94,46	85,06	83,91
Oxifluorfen	280 bandas	96,94	93,24	87,67
	Mapa de importancia	96,94	93,24	86,30
	Recursiva	97,08	93,24	86,30
	<i>Random Forest</i>	96,21	89,19	89,04
Lambda	280 bandas	95,47	90,00	92,22
	Mapa de importancia	87,00	84,51	85,52
	Recursiva	90,45	89,58	87,64
	<i>Random Forest</i>	89,79	85,91	85,29
Tebuconazol	280 bandas	86,59	80,23	80,64
	Mapa de importancia	78,81	76,41	75,11
	Recursiva	82,20	80,43	78,82
	<i>Random Forest</i>	74,72	75,11	72,35
Por Tipo	280 bandas	87,48	86,32	85,18
	Mapa de importancia	82,91	81,48	78,03
	Recursiva	92,18	87,75	87,74
	<i>Random Forest</i>	84,12	76,64	80,92

Tabla 3-22. Resultado selección de características en los datos de entrenamiento, validación y test para tres métodos: evaluación de activaciones en red 1-D (mapa de importancia), selección recursiva y *random forest*.

Se observa en la Tabla 3-22, que, en prácticamente la totalidad de los casos, los resultados obtenidos mediante la selección de características son muy similares a los obtenidos considerando todas las bandas. De forma general, los resultados son aceptables con todos los métodos, pero se puede observar que, para la selección recursiva se suelen obtener unas exactitudes más elevadas. Esto se debe a que es el método más complejo de los tres, y, además no depende de los resultados previos, como si lo hace el de mapa de importancia. Se puede observar como para el tebuconazol y la lambda-cihalotrina, en los que se obtenían peores resultados durante el entrenamiento de la red (más inestable), se obtiene peor resultado precisamente para el método que depende de ella. En el análisis por tipos de pesticida se observa una situación parecida con valores de exactitud similares a los obtenidos con todas las bandas, e incluso en la eliminación recursiva los valores de exactitud son ligeramente superiores, lo que corrobora la validez de la selección de características en este caso.

Respecto a la descripción de las bandas que quedan seleccionadas, comenzando por el **tipo de producto**, los resultados se presentan en la Ilustración 3-43. En ella se observan una serie de franjas de bandas comunes entre los distintos métodos utilizados:

- Bandas 11-30 (longitudes de onda entre 400 y 440 nm). Estas bandas de la parte inicial del visible, que se conoce en teledetección como el azul costero (*coastal/aerosol*), aparecen en diversos métodos como la eliminación recursiva, bosques aleatorios y apenas en el ANOVA, aun con variaciones y sin una definición clara.
- Bandas 71-78 (longitudes de onda entre 524 y 538 nm). Esta zona, cercana al verde (547 nm), aparece claramente en la eliminación recursiva, bosques aleatorios y en el ANOVA. Tiene bastante sentido físico pues la reflectancia en esta región está relacionada con la vegetación (color verde); y los cambios en el estado de la misma, por ejemplo, por la aplicación de fitosanitarios, puede afectar al verdor de los frutos y hojas.
- Bandas 133-155 (longitudes de onda 654-700 nm). Aunque de forma discontinua, esta región aparece con elevados índices de importancia en la eliminación recursiva, bosques aleatorios y en el ANOVA, mientras que en el

mapa de importancia derivado de la red convolucional aparece desplazado entre las bandas 161-175 (713-743 nm). Corresponde a la región conocida como límite del rojo o *red edge*, aunque en este caso ocupa parte de la región del rojo (640 nm). En toda esta zona se producen cambios igualmente relacionados con el estado de la vegetación al ser el inicio de la zona de rampa entre la baja reflectancia en el rojo y la meseta del infrarrojo cercano (NIR).

- Bandas 270-290 (longitudes de onda 952-1019 nm). Aparece prácticamente en todos los métodos y corresponde ya con el pleno infrarrojo cercano, donde la reflectancia de la vegetación es alta. Está relacionada con la estructura de las paredes vegetales, por lo que cambios en el estado de la vegetación puede afectar a esta zona y en general a toda la región del infrarrojo cercano.

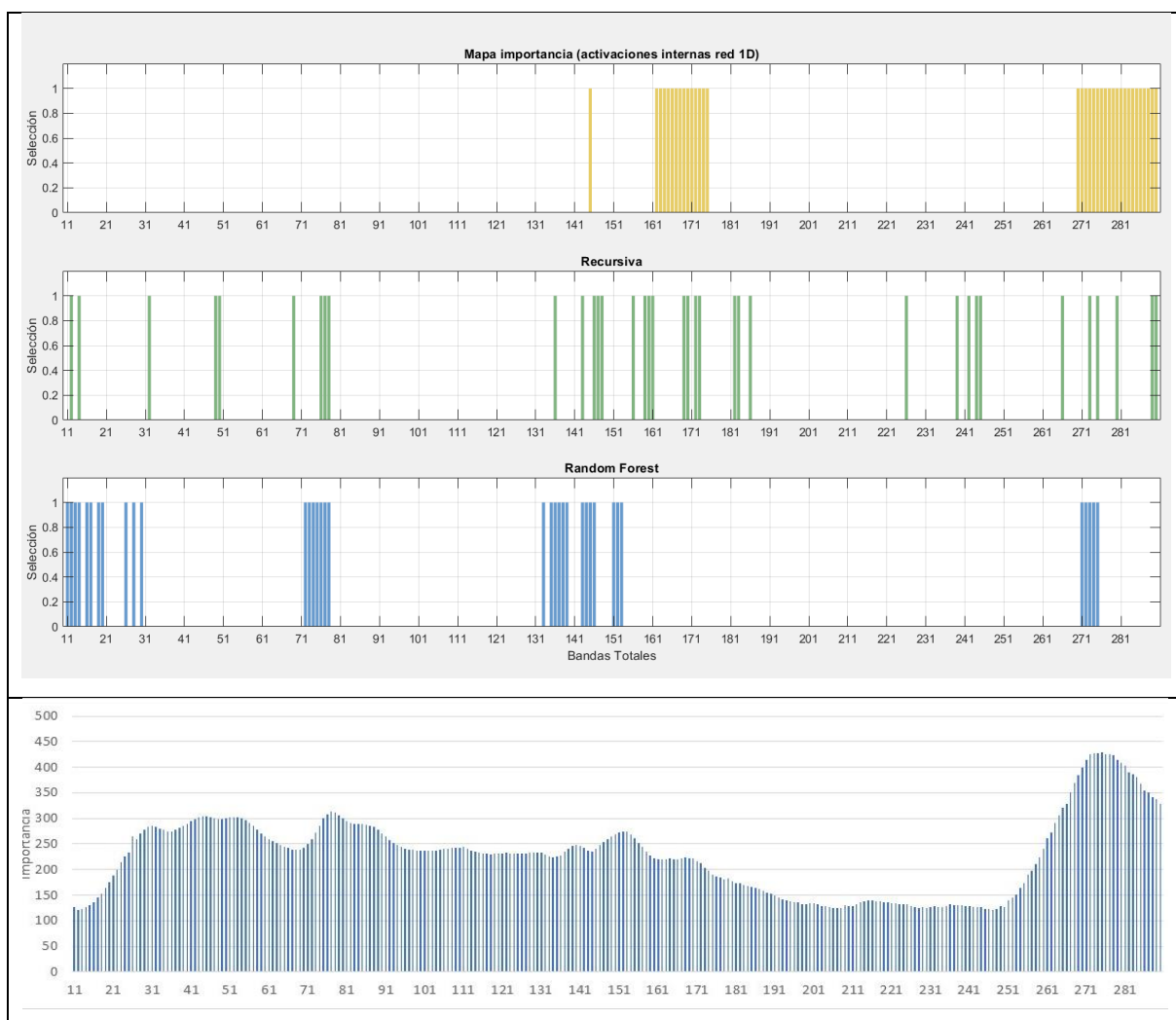


Ilustración 3-43. Selección de características mediante diversos métodos: mapa de importancia (red 1D); eliminación recursiva; bosques aleatorios; análisis de la varianza (ANOVA) para los tipos de producto.

Si se observa la curva espectral de reflectancia (Ilustración 3-44), en estas regiones se aprecian diferencias en la reflectancia media de las muestras de cada producto. Más específicamente y dejando de lado la primera región, en las otras tres se aprecian bien las diferencias. La región del verde es una zona de máximo relativo en el visible (verde), la del rojo-borde del rojo es un mínimo muy pronunciado y la última es una región del infrarrojo cercano, ya en el inicio de la caída de la meseta hacia el infrarrojo medio, muy influenciado por el estado de humedad de la vegetación.

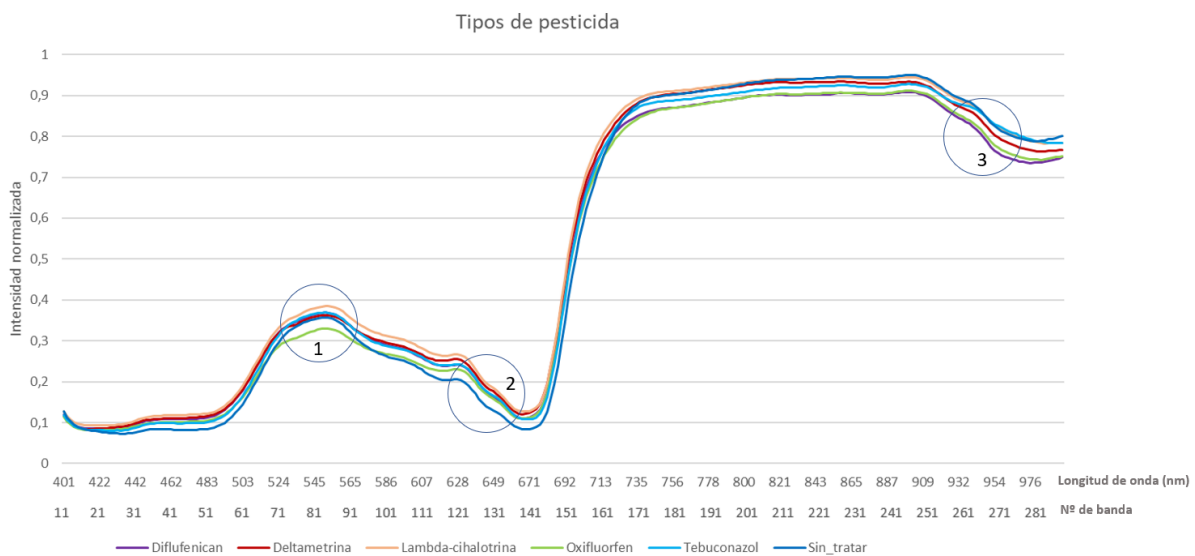


Ilustración 3-44. Regiones del espectro marcadas como más importantes en los métodos de selección de características para diferenciar por tipos de producto.

Por **nivel de producto**, los resultados de la selección de características se muestran en las figuras siguientes. Comenzando por la **deltametrina** (Ilustración 3-45), en ella se observan una serie de franjas comunes en algunos de los métodos:

- Bandas 11-25 (longitudes de onda entre 400 y 430 nm). Estas bandas de la parte inicial del visible (azul costero), aparecen prácticamente en los tres primeros métodos y apenas en el ANOVA, aun con variaciones y sin una definición clara.
- Bandas 65-80 (longitudes de onda entre 512 y 543 nm). Esta zona, cercana al verde (82 nm), aparece claramente en los tres primeros métodos, aun con cierta discontinuidad, no así en ANOVA. Tiene bastante sentido físico pues la reflectancia en esta región está relacionada con la vegetación (verde) y sus cambios de estado por alguna razón como la aplicación de fitosanitarios.

- Bandas 140-160 (longitudes de onda 670-710 nm). De nuevo, aunque de forma discontinua, esta región aparece con elevados índices de importancia en prácticamente todos los métodos, incluido el ANOVA. En este caso corresponde a la región conocida como límite del rojo o *red edge*, donde se producen cambios igualmente relacionados con el estado de la vegetación al ser el inicio de la zona de rampa entre la baja reflectancia de la misma en el rojo y la meseta de valores elevados del infrarrojo cercano.
- Bandas 270-290 (longitudes de onda 952-1019 nm). Aparece prácticamente en todos los métodos y corresponde al infrarrojo cercano, donde la reflectancia de la vegetación es alta. Los cambios en el estado de la vegetación pueden afectar a esta zona como a todo el infrarrojo cercano.

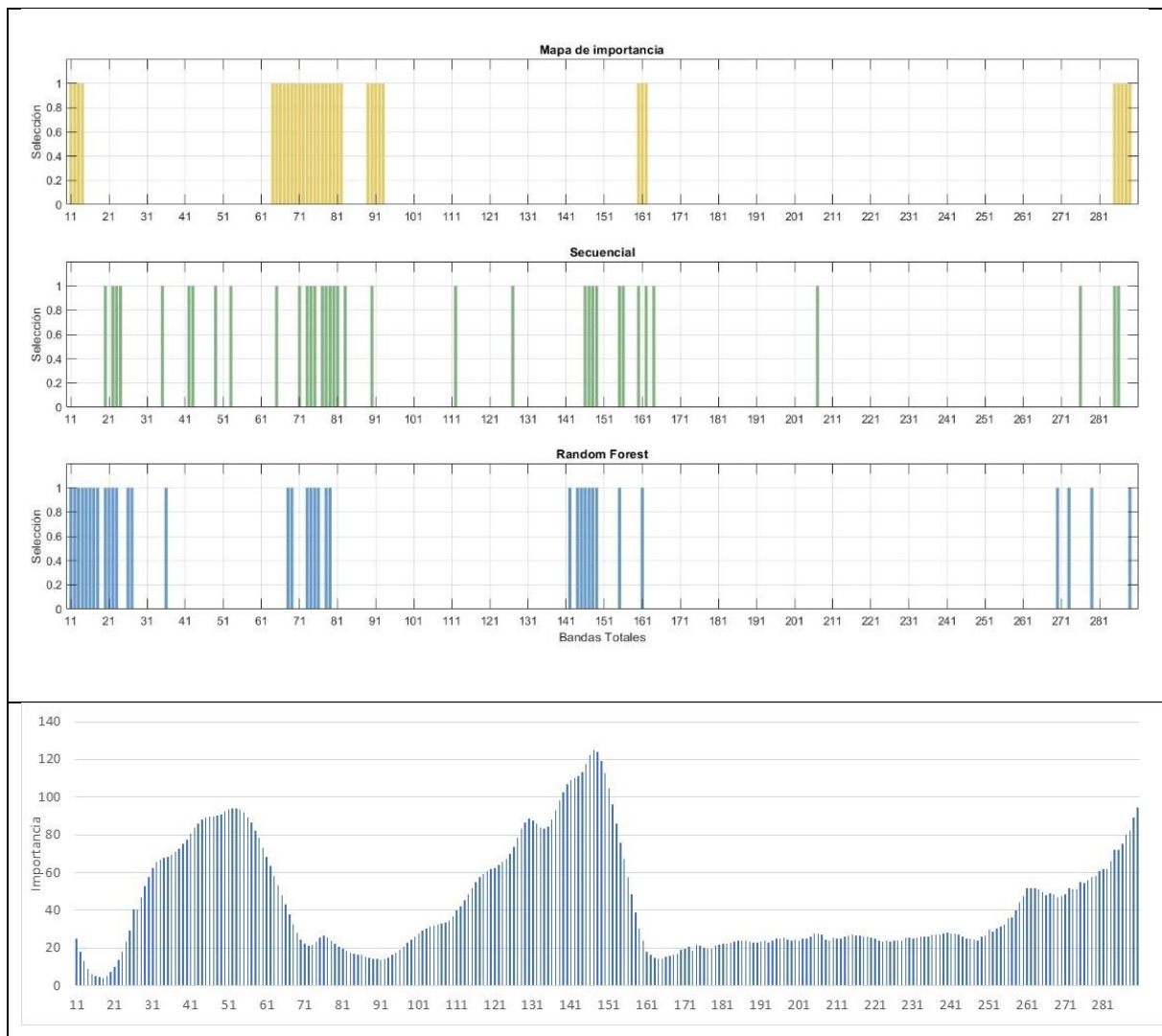


Ilustración 3-45. Selección de características mediante diversos métodos: mapa de importancia (red 1D); eliminación recursiva; bosques aleatorios; análisis de la varianza (ANOVA) para los niveles de deltametria.

Observando las curvas espectrales (Ilustración 3-46), en estas regiones se aprecian diferencias en la reflectancia de las muestras tratadas con distintos niveles de este producto. Más específicamente y dejando de lado nuevamente la primera región, en las otras tres se aprecian bien las diferencias. La región del verde es una zona de máximo relativo, la del borde del rojo es un mínimo muy pronunciado y la última es una región del infrarrojo cercano, todas ellas regiones muy afectadas por los cambios en los elementos vegetales.

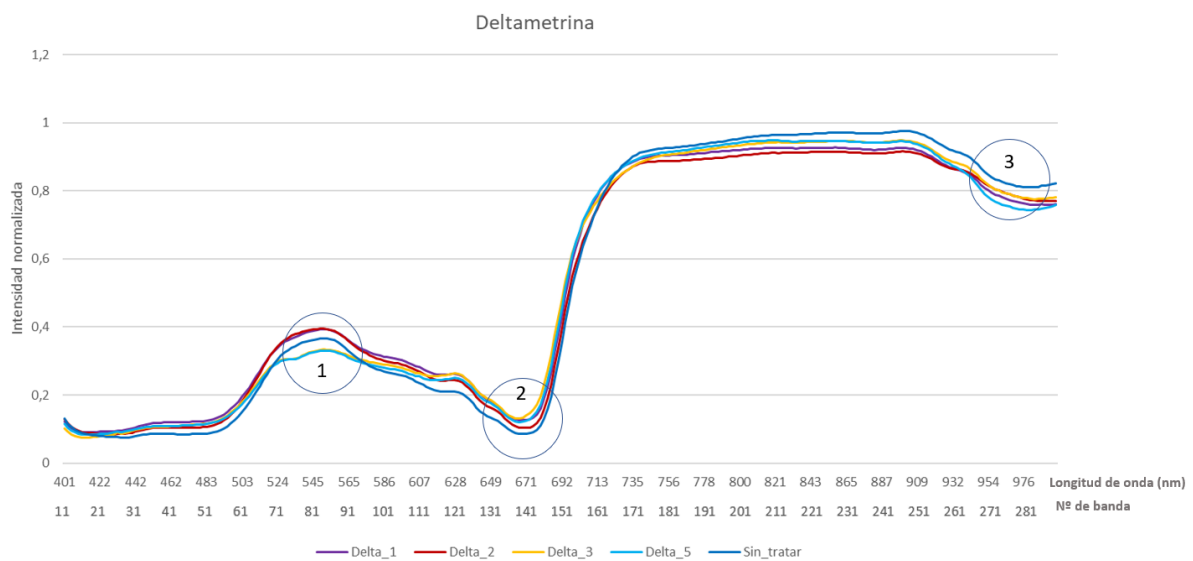


Ilustración 3-46. Regiones del espectro marcadas como más importantes en los métodos de selección de características para diferenciar por niveles de deltametrina.

En el **diflufenican** se observan como en los casos anteriores una serie de franjas comunes en algunos de los métodos (Ilustración 3-47):

- Bandas 11-30 (longitudes de onda entre 400 y 440 nm). Estas bandas de la parte inicial del visible aparecen muy claramente en la eliminación recursiva y especialmente en los bosques aleatorios e incluso en ANOVA.
- Bandas 71-80 (longitudes de onda entre 524 y 543 nm). Esta zona, cercana al verde (82 nm), aparece muy claramente en los tres primeros métodos, aun con cierta discontinuidad, no tanto en ANOVA. Tiene bastante sentido físico pues la reflectancia en esta región está relacionada con los elementos vegetales (verde) y sus cambios de estado.
- Bandas 145-160 (longitudes de onda 679-710 nm). Esta región aparece con claridad mostrando elevados índices de importancia en prácticamente todos

los métodos, incluido el ANOVA. En este caso corresponde de forma inequívoca a la región del límite del rojo, donde se producen cambios igualmente relacionados con el estado de la vegetación en la zona de transición entre el visible y el infrarrojo cercano.

- Bandas 270-290 (longitudes de onda 952-1019 nm). Aparece prácticamente en todos los métodos y corresponde al infrarrojo cercano, donde la reflectancia de la vegetación es alta, pudiendo ser afectada por los cambios.

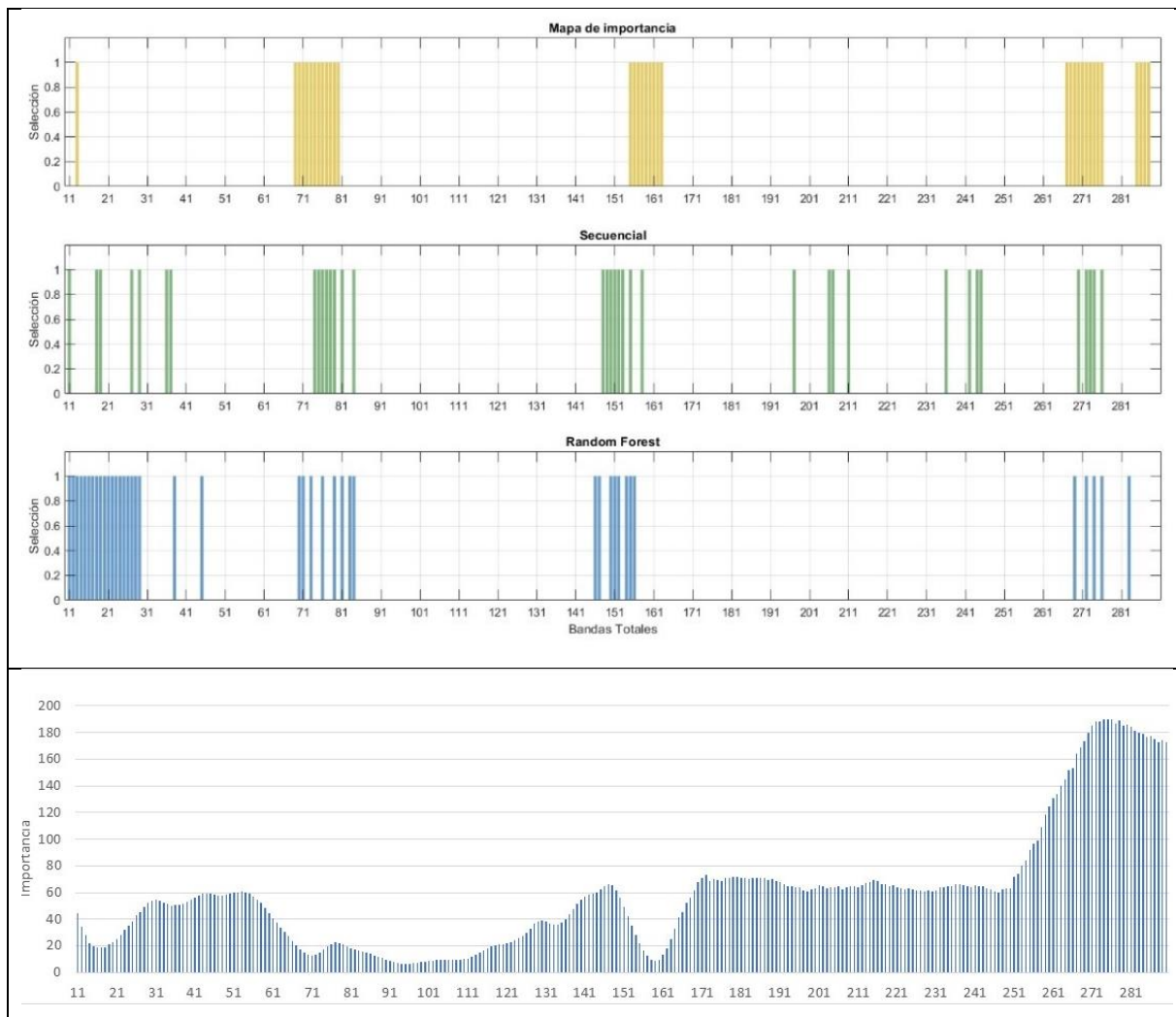


Ilustración 3-47. Selección de características mediante diversos métodos: mapa de importancia (red 1D); eliminación recursiva; bosques aleatorios; análisis de la varianza (ANOVA) para los niveles de diflufenican.

Si se observan las curvas espectrales (Ilustración 3-48), en estas regiones se aprecian diferencias en la reflectancia de las muestras tratadas con distintos niveles de este producto. Más específicamente, se puede en este caso considerar cuatro regiones, donde se aprecian dichas diferencias. La región del azul costero, la región

del verde como zona de máximo relativo, la del borde del rojo como un mínimo muy pronunciado y la región del infrarrojo cercano, todas ellas muy afectadas por los cambios en los elementos vegetales, salvo la primera donde no está muy clara su influencia.

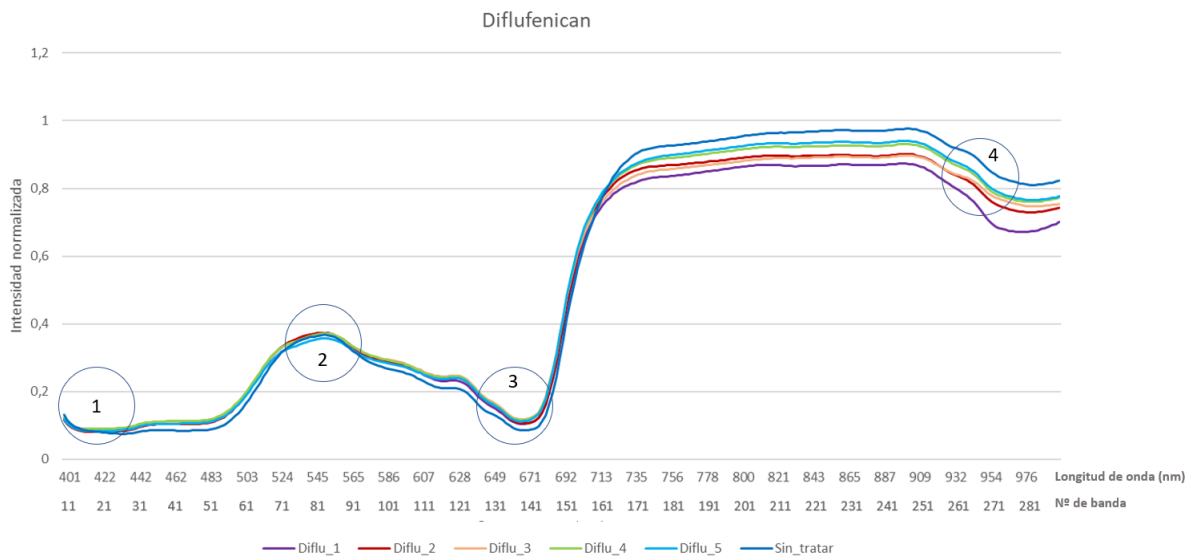


Ilustración 3-48. Regiones del espectro marcadas como más importantes en los métodos de selección de características para diferenciar por niveles de diflufenican.

En cuanto a la **lambda-cihalotrina**, las franjas comunes que se encuentran en los distintos métodos son las siguientes (Ilustración 3-49):

- Bandas 11-30 (longitudes de onda entre 400 y 440 nm). Estas bandas de la parte inicial del visible aparecen en la eliminación recursiva y especialmente en los bosques aleatorios y ahora más claramente en ANOVA.
- En este caso no aparece una región claramente definida en las inmediaciones del verde (547 nm), salvo bandas aisladas en la eliminación recursiva y los bosques aleatorios.
- Bandas 145-160 (longitudes de onda 679-710 nm). Esta región aparece con una cierta claridad mostrando elevados índices de importancia en prácticamente todos los métodos, incluido el ANOVA, aunque muy tenuemente en el mapa de importancia derivado de la red convolucional. En este caso corresponde también de forma evidente a la región del límite del rojo, donde se producen cambios relacionados con el estado de la vegetación.

- Bandas 260-290 (longitudes de onda 930-1019 nm). Aparece prácticamente en todos los métodos y corresponde al infrarrojo cercano, donde la reflectancia de la vegetación es alta, pudiendo ser afectada por los cambios. En este caso y en relación sobre todo con el mapa de importancia de la red y también con la eliminación la región puede extenderse a más longitudes de onda dentro del infrarrojo cercano.



Ilustración 3-49. Selección de características mediante diversos métodos: mapa de importancia (red 1D); eliminación recursiva; bosques aleatorios; análisis de la varianza (ANOVA) para los niveles de lambda-cihalotrina.

Si se observan las curvas espectrales (Ilustración 3-50), en estas regiones se aprecian diferencias en la reflectancia de las muestras tratadas con distintos niveles de este producto. Más específicamente, se puede en este caso considerar tres regiones que son: la región del azul costero, la del borde del rojo como un mínimo muy pronunciado y la región del infrarrojo cercano, en este caso ampliada hacia longitudes de onda menores. Todas ellas son regiones sensibles a los cambios en los elementos

vegetales, salvo la primera donde no está muy clara su influencia. Llama la atención a que en este caso no hay cambios claros en la región cercana al verde, y sí una mayor influencia de la región del azul al azul costero.

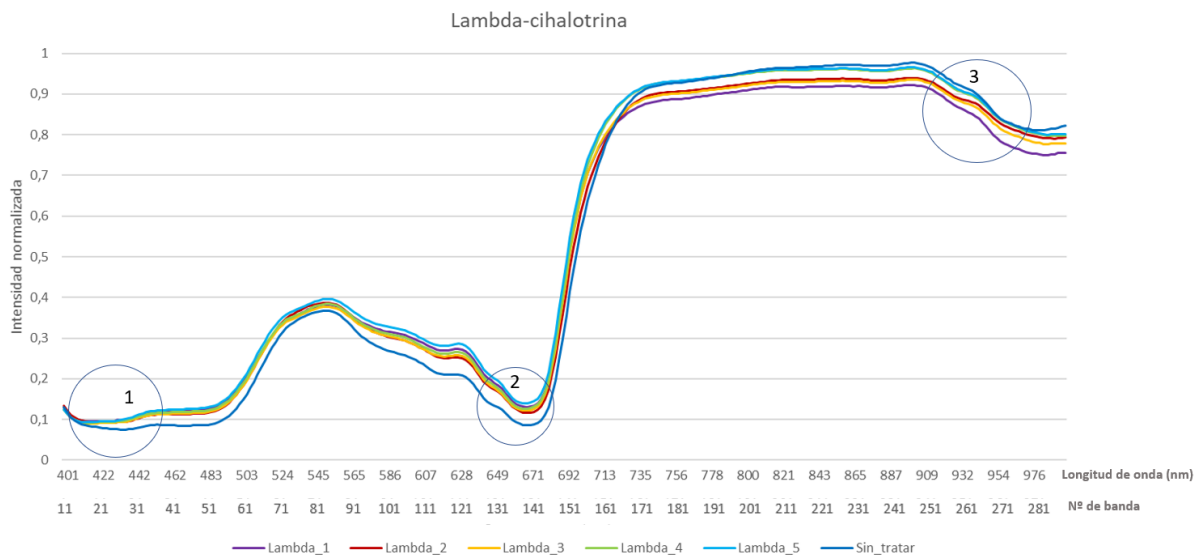


Ilustración 3-50. Regiones del espectro marcadas como más importantes en los métodos de selección de características para diferenciar por niveles de lambda-cihalotrina.

En el caso del **oxifluorfen**, se observan las siguientes franjas comunes entre los distintos métodos (Ilustración 3-51):

- Bandas 11-30 (longitudes de onda entre 400 y 440 nm). Estas bandas de la parte inicial del visible aparecen especialmente en el mapa de importancia derivado de la red y los bosques aleatorios, y en menor medida en eliminación recursiva y ANOVA.
- Bandas 70-80 (longitudes de onda entre 522 y 543 nm). Esta zona, cercana al verde (82 nm), aparece muy claramente en los tres primeros métodos, aun con cierta discontinuidad, pero no en ANOVA. Tiene bastante sentido físico pues la reflectancia en esta región está relacionada con los elementos vegetales (verde) y sus cambios de estado.
- Bandas 140-160 (longitudes de onda entre 669 y 710 nm). Esta región aparece muy claramente, mostrando elevados índices de importancia en prácticamente todos los métodos, incluido el ANOVA. En este caso corresponde también evidentemente a la región del límite del rojo, donde se producen cambios relacionados con el estado de la vegetación.

- Bandas 270-290 (longitudes de onda 952-1019 nm). Aparece prácticamente en todos los métodos, aunque no de forma tan clara como en otros casos. Corresponde al infrarrojo cercano, donde la reflectancia de la vegetación es alta, pudiendo ser afectada por los cambios.

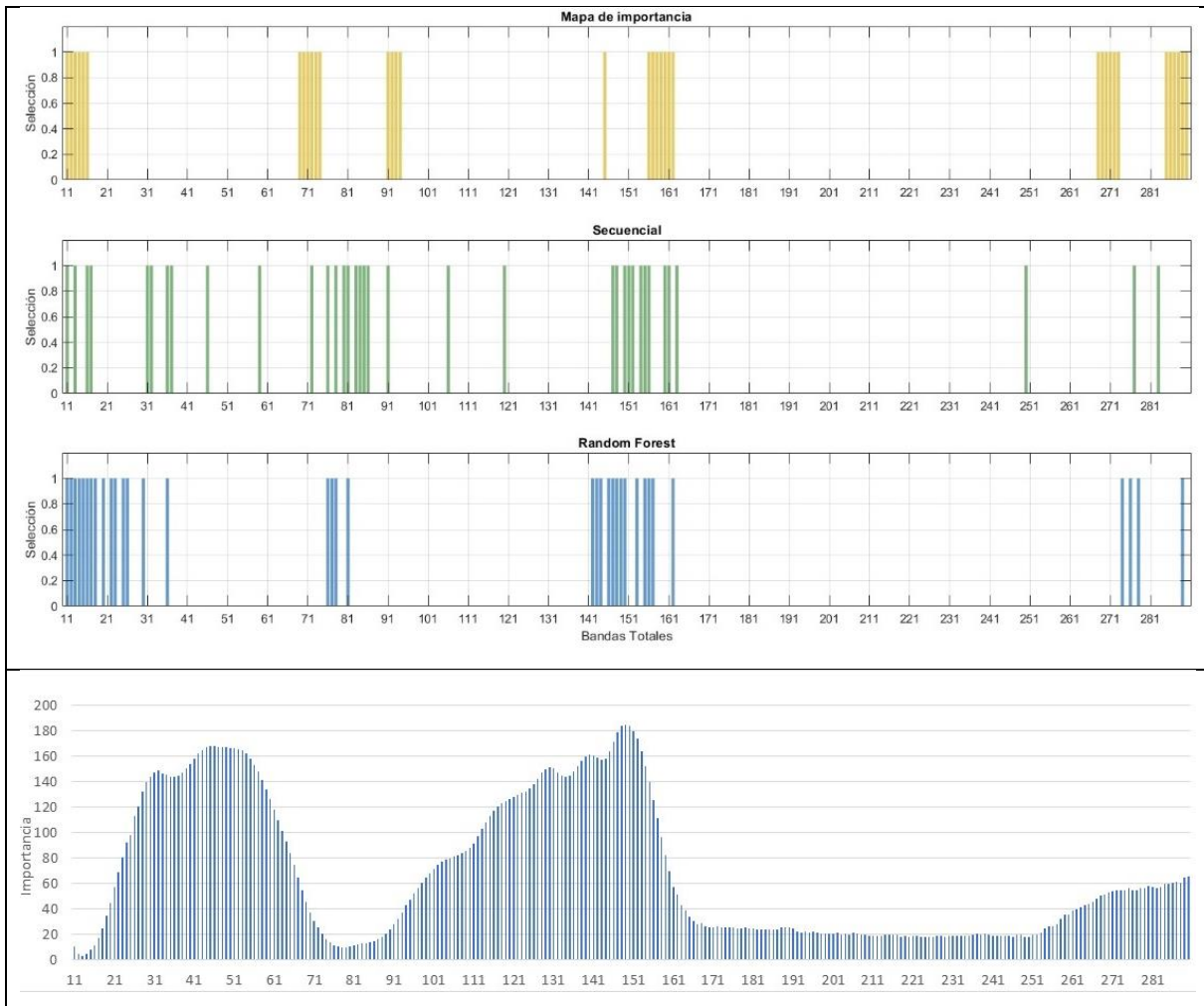


Ilustración 3-51. Selección de características mediante diversos métodos: mapa de importancia (red 1D); eliminación recursiva; bosques aleatorios; análisis de la varianza (ANOVA) para los niveles de oxifluorfen.

Si se observan las curvas espectrales (Ilustración 3-52), en estas regiones se aprecian diferencias en la reflectancia de las muestras tratadas con distintos niveles de este producto. Más específicamente y dejando de lado la primera región, que no está claramente definida, en las otras tres se aprecian bien las diferencias: la región del verde, como un máximo relativo dentro del visible; la del borde del rojo como un mínimo muy pronunciado; y la última es una región del infrarrojo cercano. Todas ellas son regiones muy afectadas por los cambios en los elementos vegetales.

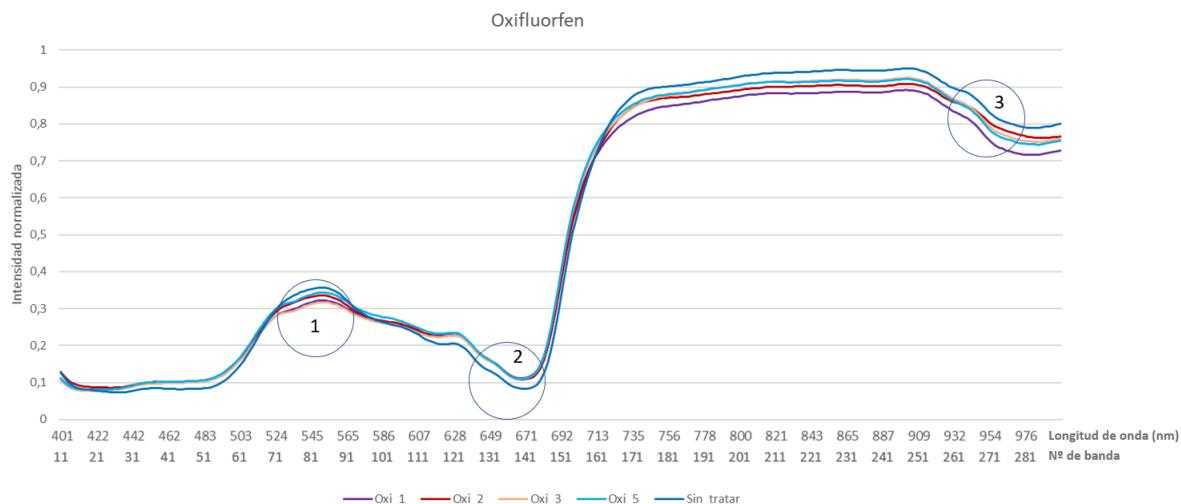


Ilustración 3-52. Regiones del espectro marcadas como más importantes en los métodos de selección de características para diferenciar por niveles de oxifluorfen.

Por último, en el caso del **tebuconazol**, se pueden observar las siguientes franjas comunes entre los distintos métodos (Ilustración 3-53):

- Bandas 11-31 (longitudes de onda entre 400 y 442 nm). Estas bandas de la parte inicial del visible aparecen con cierta claridad solo en los bosques aleatorios y en menor medida en la eliminación recursiva y ANOVA.
- En este caso, como ya ocurría en la lambda-cihalotrina tampoco aparece claramente seleccionada la región del entorno del verde (82 nm); hay algunas bandas aisladas y un cierto máximo en ANOVA, pero se ha de descartar.
- Bandas 140-170 (longitudes de onda 669-733 nm). Esta región aparece muy claramente, mostrando elevados índices de importancia en prácticamente todos los métodos, incluido el ANOVA. Corresponde a la región del límite del rojo, donde se producen cambios relacionados con el estado de la vegetación; pero en este caso más extendida hacia longitudes de onda mayores, ya en la rampa que conecta al mínimo del rojo con la meseta del infrarrojo cercano.
- Bandas 260-280 (longitudes de onda 929-974 nm). Aparece prácticamente en todos los métodos, aunque en este caso algo desplazada hacia menores longitudes de onda. Corresponde al infrarrojo cercano, donde la reflectancia de la vegetación es alta, pudiendo ser afectada por los cambios.

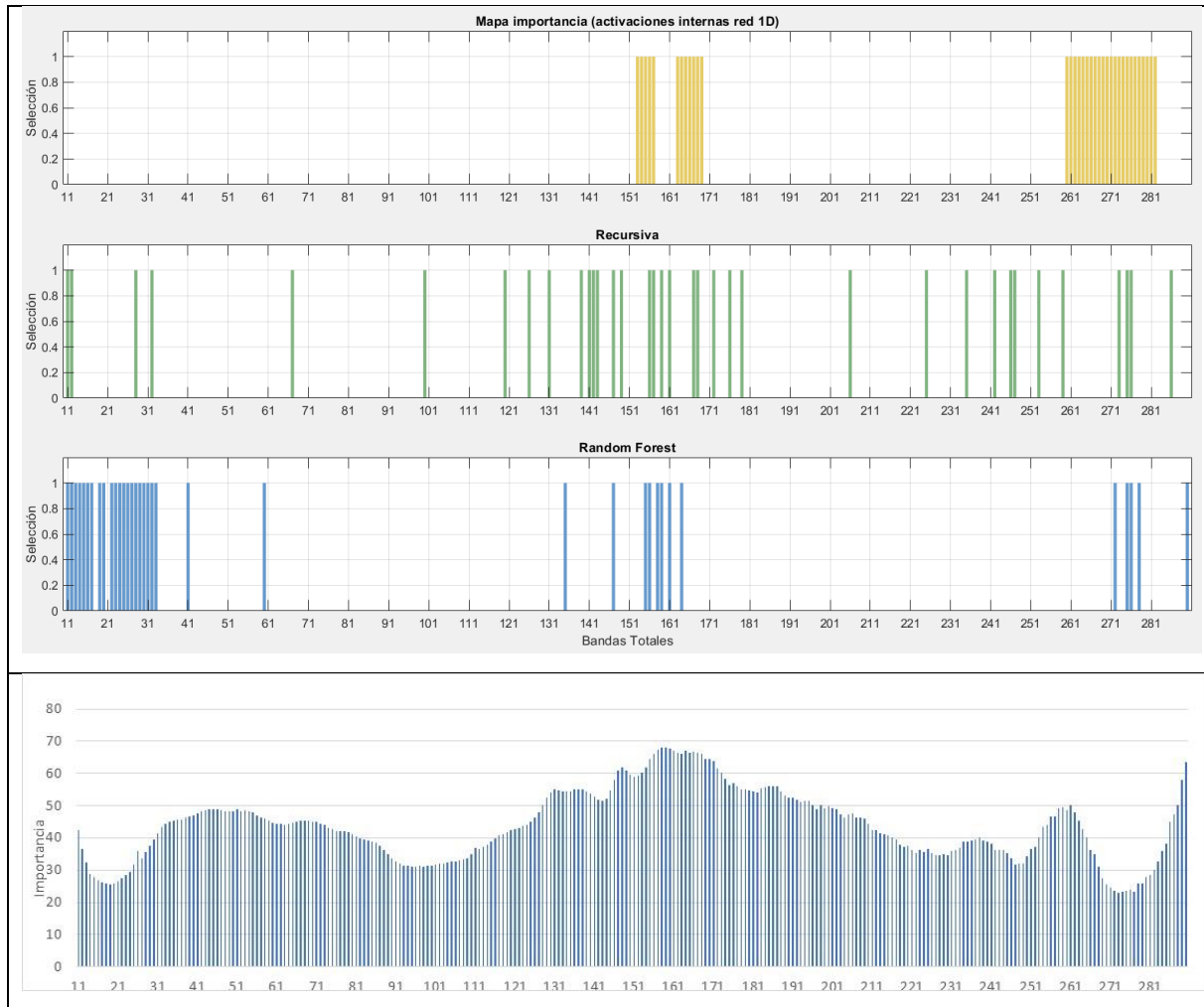


Ilustración 3-53. Selección de características mediante diversos métodos: mapa de importancia (red 1D); eliminación recursiva; bosques aleatorios; análisis de la varianza (ANOVA) para los niveles de tebuconazol.

Si se observan las curvas espectrales (Ilustración 3-54), en las regiones descritas se aprecian diferencias en la reflectancia de las muestras tratadas con distintos niveles de este producto, aunque no de la misma forma. Más específicamente y dejando de lado la primera (azul costero) y la segunda región (verde), que no están claramente definidas, en las otras dos se aprecian bien las diferencias: la región del borde del rojo como un mínimo muy pronunciado; y la región del infrarrojo cercano, en este caso más desplazado hacia longitudes de ondas menores. Ambas son regiones muy afectadas por los cambios en los elementos vegetales.

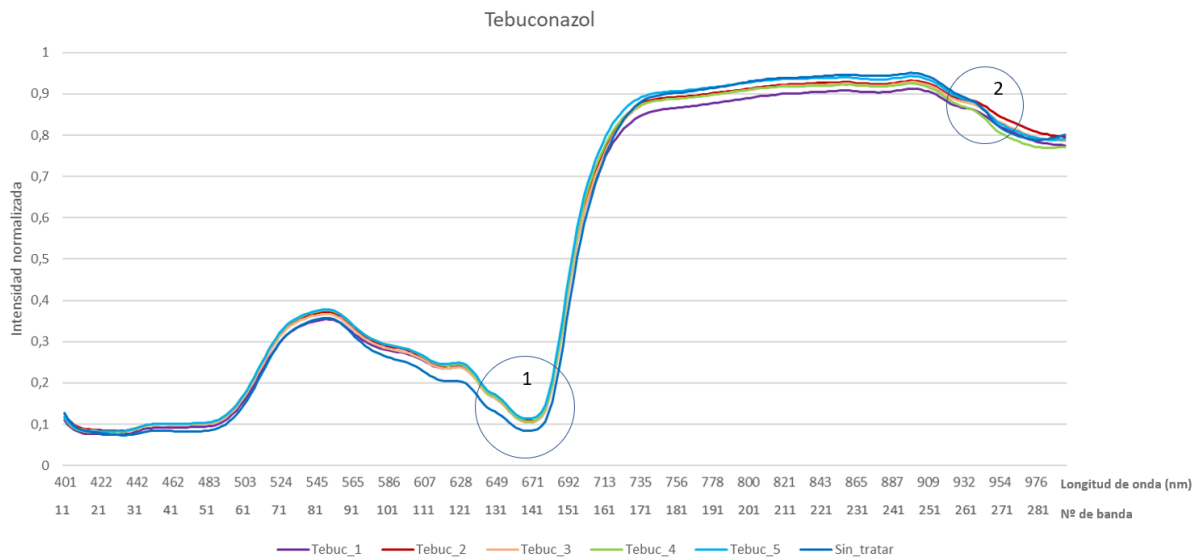


Ilustración 3-54. Regiones del espectro marcadas como más importantes en los métodos de selección de características para diferenciar por niveles de tebuconazol.

Como observaciones generales de todos los métodos de selección de factores en todos los casos analizados, por tipo y por concentración de pesticidas, se pueden extraer las siguientes consideraciones:

- A pesar de examinar imágenes de aceitunas a los que se ha aplicado distintos tipos de productos y niveles de concentración de los mismos, hay una serie de regiones donde se seleccionan en mayor medida las características (bandas espectrales) y que, por lo tanto, son las zonas más sensibles para el análisis. Estas son fundamentalmente la región del azul costero (400-440 nm), la región cercana al verde (520-542 nm), la región del borde del rojo (670-710 nm) y la región del infrarrojo próximo (930-1019).
- Las bandas de estas regiones, y de otras más aisladas, son las que seleccionan para aplicar modelos más simplificados, con buenos resultados en general, aunque en la mayor parte de los casos no se alcanzan los valores de exactitud en las que clasificaciones como aquellos modelos que emplean el espectro completo.
- Dentro de las regiones mencionadas tres de ellas tienen una clara significación física (verde, borde del rojo e infrarrojo), mientras que la otra (azul costero) puede estar más relacionada con su proximidad al extremo del espectro analizado, a pesar de que se han eliminado las 10 primeras y últimas bandas, siguiendo la recomendación de muchos investigadores.

- Llama la atención que en regiones en las que se aprecia una cierta diferencia en las curvas, como el total del infrarrojo cercano (no solo su última parte) y la propia rampa del borde del rojo, donde además hay una clara significación física, no aparecen bandas entre las seleccionadas. Más bien lo hacen en aquellas donde hay máximos, mínimos o puntos de inflexión, por lo que parece que los métodos (red convolucional, bosques aleatorios, etc.) son más sensibles a los cambios de tendencia en las curvas que a los propios valores de estas. Esta puede ser una conclusión importante, que se relaciona con la forma de trabajar de las redes convolucionales, a la hora de detectar patrones profundos en los datos.

3.4. Clasificación de las muestras de hojas

Para el estudio de las hojas se utilizan las conclusiones extraídas en el análisis de las aceitunas, por lo tanto, se entrenarán todas las redes directamente realizando una optimización bayesiana que indique cual es la mejor arquitectura y parámetros de entrenamiento para esta. Se parte de una base de datos elevada, de unas 7000 muestras aproximadamente, que permitirá entrenar correctamente las redes, pero sin existir sobre muestreo.

Para obtener los recortes de las hojas se utilizó la aplicación final, que había sido diseñada en un primer momento para mostrar los resultados finales de la clasificación. Sin embargo, como en el código se implementa una función en la que se pueden obtener definición de las áreas de interés de la imagen, se aprovecha esta función para obtenerlos automáticamente. La aplicación primero segmenta la imagen, separando las hojas del fondo. Debido a que la segmentación de las hojas es mucho más sencilla que la de la aceituna, se realiza con una umbralización, obteniendo resultados adecuados. Una vez se obtiene la máscara de los píxeles de la imagen que pertenecen a las hojas, se generan cuadrículas de 50x50 píxeles de estas máscaras, a los que se les calcula el espectro medio. De esta forma se obtienen espectros medios etiquetados para el entrenamiento de las redes convolucionales en las hojas. Estos datos se dividen en entrenamiento, validación y test, para el correcto entrenamiento y posterior validación y testeo de la red.

A continuación, se presentan los resultados globales de la aplicación de la red convolucional sobre las curvas espectrales extraídas automáticamente en el caso de las hojas (Tabla 3-23), y posteriormente se realizará el análisis detallado mostrando las curvas de evolución de la exactitud y las matrices de confusión de cada uno de los resultados.

Modelo	Entren.	Validac.	Test	Tiempo	Optimiz.
Tipos de producto	96,39	90,82	91,29	4h 40m	Bayesiana
Niveles deltametrina	98,88	97,55	97,24	25m 37s	Bayesiana
Niveles diflufenican	100,00	96,67	95,71	102m 30s	Bayesiana
Niveles lambda-cihal.	99,35	92,42	93,58	77m 32s	Bayesiana
Niveles oxifluorfen	99,67	92,22	95,33	12m 32s	Bayesiana
Niveles tebuconazol	89,56	88,19	88,81	49m 29s	Bayesiana

Tabla 3-23. Resultado para la exactitud (accuracy) de clasificación de las hojas de los datos de entrenamiento, validación y test mediante las redes entrenados

Observando la tabla, se puede determinar que la aplicación de la red convolucional 1D sobre los espectros medios de las ventanas definidas automáticamente produce unos resultados casi perfectos, siendo estos muy cercanos a los obtenidos en el último análisis de la aceituna, en el que se maximizaban los datos. En este caso, no se ha empleado un número tan excesivo de datos, en torno a unas 10 veces inferior, pero se observa que la potencia de reconocimiento de patrones de la red convolucional es suficiente para entrenar y clasificar correctamente con este número de muestras. La exactitud de los datos de validación y test son levemente inferiores a los de entrenamiento, demostrando que no existe sobreajuste. Al igual que en los casos anteriores se ha trabajado con una partición 70/15/15, común para este tipo de estudios.

Realizando un análisis más exhaustivo de los resultados, el primer caso considerado es el de los **tipos de pesticida**, que cuenta con la exactitud más reducida, aunque en torno al 90%. Se observa que el resultado es muy similar obtenido al mejor caso de las aceitunas y este valor de exactitud algo inferior al resto podía preverse debido a que la dificultad para clasificar entre los distintos productos, con distintas concentraciones es más complejo que tan solo discernir entre concentraciones de un mismo pesticida. Para la clasificación de tipos de pesticida en la hoja se obtiene mediante optimización bayesiana una red 1D de 6 capas convolucionales, 220 filtros y un tamaño de filtro de 11.

Es la red más costosa computacionalmente en el entrenamiento, porque el número de datos es más elevado, además de la complejidad que conlleva. Se observa

en la curva de entrenamiento (Ilustración 3-55) que el crecimiento no es tan exponencial, teniendo primeramente un crecimiento lineal hasta estabilizarse en torno al 91% para los datos de validación y 96% para los de entrenamiento.

Se observa en las matrices de confusión (Ilustración 3-55) que hay una mínima confusión generalizada entre las distintas clases a excepción de la sin tratar cuyo error es prácticamente nulo. La confusión más notable aparece entre la clase diflufenican y lambda, siendo normalmente clasificados los primeros erróneamente como los segundos en mayor cantidad que al contrario. Además de ello se puede observar que tanto el tebuconazol como el oxifluorfen (y el diflufenican en menor proporción) tienden a ser clasificados erróneamente como deltametrina.

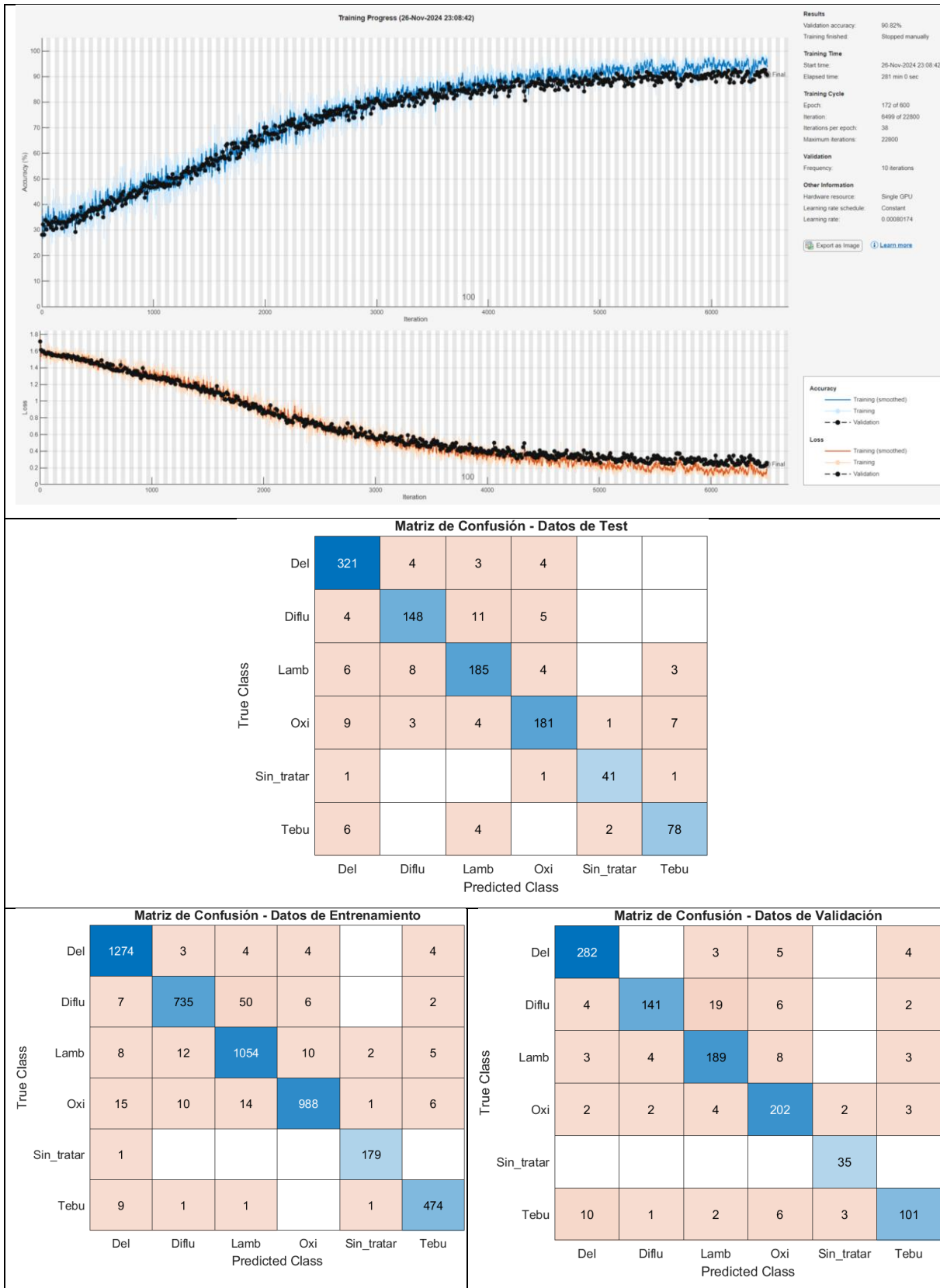


Ilustración 3-55. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para los tipos de producto en los espectros medios de las ventanas definidas automáticamente en las hojas.

En el análisis por niveles de concentración, en el caso de la **deltametrina** se obtiene una clasificación con exactitud casi perfecta, que oscila entre el 97% y el 99% para todos los conjuntos de datos.

En la gráfica de las curvas de exactitud (Ilustración 3-56) se observa un entrenamiento usual. Al inicio hay un crecimiento exponencial, donde los resultados para entrenamiento y validación se separan más, pero a partir de la iteración 1500 el modelo se ajusta, lo que provoca ambas curvas se estabilicen y se igualen en torno al 97% de exactitud.

Las matrices (Ilustración 3-56) muestran una clasificación casi perfecta, con confusiones residuales entre los distintos niveles.

Para el **diflufenican**, los resultados obtenidos son levemente inferiores que para la deltametrina, pero en todos los casos con una exactitud superior al 95%, lo que indica que la clasificación es satisfactoria y no existe sobreajuste.

La curva de entrenamiento sigue una evolución normal (Ilustración 3-57), pese a que, comparándola con el caso anterior, la curva de validación y entrenamiento una vez estabilizadas tienen una diferencia del torno al 3-4%. Sin embargo, esto no se considera preocupante puesto que se demuestra que no hay sobreajuste.

Las matrices de confusión (Ilustración 3-57) muestran solo pequeñas confusiones entre los niveles 1 y 2, ambos por encima del límite europeo.

Analizando los resultados obtenidos para el **lambda-cihalotrina**, los resultados obtenidos para validación y entrenamiento descienden levemente respecto de los anteriores; sin embargo, se supera el 92% para todos los conjuntos, lo que indica que la clasificación es satisfactoria y no existe sobreajuste.

La curva de entrenamiento es muy similar a las anteriores comentadas (Ilustración 3-58), aunque se puede observar una mayor separación entre los datos de entrenamiento y test, tal y como se puede comprobar en los resultados de exactitud.

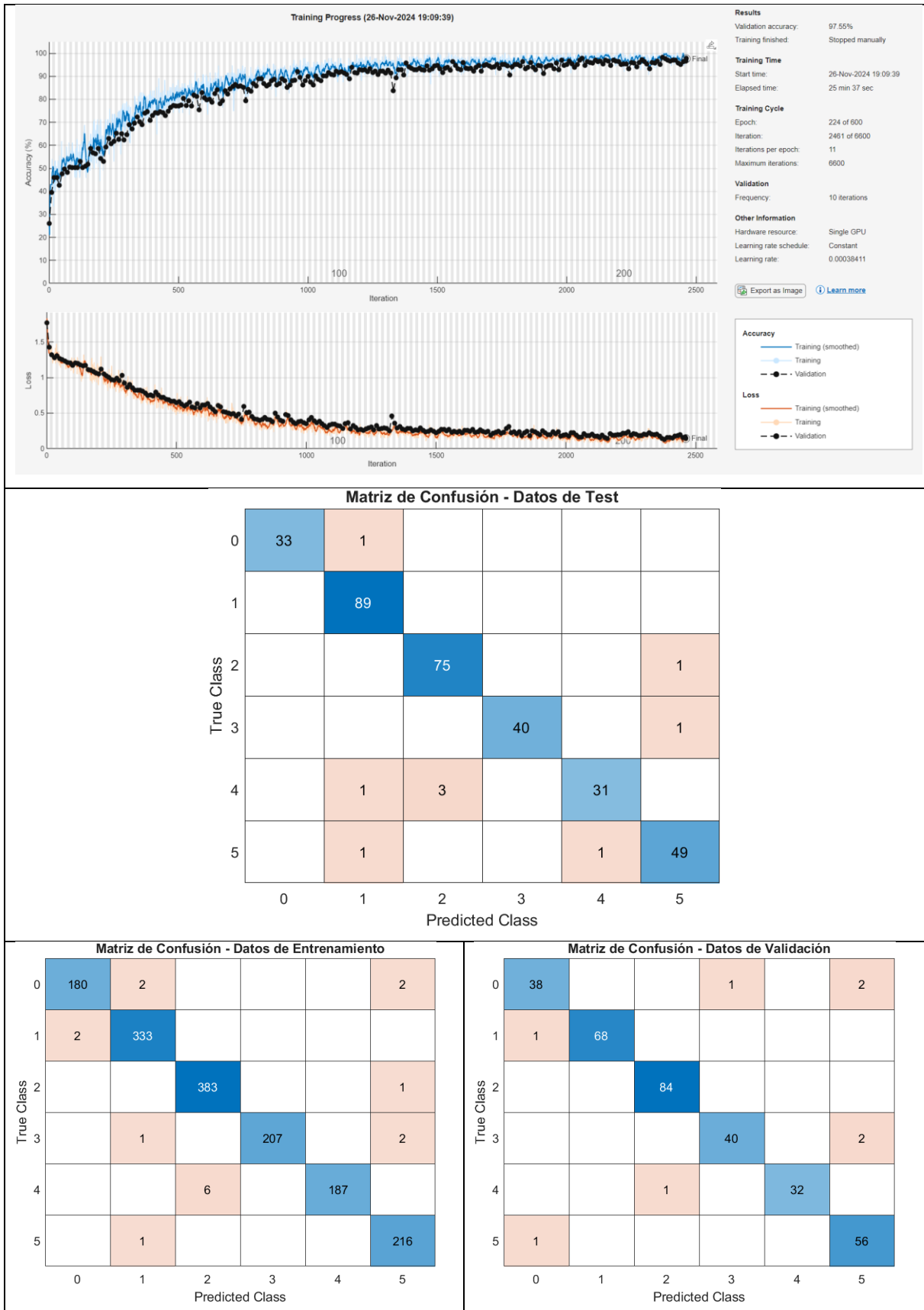


Ilustración 3-56. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para el nivel de deltametrina en los espectros medios de las ventanas definidas automáticamente en las hojas.

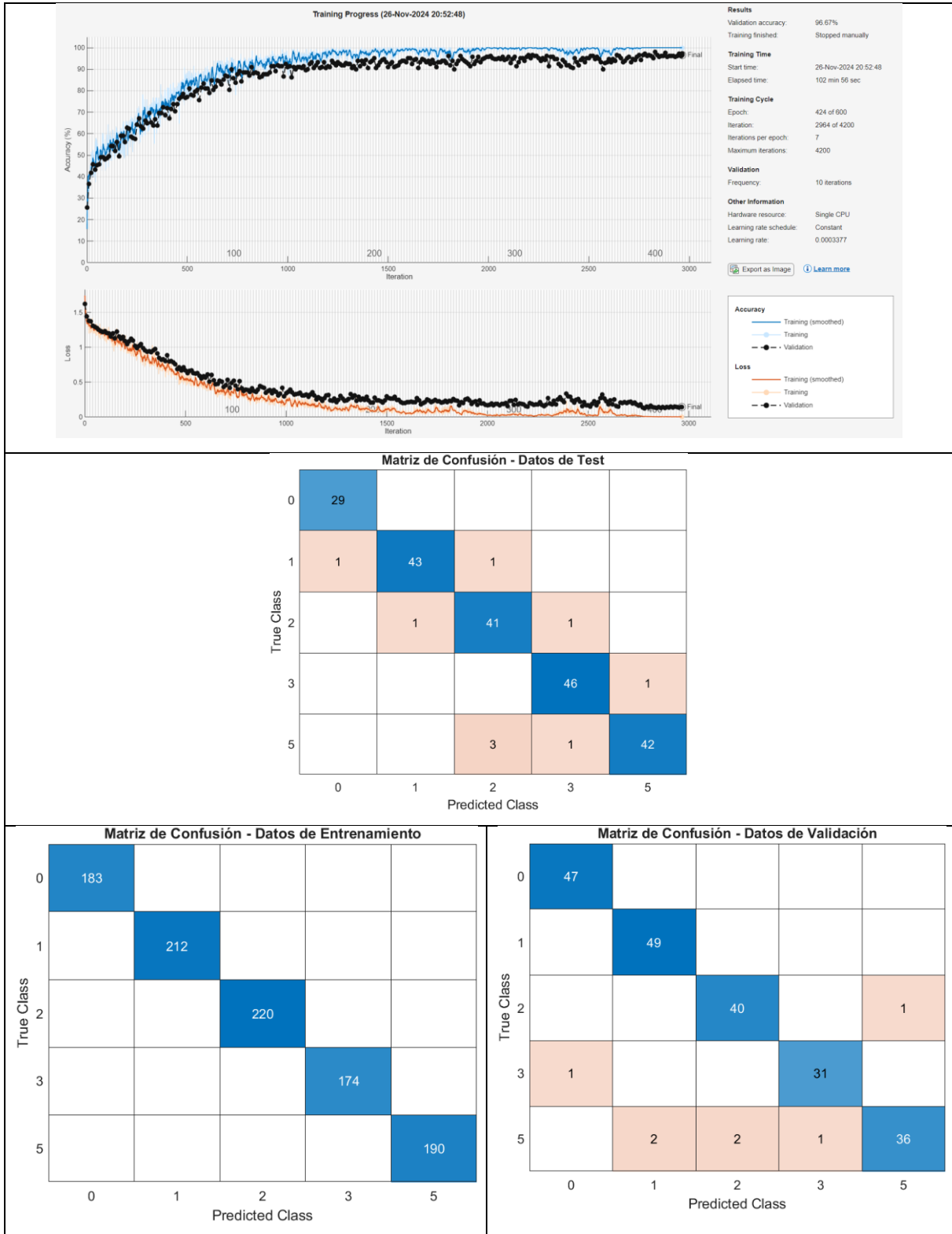


Ilustración 3-57. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para el nivel de diflufenican en los espectros medios de las ventanas definidas automáticamente en las hojas.

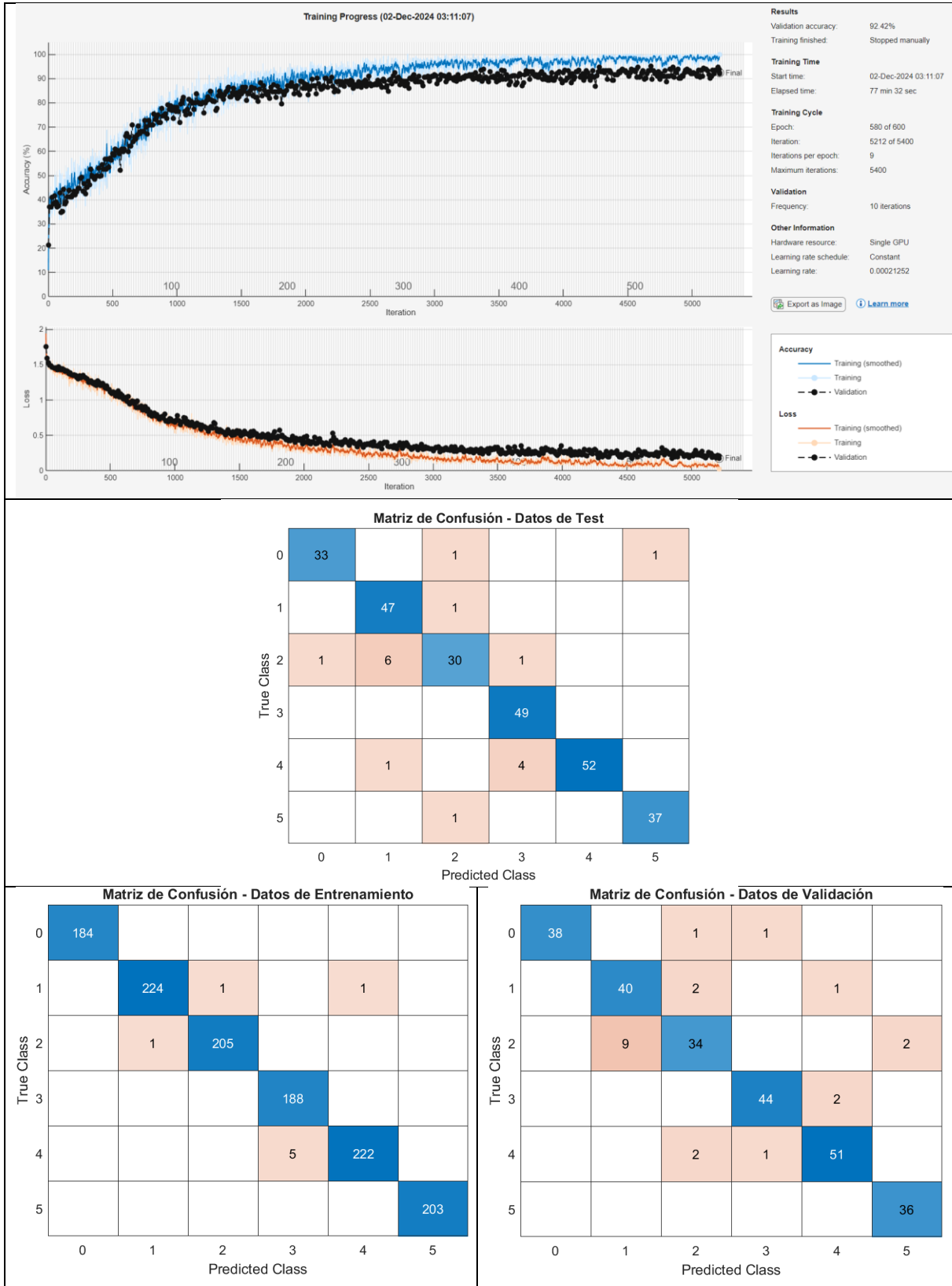


Ilustración 3-58. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para el nivel de lambda-cihalotrina en los espectros medios de las ventanas definidas automáticamente para las hojas.

Analizando las matrices (Ilustración 3-58), las clases entre las que hay una mayor confusión son entre la 2 y la 3 (que son contiguas) y la 4 y 5, que, además tienen niveles de concentración mínimos. En ambos casos, no hay problema ya que las confusiones se producen entre niveles por encima o por debajo del límite europeo.

Para el **oxifluorfen** se obtienen asimismo buenos resultados, aunque entre el entrenamiento y la validación se experimenta un descenso de exactitud del 6% aproximadamente. Sin embargo, los datos de test tienen un valor superior al 95%, por lo que no existe sobreajuste y la red está funcionando correctamente.

Esto queda expuesto en la curva de entrenamiento (Ilustración 3-59) donde se observa que las curvas de los datos de validación y entrenamiento se encuentran con una separación del torno al 5% durante todo el proceso. Hay un pequeño rizado en el entrenamiento que es algo superior en la validación.

Analizando las matrices (Ilustración 3-59), la clase con más error es la 3 (concentración intermedia), que presenta una confusión generalizada con el resto de clases en el caso de los datos de validación. Sin embargo, esta es mínima y desaparece casi al completo para el resto de conjuntos de datos.

Por último, el **tebuconazol** muestra unos resultados peores que el resto de productos, siguiendo la tendencia de este producto durante todo el análisis, por lo que se podría determinar que este es el pesticida cuya presencia es más difícil de detectar. Pese a ello, los resultados para todos los conjuntos se encuentran entre el 88 y 90.

Observando la curva de progreso de entrenamiento (Ilustración 3-60), en primer lugar, se observa que el crecimiento es más lento y menos pronunciado que en el resto de casos. Una vez se estabilizan las curvas, estas tienen un rizado de un 5% en entrenamiento y 10% en validación, haciendo esta red la menos estable.

En las matrices (Ilustración 3-60) se puede observar que hay más errores en la clasificación que en los anteriores casos. Así, cabe destacar que la concentración 2 es aquella que causa más confusión, aunque se producen fundamentalmente con el nivel 1, ambos bajo el límite europeo.

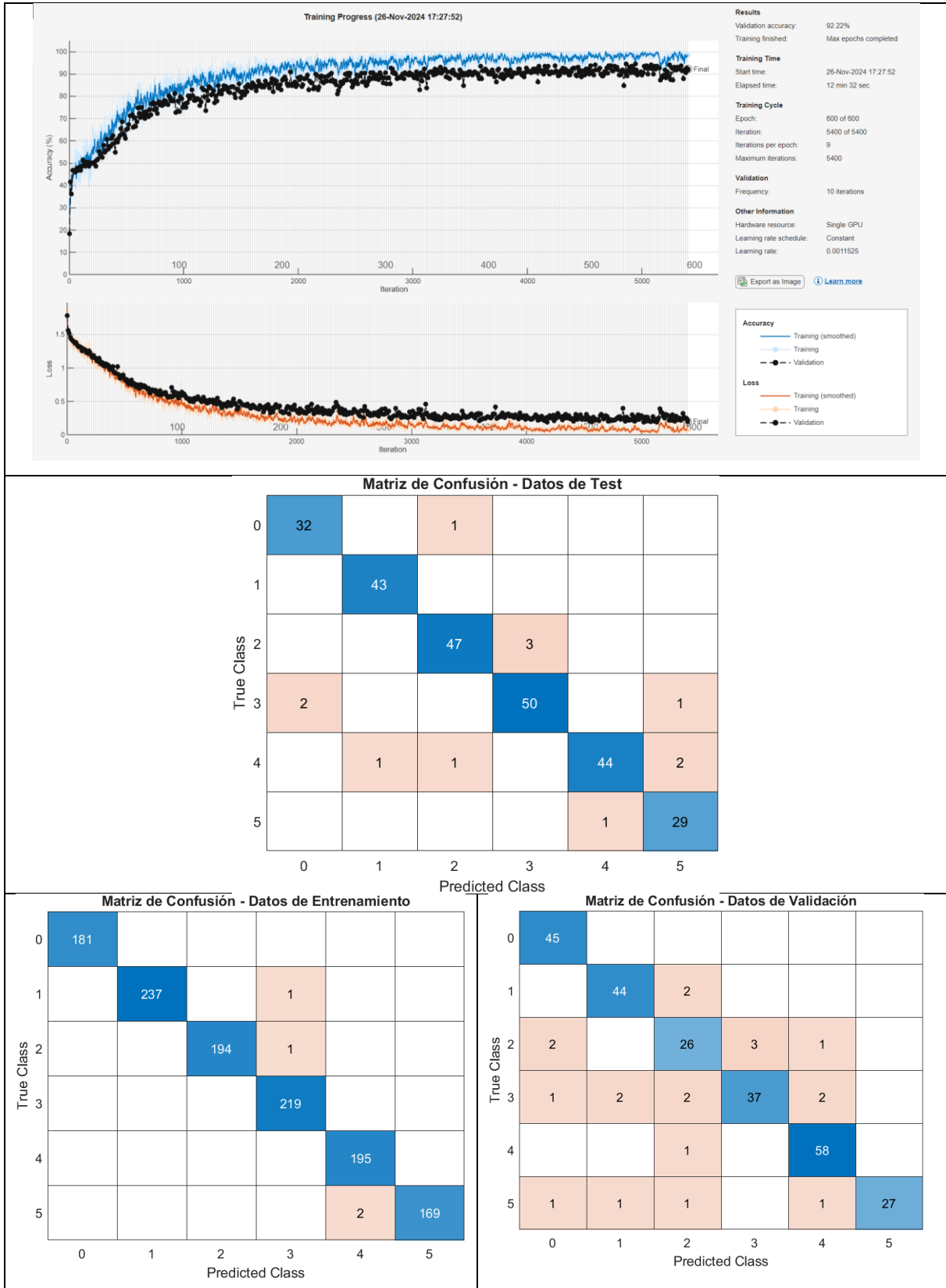


Ilustración 3-59 Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para el nivel de oxifluorfen en los espectros medios de las ventanas definidas automáticamente para las hojas.

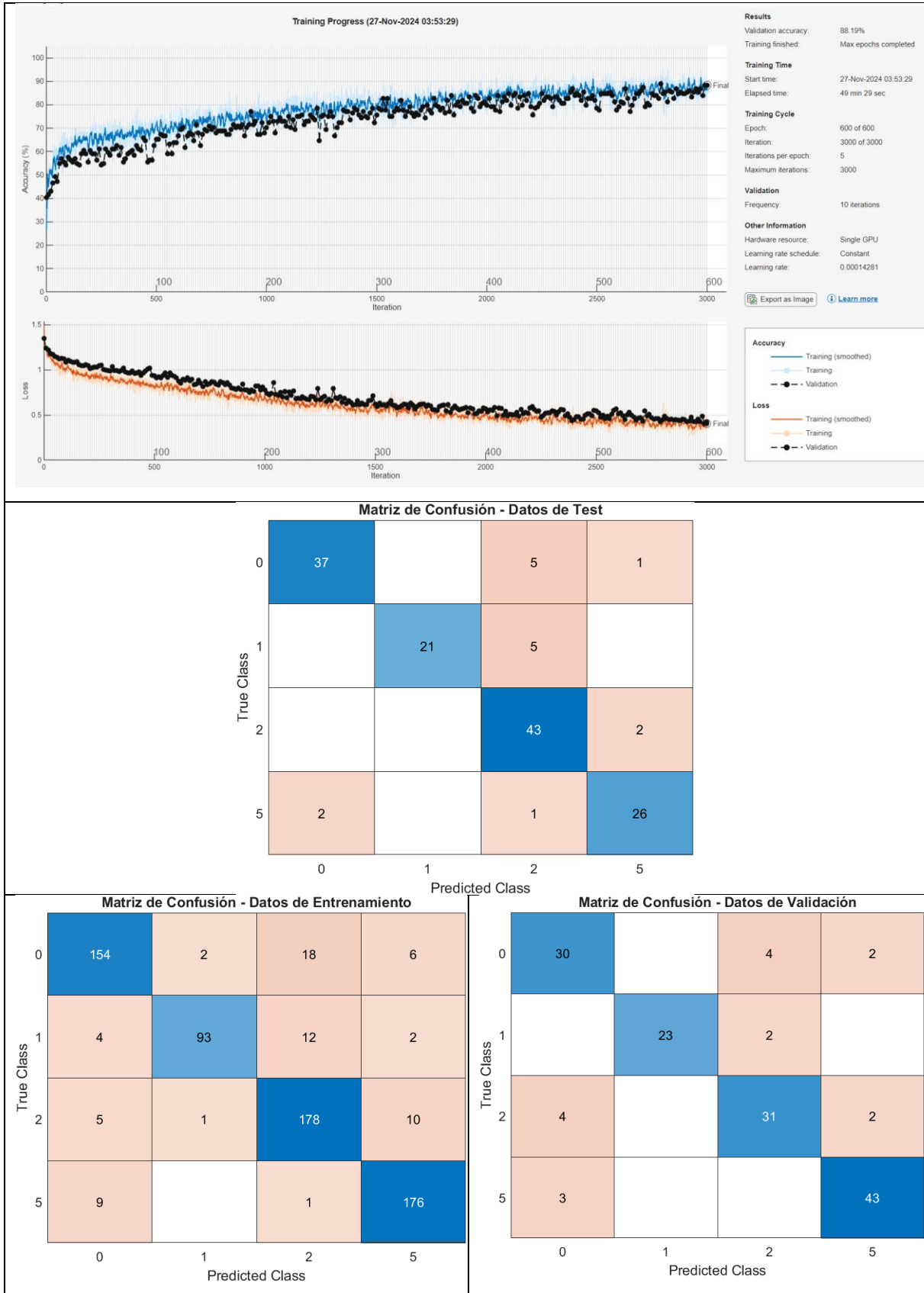


Ilustración 3-60. Proceso de entrenamiento y matrices de confusión de los datos de test, entrenamiento y validación de la red convolucional para el nivel de tebuconazol en los espectros medios de las ventanas definidas automáticamente en las hojas

3.5. Selección de características de las hojas

En el caso del análisis de las hojas, se optó por realizar la selección de características utilizando el método menos costoso desde el punto de vista computacional. Esta decisión se fundamenta en el elevado volumen de datos, el cual podría hacer que el uso de las técnicas previamente analizadas como los bosques aleatorios (*random forest*) o la eliminación recursiva de características implicara un tiempo de procesamiento considerablemente largo.

Por esta razón, se utilizó el método de análisis de activaciones de la capa 1-D o mapa de características, el cual tiene la ventaja de ser computacionalmente más eficiente. Tal como se mostrará a continuación y en el análisis correspondiente a las aceitunas, este método proporciona resultados de calidad comparable con los obtenidos por técnicas más complejas, e incluso similares a los resultados obtenidos cuando se consideran todas las características sin realizar ninguna selección previa.

En la Tabla 3-24 se muestra el resultado de las redes anteriormente obtenidas, reentrenadas teniendo solo en cuenta las características seleccionadas (bandas) para cada una de las muestras. Al igual que con las aceitunas, se toma un 90% del total de las características, que representan un total de 30 bandas.

Modelo	Entren.	Validac.	Test	Tiempo	Optimiz.
Tipos de producto	99,34	95,12	95,41	34m 46s	Bayesiana
Niveles deltametrina	99,28	96,93	95,71	21m 48s	Bayesiana
Niveles diflufenican	99,79	95,71	93,81	113m 49s	Bayesiana
Niveles lambda-cihal.	97,49	93,18	92,45	55m 42s	Bayesiana
Niveles oxifluorfen	98,58	93,77	94,16	17m 23s	Bayesiana
Niveles tebuconazol	88,08	86,81	86,01	10m 06s	Bayesiana

Tabla 3-24. Resultado para la exactitud (accuracy) de clasificación de las hojas de los datos de entrenamiento, validación y test mediante las redes entrenadas tras la selección de características

En ella se muestra las 30 bandas seleccionadas (10% de las totales) para cada una de las redes entrenadas: 5 redes para la clasificación de las concentraciones de los distintos pesticidas y 1 para diferenciar entre tipos de pesticidas.

En el análisis conjunto de todos los productos y bandas, que se muestra en la Ilustración 3-61, se pueden extraer las siguientes observaciones:

- Bandas 120-150 (longitudes de onda 626-690 nm). Corresponde preferentemente a la región del borde del rojo (*red edge*) que, como se indicó anteriormente, es bastante sensible a los cambios en los elementos de la vegetación.
- Bandas aisladas entre la 170-190 (733-776 nm) o por encima de la 200 (800 nm), ya en el infrarrojo cercano, igualmente una región muy importante para definir el estado de la vegetación.

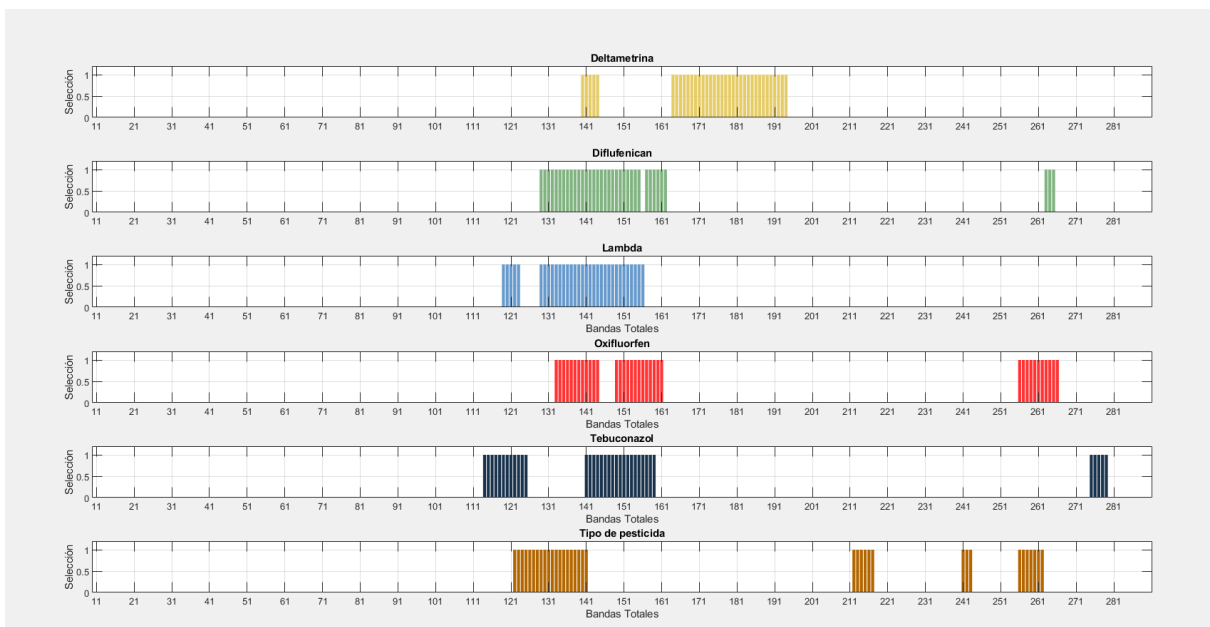


Ilustración 3-61. Bandas seleccionadas para cada una de las redes entrenadas

4. CONCLUSIONES

Este trabajo tiene como objetivo la detección y clasificación del tipo de productos fitosanitarios y, sobre todo, su concentración en aceitunas y hojas de olivo, a partir de imágenes hiperespectrales y sus correspondientes curvas espectrales medias. Para ello se ha contemplado la aplicación de las siguientes técnicas:

- Aprendizaje profundo sobre las muestras de aceitunas previamente etiquetadas, desarrollando redes neuronales convolucionales (CNN) 1D, que son las habitualmente recomendadas en la amplia bibliografía consultada para la realizar la clasificación mencionada.
- Los resultados que se han ido obteniendo han conducido a distintas aproximaciones al muestreo de los datos, comenzando por la obtención de los espectros medios de las aceitunas; pasando después a la extracción de los espectros medios en ventanas definidas manualmente, con el fin de aumentar el tamaño de la muestra; y finalmente a la extracción de los espectros medios sobre cuadrículas generadas automáticamente.
- Sobre todos estos conjuntos de datos se han aplicado además distintos métodos de clasificación mediante aprendizaje automático, con el propósito de comparar y comprobar los resultados obtenidos con el aprendizaje profundo.
- Adicionalmente, se ensayaron varios métodos de selección de características, y reducción de la dimensionalidad, evaluando y comparando los resultados, con la finalidad de extraer conclusiones sobre las bandas más influyentes en la detección de las distintas concentraciones y tipos de fitosanitarios.
- Como culminación del trabajo, se ha desarrollado una aplicación automática que clasifique las concentraciones y tipos de pesticida con el que se han tratado las aceituna u hojas, por zonas de la imagen.
- Como tarea inicial fue necesario segmentar la parte sana de las aceitunas y hojas de otros elementos de la imagen. En el caso de las aceitunas se entrenó una red CNN 2D con las imágenes RGB, cuyos resultados fueron óptimos, con una exactitud superior al 90%. Se decidió hacerlo con este método debido a la complejidad de estas imágenes, en las que había fondos de distintos colores, defectos en las aceitunas, etc. Las hojas, por su parte, se segmentaron mediante una umbralización a partir del índice NDVI.

En cuanto a los resultados obtenidos a partir de los espectros medios de las aceitunas se extraen las siguientes conclusiones:

- El aprendizaje automático es más sencillo de implementar, desde la interfaz desde el *Classification Learner* y sus tiempos de entrenamiento son los más reducidos, entre 10 y 20 segundos.
- En el tipo de pesticida aparecen buenos resultados con métodos sencillos (análisis discriminante y k-NN), con los que se obtienen exactitudes de validación y test superiores al 80%.
- En la clasificación por concentraciones llama la atención que en algunos casos se alcanzan exactitudes de la muestra de test de hasta el 100%, lo que se puede deber al número reducido de muestras. Con los datos de validación, los métodos con mejores resultados son el análisis discriminante y SVM, y los más mediocres son k-NN y los árboles de decisión. Por productos, se observó que los peores resultados aparecen para la deltametrina, (60%), mientras que para el resto se supera el 80%.
- La aplicación de una red CNN-1D con tres capas convolucionales y ajuste manual, emplea un tiempo de entrenamiento del orden de minutos, lo que no es un problema importante, puesto que la red, una vez entrenada, clasifica instantáneamente. Los resultados son equiparables y en algunos casos superiores a los obtenidos con el aprendizaje automático, con exactitudes más altas en la clasificación por nivel de concentración que por tipo de pesticida.
- El escaso número de datos de la muestra hace sin embargo que los resultados no sean consistentes cada vez que se entrena la red. Esto se debe a que la curva de exactitud del entrenamiento sube rápidamente casi al 100%, en un caso claro de sobreajuste, con un entrenamiento deficiente y valor bajo en la exactitud en los datos de validación y test.
- La **conclusión final para este primer análisis** es que el uso de los espectros medios de las aceitunas complejas no es aconsejable en este caso puesto que el número de muestras no es suficiente para entrenar correctamente la red convolucional 1-D, la cual se sobreajusta. El número de muestras puede ser más adecuado para técnicas de aprendizaje automático como el análisis discriminante o SVM.

Debido a lo explicado, el siguiente paso fue aumentar los datos, para lo que se desarrolla una aplicación en la que el usuario selecciona manualmente ventanas de la imagen. Se tomaron un total de 10 muestras por imagen, multiplicando así el conjunto de datos por 10. Los resultados obtenidos son:

- En el aprendizaje automático se mantienen los tiempos de entrenamiento reducidos (alrededor de los 20 segundos). Los resultados son similares a los obtenidos en los espectros medios, aunque mucho más coherentes, debido al mayor tamaño de la muestra.
- Así, para la clasificación atendiendo al tipo de producto se encuentran valores de exactitud entre el 60% (árbol de decisión) y el 90% (análisis discriminante). En los niveles de producto, hay aún mayor variabilidad, aunque los mejores resultados se alcanzan de nuevo en el análisis discriminante, SVM y las redes neuronales, y los peores en k-NN y los árboles de decisión. Llama la atención los buenos resultados de métodos sencillos como el análisis discriminante, pero este hecho se produce igualmente en otros trabajos consultados.
- En las redes CNN-1D, se obtienen exactitudes muy satisfactorias para todos los casos y superiores a los obtenidos anteriormente. Así para diflufenican y oxifluorfen (herbicidas), así como para deltametrina y lambda-cihalotrina (insecticidas) se obtienen valores de exactitud del 90%; para la clasificación entre los distintos tipos, del 85%; y para tebuconazol (fungicida), del 80%.
- Debido a su mayor complejidad, en algunos casos se utiliza una optimización bayesiana que ajusta el número de capas convolucionales, tamaño y número de filtros, así como los hiperparámetros de la red.
- La gráfica de exactitud de entrenamiento y validación, así como las matrices de confusión determinan que los resultados son consistentes y no hay sobreajuste, aunque los tiempos de entrenamiento son mucho mayores (20 y 50 minutos); esto no representa un problema, como ya se ha indicado antes.
- Por lo tanto, un aumento de datos permite la obtención de mejores resultados, pero sobre todo más fiables y consistentes. Además, se puede sacar un mayor partido a las redes respecto a los métodos de aprendizaje automático. Y si se consideran los límites europeos para determinar si una muestra está o no tratada, la red puede discernir apenas sin error entre ambos conjuntos.

Por último, con la intención de automatizar el problema y también de mejorar los resultados, se realiza un tercer proceso de entrenamiento, para el que, en lugar de obtener ventanas manualmente, se extraen automáticamente cuadrículas de la imagen. En cada una de ellas se calcula el espectro medio, aumentando en unas 30 veces los datos usados en el caso anterior (65800). Los resultados se resumen en:

- Para los métodos de aprendizaje automático se obtienen resultados muy satisfactorios, alcanzándose valores superiores al 90% para la mayor parte de las clasificaciones. Vuelve a destacar el análisis discriminante, la red neuronal y SVM, con resultados algo inferiores en los árboles de decisión.
- El aumento de la muestra ha conseguido mejorar alguno de los resultados, manteniendo la coherencia con los obtenidos para los espectros medios y las ventanas de extracción manual, pero el tiempo de entrenamiento asciende considerablemente hasta casi 10 minutos e incluso hasta 1 hora.
- Las redes convolucionales 1D utilizaron en varios casos la optimización bayesiana para definir la arquitectura de la red y los parámetros de entrenamiento, obteniendo redes de 3 a 6 capas convolucionales, con tiempos de entrenamiento que fluctúan entre 20 minutos a 2h en algunos de ellos.
- Tanto para la clasificación por tipos como por concentraciones se supera el 90%, excepto para el tebuconazol que queda en el 80%. Pese a los excelentes resultados obtenidos, se observa que, en general, son iguales o algo inferiores a los obtenidos con alguno de los métodos de aprendizaje automático, en especial el análisis discriminante. Esto resulta algo sorprendente y hace pensar en algún tipo de sobreajuste, aunque estudios realizados con los mismos datos de partida (Martínez-Gila, et al., 2024) son consistentes con estos.
- No obstante, si se consideran estos resultados junto con los del muestreo anterior, se puede concluir que las redes convolucionales 1D con un conjunto de datos suficiente (ventanas) y con una arquitectura y parámetros de entrenamiento correctos (optimización bayesiana o ajuste manual) producen resultados bastante estables y consistentes, sin necesidad de recurrir a una muestra demasiado grande que dificulte los procesos. Son muy satisfactorios los valores de exactitud en el test, superando el 80-90% y discriminando prácticamente a la perfección entre los niveles sobre y bajo el límite europeo.

El análisis de las hojas, con el objeto de agilizar el proceso, se realizó únicamente con los métodos que se consideraron óptimos para las aceitunas. Por ello, se hizo un muestreo automático con cuadrículas, pero en este caso, se tomaron unas 6500, una cantidad suficiente para un entrenamiento consistente, pero no demasiado elevada para evitar excesivos tiempos de entrenamiento. Los resultados se resumen en:

- Los valores de exactitud en test son excelentes, superando con facilidad el 90% para todos los casos y alcanzando este mismo porcentaje en el peor de los productos analizados, que nuevamente es el tebuconazol. Este parece ser el producto cuya concentración es más compleja de determinar, puesto que sistemáticamente presenta peores resultados en aceitunas y hojas.
- Los tiempos de entrenamiento oscilan entre unos 20 minutos hasta 4h en el caso más complejo. Se aplica una optimización bayesiana para la construcción de todas las redes, obteniendo redes 1-D con entre 3 y 6 capas convolucionales, cuyas gráficas de exactitud muestran en todos los casos procesos de entrenamiento y resultados de validación correctos.
- Adicionalmente, en la mayoría de los casos aparece una diferencia algo mayor entre los resultados de entrenamiento y validación y test, que los obtenidos anteriormente. Reduciendo la complejidad de las redes se podría disminuir estas diferencias; sin embargo, esto disminuiría también los valores de exactitud obtenidos, lo que no se considera necesario puesto que las redes no se pueden catalogar ni siquiera como sobreajustadas y los resultados son muy satisfactorios.

Finalmente, para la selección de características se ensayan diferentes métodos, ANOVA y análisis de componentes principales (PCA) para el aprendizaje automático; y análisis de activaciones de las redes convolucionales, eliminación recursiva y *random forest* para las redes 1D. Las conclusiones en este aspecto son las siguientes:

- En el aprendizaje automático de las aceitunas para evitar una reducción significativa de la exactitud es necesario recurrir a una varianza del 99% en el PCA, apareciendo entre 4 y 6 componentes principales; y mantener 30 características con ANOVA. Con estas condiciones, los resultados son aceptables, sobre todo en el caso del muestreo con ventanas automáticas.

- Las técnicas empleadas para la selección de características de las redes 1D son más complejas, sobre todo el método de eliminación recursiva y *random forest*, necesitando largos tiempos de procesamiento de alrededor de 1,5 h, por lo que solo se realiza el estudio para el que se considera el muestreo óptimo (ventanas). Al igual que con ANOVA, manteniendo un número de características igual a 30, se consiguen resultados de exactitud muy satisfactorios. Los resultados son aceptables para todos los métodos excepto en el tebuconazol, corroborando la complejidad que presenta su clasificación.
- Analizando las bandas seleccionadas por los distintos métodos se observan una serie de regiones con mayor poder de discriminación en los tipos y niveles de producto: azul costero (400-440 mm), región cercana al verde (520-542 mm), región del borde del rojo (670-710 mm) y región del infrarrojo próximo (930-1019 mm); en ellas se seleccionan en mayor medida las características.
- Dentro de estas regiones, tres de ellas (verde, rojo e infrarrojo) tienen una clara significación física en muestras de elementos de la vegetación, mientras que la del azul costero puede deberse más a su proximidad con el extremo del espectro asociado, pese a haber eliminado las 10 primeras y últimas bandas.
- Sin embargo, llama la atención de que zonas donde se puede apreciar mayor diferencia, incluso visualmente, como el infrarrojo cercano en su totalidad (y no solo la última parte) y la propia rampa del borde rojo, no aparecen bandas seleccionadas, sino que estas aparecen en aquellas regiones con máximos, mínimos o puntos de inflexión. Así, estos métodos parecen ser más sensibles a los cambios de tendencias en las curvas que a los propios valores. Esta idea puede ser relevante, ya que conecta con el modo en que las redes convolucionales funcionan al identificar patrones complejos en los datos.
- En el caso de las muestras de las hojas se empleó el método óptimo en relación exactitud-tiempo de procesamiento, que es el mapa de importancia derivado de la red convolucional 1D. En este análisis se obtienen resultados excelentes, manteniendo una diferencia no superior al 2% en la exactitud obtenida con todas las bandas, e incluso aumentándola.
- La región más significativa corresponde al borde del rojo (626-690 nm), que como en las aceitunas es bastante sensible a los cambios, así como el infrarrojo cercano, en el que aparecen algunas bandas aisladas.

Finalmente, como mejoras del trabajo se podrían considerar las siguientes:

- Disponer de un número mayor de muestras (imágenes) de aceitunas, con objeto de aplicar el entrenamiento sobre las aceitunas u hojas completas con su espectro medio, ya que el objetivo de la clasificación es clasificar estas muestras completas para validar o desechar las mismas en el control de calidad.
- Analizar el comportamiento de las redes convolucionales en problemas más complejos como la mezcla de productos, incluyendo no solo pesticidas (herbicidas, insecticidas y fungicidas) sino fertilizantes, plaguicidas, o incluso la detección de enfermedades, patógenos, disfunciones, etc.
- Considerar la utilización de verdaderas redes convolucionales 3D que permitieran realizar el análisis por partes o segmentos de los órganos estudiados, refinando así el proceso de clasificación en los casos que fuera de interés. Este tipo de redes, no obstante, requieren de un procesamiento fuera del alcance de los ordenadores personales, por lo que su entrenamiento habría que llevarlo a cabo en superordenadores, aunque una vez entrenadas las redes se podrían aplicar en ordenadores convencionales.
- Automatizar el proceso, no solo en lo planteado en este trabajo en la aplicación desarrollada para visualizar la clasificación, sino para realmente implementar las técnicas de reconocimiento y clasificación en una cadena de producción, con las imágenes obtenidas en tiempo real, por ejemplo, en una cinta transportadora.

Bibliografía

- Adán, A., Merchán, P. & Salamanca, S., 2016. Representación de la información 3D. En: E. Alegre, G. Pajares & A. Escalera, eds. *Conceptos y métodos en Visión por Computador*. s.l.:Grupo de Visión del Comité Español de Automática, p. 201.
- Afonso, M. y otros, 2020. Tomato Fruit Detection and Counting in Greenhouses Using Deep Learning. *Front Plant Sci.*, Volumen 11, p. 571299.
- Agarwal, M. y otros, 2020. Identification and diagnosis of whole body and fragments of *Trogoderma granarium* and *Trogoderma variabile* using visible near infrared hyperspectral imaging technique coupled with deep learning. *Comput Electron Agric*, Volumen 173, p. 105438. <https://doi.org/10.1016/J.COMPAG.2020.105438>.
- Aggarwal, M. y otros, 2023. Pre-Trained Deep Neural Network-Based Features Selection Supported Machine Learning for Rice Leaf Disease Classification. *Agriculture*, Volumen 13, p. 936. <https://doi.org/10.3390/agriculture13050936>.
- Agriopoulou, S. y otros, 2021. Authentication and Chemometric aDiscrimination of Six Greek PDO Table Olive Varieties through Morphological Characteristics of Their Stones. *Foods*, Volumen 10, p. 1829.
- Aguilera Puerto, D., Cáceres Moreno, Ó. & Martínez Gila, D. a., 2019. Online system for the identification and classification of olive fruits for the olive oil production process. *Food Measure*, Volumen 13, pp. 716–727. <https://doi.org/10.1007/s11694-018-9984-0>.
- Aguilera Puerto, D., Martínez Gila, D., Gámez García, J. & J., O., 2015. Sorting olive batches for the milling process using image. *Sensors*, Volumen 15, p. 15738–15754.
- Alegre, E. & Fernández-Robles, L., 2016. SIFT (Scale Invariant Feature Transform). En: E. Alegre, G. Pajares & A. Escalera, eds. *Conceptos y métodos en Visión por Computador*. s.l.:Grupo de Visión del Comité Español de Automática, p. 131.
- Ang, K. & Seng, J., 2021 . Big data and machine learning with hyperspectral information in agriculture. *IEEE Access*, Volumen 9, p. 36699–36718. <https://doi.org/10.1109/ACCESS.2021.3051196>..
- Aquino, A., Ponce, J. M. & Andújar, J., 2020. Identification of olive fruit, in intensive olive orchards, by means of its morphological structure using convolutional neural networks. *Computers and Electronics in Agriculture*, Volumen 176, p. 105616.
- Asensio Jiménez, P. y otros, 2022. Predicción automática de la carga frutal de olivos empleando UAV y redes convolucionales. *XLIII Jornadas de Automática: libro d libro de actas*, pp. 956-963 <https://doi.org/10.17979/spudc.9788497498418.0956>.
- Bai, X. y otros, 2022. Evaluation of rice bacterial blight severity from lab to field with hyperspectral imaging technique. *Front Plant Sci*, Volumen 13, p. 1037774. <https://doi.org/10.3389/FPLS.2022.1037774/BIBTEX>.
- Bannari, A., Morin, D. & Bonn, F., 1995. A review of vegetation indices. *Remote Sensing Reviews*, 13(1), p. 95 — 120.

- Barbedo, J., 2023. A review on the combination of deep learning techniques with proximal hyperspectral images in agriculture. *Comput Electron Agric*, Volumen 210, p. 107920.
- Barbedo, J., Tibola, C. & Fernandes, J., 2015. Detecting Fusarium head blight in wheat kernels using hyperspectral imaging. *Biosyst Eng*, Volumen 131, p. 65–76. <https://doi.org/10.1016/J.BIOSYSTEMSENG.2015.01.003>.
- Belgiu, M. & Drăgu, L., 2016. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, Volumen 114, pp. 24-31.
- Beltrán Ortega, J. y otros, 2016. Novel technologies for monitoring the in-line quality of virgin olive oil during manufacturing and storage. *J Sci Food Agric.*, 96(14), pp. 4644-4662.
- Benkler, Y., 2012. *El Pingüino y el Leviatán: Por qué la cooperación es nuestra arma más valiosa para mejorar el bienestar de la sociedad*. s.l.:s.n.
- Beucher, S., 1992. *The Watershed Transformation Applied To Image Segmentation*, France: Centre de Morphologie Mathematique.
- Bhargava, A. y otros, 2024. Hyperspectral imaging and its applications: A review. *Heliyon*, 10(12), p. e33208. doi: 10.1016/j.heliyon.2024.e33208. PMID: 39021975; PMCID: PMC11253060.
- Bishop, C., 2006. *Pattern Recognition and Machine Learning*. New York Inc.: Springer-Verlag.
- Bosch, A., Casas, J. & Lozano, T., 2019. *Deep Learning: Principios y fundamentos*. s.l.:Reverté-Aguilar.
- Brakenridge, G., Anderson, E. & Nghiem, S., 2007. Satellite microwave detection and measurement of river floods. En: *AGU Spring Meeting Abstracts*. s.l.:s.n., pp. H23A-05.
- Breiman, L., 2001. Random Forests. *Machine Learning*, 45(1), pp. 5-32.
- Breiman, L., Friedman, J., Olshen, R. & Stone, C., 1986. *Classification and Regression Trees*. s.l.:Wadsworth and Brooks/Cole.
- Brosnan, T. & Sun, D., 2002. Inspection and grading of agricultural and food products by computer vision systems – A review. *Comput Electron Agric*, Volumen 36, p. 193–213.
- Canny, J., 1986. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI*, 8(6), p. 679–698.
- Cano Marchal, P., Martínez Gila, D., Gámez García, J. & Gómez Ortega, J., 2013. Expert system based on computer vision to estimate the content of impurities in olive oil samples. *J Food Eng*, Volumen 220–228, p. 119.
- Cano Marchal, P. y otros, 2021b. Assessment of the Nutritional State for Olive Trees Using UAVs.. En: *In: Gonçalves, J.A., Braz-César, M., Coelho, J.P. (eds) CONTROLO 2020. Lecture Notes in Electrical Engineering*. s.l.:Springer, p. 695.

- Cano-Marchal, P., Satorres Martínez, S., Gómez Ortega, J. & Gámez García, J., 2021a. Automatic System for the Detection of Defects on Olive Fruits in an Oil Mill. *Applied Sciences*, Volumen 11, p. 8167.
- Cao, F. & Guo, W., 2020. Deep hybrid dilated residual networks for hyperspectral image classification. *Neurocomputing*, Volumen 384, p. 170–181. <https://doi.org/10.1016/J.NEUCOM.2019.11.092>.
- Carfagni, M., Daou, M., Furferi, R. & Marta, S., 2008. Real-time estimation of olive oil quality parameters: A combined approach based on ANNs and Machine Vision. *Int Conf Comput*, Volumen 12, p. 387–392.
- Chen, F. y otros, 2022. Rice False Smut Detection and Prescription Map Generation in a Complex Planting Environment, with Mixed Methods, Based on Near Earth Remote Sensing. *Remote Sensing*, Volumen 14, p. 945. <https://doi.org/10.3390/RS14040945>.
- Chen, L. y otros, 2015. Semantic image segmentation with deep convolutional nets and fully connected CFRS.. En: *Proc. Int. Conf. Learn. Representations*. s.l.:s.n.
- Chen, L. y otros, 2017. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs.. En: *TPAMI*. s.l.:s.n.
- Chen, L. y otros, 2018. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4), pp. 834-848..
- Chen, S., Chiu, M. & Zou, X., 2022. Real-time defect inspection of green coffee beans using NIR snapshot hyperspectral imaging. *Comput Electron Agric*, Volumen 197, p. 106970. <https://doi.org/10.1016/J.COMPAG.2022.106970>.
- Chen, S., Wang, Y. & C., C., 2012. Key Issues in Modeling of Complex 3D Structures from Video Sequences. *Mathematical Problems in Engineering*, Volumen 2012, pp. 1-17.
- Chen, X. & Jeong, J., 2007. Enhanced recursive feature elimination. En: *Sixth international conference on machine learning and applications (ICMLA 2007)*. s.l.:s.n., pp. 429-435..
- Chu, H. y otros, 2022. Hyperspectral imaging with shallow convolutional neural networks (SCNN) predicts the early herbicide stress in wheat cultivars. *J Hazard Mater*, Volumen 421, p. 126706. <https://doi.org/10.1016/J.JHAZMAT.2021.126706>.
- Congalton, R., 1991. A review of assessing the accuracy of classifications of remotely sensed data.. *Remote Sensing of Environment*, Volumen 37, pp. 35-46.
- Cui, R. y otros, 2022. Hyperspectral imaging coupled with Dual-channel convolutional neural network for early detection of apple valsa canker. *Comput Electron Agric*, Volumen 202, p. 107411. <https://doi.org/10.1016/J.COMPAG.2022.107411>.
- de Souza, R. y otros, 2020. Occurrence, impacts and general aspects of pesticides in surface water: A review. *Process Safety and Environmental Protection*, Volumen 135, pp. 22-37.

- Devijver, P. & Kittler, K., 1982. *Pattern Recognition: A Statistical Approach*. Londres: Prentice-Hall.
- Díaz, R. y otros, 2000. The application of a fast algorithm for the classification of olives by machine vision. *Food Res Int.*, Volumen 33, p. 305–309.
- Díaz, R. y otros, 2004. Comparison of three algorithms in the classification of table olives by means of computer vision. *J Food Eng*, Volumen 61, p. 101–107.
- Domínguez-Cid, S. y otros, 2023. In-field hyperspectral imaging dataset of Manzanilla and Gordal olive varieties throughout the season. *Data Brief*, Volumen 46, p. 108812. <https://doi.org/10.1016/J.DIB.2022.108812>.
- Dong, F. y otros, 2022. Identification of the proximate geographical origin of wolfberries by two-dimensional correlation spectroscopy combined with deep learning. *Comput. Electron. Agric.*, Volumen 198, p. 107027.
- Du, Z., Jeong, M. & Kong, S., 2007. Band selection of hyperspectral images for automatic detection of poultry skin tumors. *IEEE Trans. Autom. Sci. Eng.*, 4(3), pp. 332–339, <https://doi.org/10.1109/TASE.2006.888048>.
- Elahi, E., Weijun, C., Zhang, H. & Nazeer, M., 2019. Agricultural intensification and damages to human health in relation to agrochemicals: Application of artificial intelligence. *Land use policy*, Volumen 83, pp. 461-474.
- Elhashash, M., Albanwan, H. & Qin, R., 2022. A review of mobile mapping systems: From sensors to applications. *Sensors*, 22(11), p. 4262.
- El-Nahhal, Y. & El-Nahhal, I., 2021. Cardiotoxicity of some pesticides and their amelioration. *Environmental Science and Pollution Research*, Volumen 28, pp. 44726-44754.
- Escalera, A., 2001. *Visión por computador fundamentos y métodos*. s.l.:Prentice Hall, 274 p..
- Espinosa-Aranda, J., Fernández-Carroble, M. & Vález, N., 2016. Operaciones sobre el histograma y filtrado de la imagen. En: E. Alegre, G. Pajares & A. Escalera, edits. *Conceptos y métodos en Visión por Computador*. s.l.:Grupo de Visión del Comité Español de Automática, p. 31.
- Fazari, A. et al., 2018. Application of deep convolutional neural networks for the detection of anthracnose in olives using VIS/NIR hyperspectral images. *Comput Electron Agric* 187, 106252. <https://doi.org/10.1016/J.>
- Fei, S. y otros, 2022. Combining novel feature selection strategy and hyperspectral vegetation indices to predict crop yield. *Plant Methods*, Volumen 18, pp. 1–13. <https://doi.org/10.1186/S13007-022-00949-0/TABLES/2..>
- Feng, S. y otros, 2022. A deep convolutional neural network-based wavelength selection method for spectral characteristics of rice blast disease. *Comput Electron Agric*, Volumen 199, p. 107199. <https://doi.org/10.1016/J.COMPAG.2022.107199>.

- Fernández Ruiz, M., 2022. *Aplicación de técnicas de Inteligencia Artificial para identificar el porcentaje de suciedad en lotes de aceitunas en almazaras*. Jaén: TFG, Universidad de Jaén, 259 p..
- Ferrer, C. y otros, 2005. Determination of pesticide residues in olives and olive oil by matrix solid-phase dispersion followed by gas chromatography/mass spectrometry. *Journal of Chromatography*, 1069(2), pp. 183-194.
- Foudan, S. y otros, 2001. Hyperspectral image analysis for oil spill detection. En: *Summaries of NASA/JPL Airborne Earth Science Workshop*. Pasadena, CA: s.n., p. 5–9.
- Fraga, H., Pinto, J., Viola, F. & Santos, J., 2020. Climate change projections for olive yields in the Mediterranean Basin. *International Journal of Climatology*, 40(2), pp. 769-781.
- Furferi, R. & Carfagni, M., 2010. A machine vision system for real-time and automatic assessment of olives colour and surface defects. *Int J Comput Res*, Volumen 17, p. 303–320.
- Gao, B. y otros, 2023. Early Detection of *Dendroctonus valens* Infestation at Tree Level with a Hyperspectral UAV Image. *Remote Sens*, Volumen 15, p. 407. <https://doi.org/10.3390/RS15020407>.
- Gao, T. y otros, 2021. HyperSeed: An End-to-End Method to Process Hyperspectral Images of Seeds. *Sensors*, Volumen 21, pp. 8184–8184. <https://doi.org/10.3390/S21248184>.
- Gao, Z. y otros, 2020. Real-time hyperspectral imaging for the in-field estimation of strawberry ripeness with deep learning. *Artificial Intelligence in Agriculture*, Volumen 4, p. 31–38. <https://doi.org/10.1016/J.AIIA.2020.04.003>.
- García-Berná, J. y otros, 2020. Systematic Mapping Study on Remote Sensing in Agriculture. *Applied Sciences*, Volumen 10, p. 3456. <https://doi.org/10.3390/APP10103456>.
- García-Olalla, O. & Alegre, E., 2016. Descripción de textura en imágenes utilizando Local Binary Pattern (LPB). En: Alegre, G. Pajares & A. Escalera, edits. *Conceptos y métodos en Visión por Computador*. s.l.:Grupo de Visión del Comité Español de Automática, p. 115.
- García-Ordás, M., Alaiz-Rodríguez, R. & Alegre, E., 2016. Clasificación y reconocimiento de patrones. En: *Conceptos y métodos en Visión por Computador*. s.l.:s.n., pp. 159-179.
- García-Reyes, J. y otros, 2007. Determination of pesticide residues in olive oil and olives. *Trends in analytical chemistry*, 26(3), pp. 239-251.
- Gatica, G., Best, S., Ceroni, J. & Lefranc, G., 2013. Olive fruits recognition using neural networks. *Procedia Comput Sci.*, Volumen 17, p. 412–419.
- Goetz, A., Vane, G., Solomon, J. & Rock, B., 1985. Imaging spectrometry for earth remote sensing. *Science*, 228(4704), pp. 1147-1153.

Golhani, K., Balasundram, S., Vadamalai, G. & Pradhan, B., 2018. A review of neural networks in plant disease detection using hyperspectral data. *Information Processing in Agriculture*, Volumen 5, p. 354–371. <https://doi.org/10.1016/J.INPA.2018.05.002>.

Guerri, M. D. C., Spagnolo, P., Bougourzi, F. & Taleb-Ahmed, A., 2023. Deep Learning Techniques for Hyperspectral Image Analysis in Agriculture: A Review. *ArXiv*, p. abs/2304.13880.

Guijarro, M., Herrera, P. & Montalvo, M., 2016. Segmentación de regiones. En: E. Alegre, G. Pajares & A. Escalera, edits. *Conceptos y métodos en Visión por Computador*. s.l.:Grupo de Visión del Comité Español de Automática, p. 99.

Guyon, I. & Elisseeff, A., 2003. An introduction to feature extraction. En: *In Feature Extraction, Foundations and Applications*. s.l.:Springer.

Guzmán, E., Baeten, V., Pierna, J. & García-Mesa, J., 2013a. Infrared machine vision system for the automatic detection of olive fruit quality. *Talanta*, Volumen 116, p. 894–898.

Guzmán, E., Baeten, V., Pierna, J. & García-Mesa, J., 2013b. Using a visible vision system for on-line determination of quality parameters of olive fruits. *Food Nutr Sci*, Volumen 4, p. 90–98.

Guzmán, E., Baeten, V., Pierna, J. & García-Mesa, J., 2015. Determination of the olive maturity index of intact fruits using image analysis. *J Food Sci Technol*, Volumen 52, p. 1462–1470.

Han, Z. & Gao, J., 2019. Pixel-level aflatoxin detecting based on deep learning and hyperspectral imaging. *Comput Electron Agric*, Volumen 164, p. 104888. <https://doi.org/10.1016/J.COMPAG.2019.104888>.

Hastie, T., Tibshirani, R. & Friedman, J., 2009. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. s.l.:Springer.

Haut, J. y otros, 2019. Hyperspectral Image Classification Using Random Occlusion Data Augmentation. *IEEE Geoscience and Remote Sensing Letters*, Volumen 16, p. 1751–1755. <https://doi.org/10.1109/LGRS.2019.2909495>.

He, K., Zhang, X., Ren, S. & Sun, J., 2015. Deep residual learning for image recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, p. 770–778.

He, M., Li, B. & Chen, H., 2018. Multi-scale 3D deep convolutional neural network for hyperspectral image classification. En: *Proceedings - International Conference on Image Processing*. s.l.:s.n., p. <https://doi.org/10.1109/ICIP.2017.8297014>.

Herrera, P., Guijarro, M. & Guerrero, J., 2016. Operaciones de transformación de imágenes. En: E. Alegre, G. Pajares & A. Escalera, edits. *Conceptos y métodos en Visión por Computador*. s.l.:Grupo de Visión del Comité Español de Automática, p. 61.

He, W. y otros, 2021. Non-destructive detection and recognition of pesticide residues on garlic chive (*Allium tuberosum*) leaves based on short wave infrared hyperspectral imaging and one-dimensional convolutional neural network. *Journal of Food Measurement and*

Characterization, Volumen 15, pp. 4497–4507. <https://doi.org/10.1007/S11694-021-01012-7/REFERENCES>.

Hochreiter, S. & Schmidhuber, J., 1997. Long short-term memory. *Neural Computation*, 9(8), pp. 1735-1780.

Hong, S. y otros, 2023. Nondestructive prediction of pepper seed viability using single and fusion information of hyperspectral and X-ray images. *Sens Actuators A Phys*, Volumen 350, p. 114151. <https://doi.org/10.1016/J.SNA.2022.114151>.

Hu, M. y otros, 2014. Reduction of postharvest anthracnose and enhancement of disease resistance in ripening mango fruit by nitric oxide treatment. *Postharvest Biology and Technology*, Volumen 97, pp. 115-122.

Hu, Z. y otros, 2022. Research on Identification Technology of Field Pests with Protective Color Characteristics. *Applied Sciences*, Volumen 12, p. 3810.

Illana Rico, S., Martínez Gila, D., Cano Marchal, P. & Gómez Ortega, J., 2022. Automatic Detection of Olive Tree Canopies for Groves with Thick Plant Cover on the Ground. *Sensors*, 22(16), p. 6219. <https://doi.org/10.3390/s22166219>.

Ito, T., Minamino, K. & Umeki, S., 2023. Analysis of Fertilization Effects on Rice and Wheat by Time-Series Clustering of Vegetation Index Data. En: *The 5th International Symposium on Advanced Technologies and Applications in the Internet of Things (ATAIT)*. s.l.:s.n., pp. 55-65.

James, G., Witten, D., Hastie, T. & Tibshirani, R., 2017. *An introduction to statistical learning*. 7th ed. s.l.:Springer.

Jänicke, C. y otros, 2020. Brightness gradient-corrected hyperspectral image mosaics for fractional vegetation cover mapping in Northern California. *Remote Sens. Lett.*, 11(1), p. 1–10.

Jensen, J., 2006. *Remote Sensing of the Environment: An Earth Resource Perspective*. 2nd edition ed. s.l.:Pearson College Div.

Jensen, J., 2009. *Remote Sensing of the Environment: An Earth Resource Perspective*. 2 ed. Pearson Education: New Delhi, 613 pp..

Jensen, J. & Lulla, K., 1987. *Introductory digital image processing: a remote sensing perspective*. s.l.:s.n.

Jia, S. y otros, 2021. A survey: Deep learning for hyperspectral image classification with few labeled samples. *Neurocomputing*, Volumen 448, p. 179–204. <https://doi.org/10.1016/J.NEUCOM.2021.03.035>.

Jiménez, L., 2020. *Visión por Computador. Grado Ingeniería Electrónica y Automática Industrial (Material Docente)*. s.l.:s.n.

- Jin, X. y otros, 2018. Classifying Wheat Hyperspectral Pixels of Healthy Heads and Fusarium Head Blight Disease Using a Deep Neural Network in the Wild Field. *Remote Sensing*, Volumen 10, p. 395. <https://doi.org/10.3390/RS10030395>.
- Johnson, R. & Wichern, D., 2007. *Applied Multivariate Statistical Analysis*. s.l.:Prentice Hall.
- Jolliffe, I., 2002. *Principal component analysis*. Nueva York: Springer-Verlag: Springer Series in Statistics.
- Kanso, A. & Ghebleh, M., 2017. An efficient (t,n)–threshold secret image sharing scheme. *Multimed Tools Appl*, Volumen 76,, p. 16369–16388.
- Kapur, J., Sahoo, P. & Wong, A., 1985. A new method for gray-level picture thresholding using the entropy of the histogram. *Comput. Vis. Graph. Image Process*, 29(3), pp. 273-285..
- Karila, K. y otros, 2022 . Estimating Grass Sward Quality and Quantity Parameters Using Drone Remote Sensing with Deep Neural Networks. *Remote Sensing*, Volumen 14, p. 2692. <https://doi.org/10.3390/RS14112692>..
- Kass, M., Witkin, A. & Terzopoulos, D., 1988. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4), p. 321–331.
- Kaur, R., Kumar, M. & M., J., 2017. Fruit grading By Computer Vision-A Review. *International Journal of Recent Research Aspects*, 4(3), pp. 54-58.
- Khan, M. y otros, 2018. Modern Trends in Hyperspectral Image Analysis: A Review. *IEEE Access*, Volumen 6, p. 14118–14129.
- Krauss, K., 2007. *Photogrammetry: geometry from images and laser scans*. Walter De Gruyter.. s.l.:Walter De Gruyter.
- Krizhevsky, A., Sutskever, I. & Hinton, G., 2012. Imagenet classification with deep convolutional neural networks. *International Conference on Neural Information Processing Systems*, p. 1097–1105.
- Kuflik, P. & Rotman, S., 2012. Band selection for gas detection in hyperspectral images. En: *IEEE 27th Convention of Electrical and Electronics Engineers in Israel*. s.l.:s.n., p. 1–4. doi: 10.1109/EEEI.2012.6376973.
- Kumar, S. y otros, 2013. Change in the microenvironment of breast cancer studied by FTIR imaging. *Analyst*, 138(14), pp. 4058-4065.
- Kurtulmus, F., Lee, W. S. & Vardar, A., 2011. Green citrus detection using 'eigenfruit', color and circular Gabor texture features under natural outdoor conditions. *Computers and Electronics in Agriculture*, 78(2), p. 140–149.
- Kuswidiyanto, L. N. H. H. X., 2022. Plant Disease Diagnosis Using Deep Learning Based on Aerial Hyperspectral Images: A Review. *Remote Sensing*, Volumen 14, p. 6031. <https://doi.org/10.3390/RS14236031> . .

- Ladi, S. y otros, 2022. A Novel Grey Wolf Optimisation based CNN Classifier for Hyperspectral Image classification. *Multimed Tools Appl*, Volumen 81, pp. 28207–28230. <https://doi.org/10.1007/S11042-022-12628-2/TABLES/9>.
- Lantero, E., Matallanas, B. & Callejas, C., 2023. Current status of the main olive pests: Useful integrated pest management strategies and genetic tools. *Applied Sciences*, 13(21), p. 12078.
- Lecun, Y., Bottou, L., Bengio, Y. & Haffner, P., 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), p. 2278–2324.
- Lentza-Rizos, C. & Avramides, E. J., 1995. Pesticide residues in olive oil. *Reviews of Environmental Contamination and Toxicology: Continuation of Residue Reviews*, pp. 111-134.
- Li, H., Lee, W. & Wang, K., 2016. Immature green citrus fruit detection and counting based on fast normalized cross correlation (FNCC) using natural outdoor colour images. *Precision agriculture*, 17(678-697), p. 6.
- Lillesand, T., Kiefer, R. & Chipman, J., 2015. *Remote Sensing and Image Interpretation*. 7 ed. s.l.:Wiley.
- Lin, G. y otros, 2019. Fruit detection in natural environment using partial shape matching and probabilistic Hough transform. *Precision Agriculture*, Volumen 21, pp. 160-177.
- Lin, T. y otros, 2014. Microsoft COCO: Common Objects in Context.. En: D. P. T. S. B. T. T. (. Fleet, ed. *Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science*. s.l.:Springer, Cham. <https://doi.org/10.1007/97>, p. 8693.
- Liu, C. y otros, 2020a. Determination of starch content in single kernel using near-infrared hyperspectral images from two sides of corn seeds. *Infrared Phys Technol*, Volumen 110, p. 103462. <https://doi.org/10.1016/J.INFRARED.2020.103462>.
- Liu, L. y otros, 2015. Photoacoustic spectroscopy as a non-destructive tool for quantification of pesticide residue in apple cuticle. *International Journal of Thermophysics*, Volumen 36, pp. 868-872.
- Li, Y., Pang, Z., Li, W. & Jia, S., 2018. Tree species classification of airborne hyperspectral image in cloud shadow area. En: *International Symposium of Space Optical Instrument and Application*. s.l.:Springer, p. 389–398.
- Li, Y., Zhang, H. & Shen, Q., 2017. Spectral–Spatial Classification of Hyperspectral Imagery with 3D Convolutional Neural Network. *Remote Sensing*, Volumen 9 , p. 67. <https://doi.org/10.3390/RS9010067>.
- Lloyd, S., 1982. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, Volumen 28, p. 129–137.
- Lowe, 2015. *The Computer Vision Industry*. [En línea] Available at: <https://www.cs.ubc.ca/~lowe/vision.html> (último acceso el 01/07/2022). [Último acceso: 10 10 2024].

- Lowe, A., Harrison, N. & French, A., 2017. Hyperspectral image analysis techniques for the detection and classification of the early onset of plant disease and stress. *Plant Methods*, Volumen 13, pp. 1–12. <https://doi.org/10.1186/S13007-017-0233-Z/TABLES/2>.
- Lowe, D., 2004. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, Volumen 60, p. 91–110.
- Lu, B. y otros, 2020. Recent Advances of Hyperspectral Imaging Technology and Applications in Agriculture. *Remote Sensing*, Volumen 12, p. 2659. <https://doi.org/10.3390/RS12162659>.
- Luchetti, F., 2002. Importance and future of olive oil in the world market—An introduction to olive oil. *European Journal of Lipid Science and Technology*, 104(9-10), pp. 559-563.
- Lu, G. & Fei, B., 2014. Medical hyperspectral imaging: a review.. *Journal of biomedical optics*, 19(1), pp. 010901-010901.
- Lv, W. & Wang, X., 2020. Overview of Hyperspectral Image Classification. *J. Sensors*, Volumen 4817234, pp. 1-13.
- Mahlein, A., 2016. Plant disease detection by imaging sensors— parallels and specific demands for precision agriculture and plant phenotyping. *Plant Dis*, Volumen 100, p. 241–251.
- Malik, M. y otros, 2018. Mature tomato fruit detection algorithm based on improved HSV and watershed algorithm. *IFAC-PapersOnLine*, Volumen 51, pp. 431-436.
- MAPA, 2024. *Ministerio de Agricultura, Pesca y Alimentación, Agricultura, Producciones agrícolas, Aceite y aceituna de mesa*. [En línea]
Available at: <https://www.mapa.gob.es/es/agricultura/temas/producciones-agricolas/aceite-oliva-y-aceituna-mesa/aceite.aspx>
[Último acceso: 15 11 2024].
- Marr, D., 1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: W. H. Freeman.
- Marr, D. & Hildreth, E., 1980. Theory of edge detection. *Proceedings of the Royal Society of London*, Volumen B 207, p. 187–217.
- Martínez Gila, D., Cano Marchal, P., Gaméz García, J. & Gómez Ortega, J., 2013. Control System of the Malaxing State for the Olive Paste Based on Computer Vision and Fuzzy Logic. *IEEE International Conference on Systems, Man, and Cybernetics, Manchester*, p. 1.
- Martínez Gila, D., Cano Marchal, P., Gámez García, J. & Gómez Ortega, J., 2015b. On-line system based on hyperspectral information to estimate acidity, moisture and peroxides in olive oil samples. *Comput Electron Agric*, Volumen 116, p. 1–7.
- Martínez Gila, D., Cano Marchal, P., Gómez Ortega, J. & Gámez García, J., 2018a. Expert System for Monitoring the Malaxing State of the Olive Paste Based on Computer Vision. *Sensors*, 18(7), p. 2227.

Martínez Gila, D., Cano Marchal, P., Gómez Ortega, J. & Gámez García, J., 2018b. Non-Invasive Methodology to Estimate Polyphenol Content in Extra Virgin Olive Oil Based on Stepwise Multilinear Regression. *Sensors*, Volumen 18,, p. 975.

<https://doi.org/10.3390/s18040975>.

Martínez Gila, D., Estévez, E., Gómez-Ortega, J. & Gamez García, J., 2023. Application of a lab-made voltammetric electronic tongue to identify musty and vinegary defects in olive oils. *Food Measure*, Volumen 17, pp. 1402–1412. <https://doi.org/10.1007/s11694-022-01694-7>.

Martínez Gila, D., Navarro Soto, J. & Satorres Martínez, S. a., 2022. The Advantage of Multispectral Images in Fruit Quality Control for Extra Virgin Olive Oil Production. *Food Anal. Methods*, Volumen 15, pp. 75–84. <https://doi.org/10.1007/s12161-021-02099-w>.

Martínez Gila, D., Puerto, D., Gamez García, J. & Gomez Ortega, J., 2015a. Automatic classification of olives for oil production using computer vision. *IEEE International Conference on Industrial Technology (ICIT), Seville, Spain.*, pp. 1651-1656..

Martínez Gila, D. y otros, 2021.). Classification of olive fruits and oils based on their fatty acid ethyl esters content using electronic nose technology. *Journal of Food Measurement and Characterization*, 15(6), pp. 5427-5438.

Martínez-Gila, D. y otros, 2024. Non-invasive detection of pesticide residues in freshly harvested olives using hyperspectral imaging technology. *Smart Agricultural Technology*, Volumen 9, p. 100644.

Masci, J., Meier, U., Cireşan, D. & Schmidhuber, J., 2011. Stacked Convolutional Auto-Encoders for Hierarchical Feature Extraction. En: T. D. W. G. M. K. S. Honkela, ed. *Artificial Neural Networks and Machine Learning – ICANN 2011. ICANN 2011. Lecture Notes in Computer Science*. Berlin, Heidelberg.: Springer, p. 6791.

Meng, S. y otros, 2021. Deep learning-based crop mapping in the cloudy season using one-shot hyperspectral satellite imagery. *Comput Electron Agric*, Volumen 186, p. 106188. <https://doi.org/10.1016/J.COMPAG.2021.106188>.

Min, D. y otros, 2024. Application of methyl jasmonate to control disease of postharvest fruit and vegetables: A Meta-analysis. *Postharvest Biology and Technology*, Volumen 208, p. 112667.

Moghimi, A., Yang, C. & Anderson, J., 2020. Aerial hyperspectral imagery and deep neural networks for high-throughput yield phenotyping in wheat. *Comput Electron Agric*, Volumen 172, p. 105299. <https://doi.org/10.1016/J.COMPAG.2020.105299>..

Moharram, M. & Sundaram, D., 2023. Land use and land cover classification with hyperspectral data: A comprehensive review of methods, challenges and future directions. *Neurocomputing*, Volumen 536, p. 90–113. <https://doi.org/10.1016/J.NEUCOM.2023.03.025>..

Mortensen, E. N. & Barrett, W. A., 1998. Interactive segmentation with intelligent scissors.. *Graphical Models and Image Processing*, 60(5), p. 349–384.

- Mou, L., Ghamisi, P. & Zhu, X., 2017. Deep recurrent neural networks for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, Volumen 55, p. 3639–3655. <https://doi.org/10.1109/TGRS.2016.2636241>.
- Narendra, V. & Hareesha, K., 2010. Quality Inspection and Grading of Agricultural and Food Products by Computer Vision- A Review. *International Journal of Computer Applications*, 2(1), pp. 43-65.
- Navarro Soto, J., Rico, S., Martínez Gila, D. & Satorres Martínez, S., 2024. Influence of the Degree of Fruitiness on the Quality Assessment of Virgin Olive Oils Using Electronic Nose Technology. *Sensors*, 24(8), p. 2565. <https://doi.org/10.3390/s24082565>.
- Navarro Soto, J. y otros, 2018. Fast and Reliable Determination of Virgin Olive Oil Quality by Fruit Inspection Using Computer Vision. *Sensors*, 18(11), p. 3826. <https://doi.org/10.3390/>
- Neupane, K. & Baysal-Gurel, F., 2021. Automatic Identification and Monitoring of Plant Diseases Using Unmanned Aerial Vehicles: A Review. *Remote Sensing*, Volumen 13, p. 3841. <https://doi.org/10.3390/RS13193841>.
- Nguyen, C., Sagan, V., Bhadra, S. & Moose, S., 2023. UAV Multisensory Data Fusion and Multi-Task Deep Learning for High-Throughput Maize Phenotyping. *Sensors*, Volumen 23, p. 1827. <https://doi.org/10.3390/S23041827>.
- Nguyen, C. y otros, 2021. Early Detection of Plant Viral Disease Using Hyperspectral Imaging and Deep Learning. *Sensors*, Volumen 21, p. 1–23. <https://doi.org/10.3390/S21030742>.
- Ojeda-Magaña, B., Ruelas, R., Quintanilla-Domínguez, J. & Andina, D., 2010. Color image segmentation by partitional clustering algorithms. *IECON 2010 - 36th Annual Conference on IEEE Industrial Electronics Society*, pp. 2828-2833.
- Onmankhong, J. y otros, 2022. Cognitive spectroscopy for the classification of rice varieties: A comparison of machine learning and deep learning approaches in analysing long-wave near-infrared hyperspectral images of brown and milled samples. *Infrared Phys Technol*, Volumen 123, p. 104100. <https://doi.org/10.1016/J.INFRARED.2022.104100>.
- Oscó, L. y otros, 2022. An impact analysis of pre-processing techniques in spectroscopy data to classify insect-damaged in soybean plants with machine and deep learning methods. *Infrared Phys Technol*, Volumen 123, p. 104203.
- Otsu, N., 1979. A threshold selection method from gray-level histograms. *IEEE Trans. Sys., Man., Cyber.*, Volumen 9, pp. 62-66.
- Pang, L. y otros, 2022. Rapid seed viability prediction of *Sophora japonica* by improved successive projection algorithm and hyperspectral imaging. *Infrared Phys Technol*, Volumen 123, p. 104143. <https://doi.org/10.1016/J.INFRARED.2022.104143>.
- Paoletti, M., Haut, J., Plaza, J. & Plaza, A., 2019. Deep learning classifiers for hyperspectral imaging: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, Volumen 158, pp. 279-317.

- Patel, M. y otros, 2023. Retrieving canopy nitrogen concentration and aboveground biomass with deep learning for ryegrass and barley: Comparing models and determining waveband contribution. *Field Crops Res*, Volumen 294, p. 108859. <https://doi.org/10.1016/J.FCR.2023.108859>.
- Patil, N. M. V. & Yadahalli, R., 2011. Color and Texture Based Identification and Classification of food Grains using different Color Models and Haralick features. *International Journal on Computer Science and Engineering*, 3(12).
- Paty, D. y otros, 1988. MRI in the diagnosis of MS: a prospective study with comparison of clinical evaluation, evoked potentials, oligoclonal banding, and CT. *Neurology*, 38(2), p. 180.
- Paul, S. y otros, 2020. Canopy Averaged Chlorophyll Content Prediction of Pear Trees Using Convolutional Autoencoder on Hyperspectral Data. *IEEE J Sel Top Appl Earth Obs Remote Sens*, Volumen 13, p. 1426–1437. <https://doi.org/10.1109/JSTARS.2020.2983000>.
- Peñuelas, J. & Filella, I., 1998. Visible and near-infrared reflectance techniques for diagnosing plant physiological status. *Trends in Plant Science*, 3(4), pp. 151-156.
- Platero, C., 2008. *Apuntes de visión artificial*. s.l.:Universidad Politécnica de Madrid.
- Polder, G. y otros, 2019. Potato virus Y detection in seed potatoes using deep learning on hyperspectral images. *Front Plant Sci*, Volumen 10, p. 434052. <https://doi.org/10.3389/FPLS.2019.00209/BIBTEX>.
- Pullanagari, R., Dehghan-Shoar, M., Yule, I. & Bhatia, N., 2021. Field spectroscopy of canopy nitrogen concentration in temperate grasslands using a convolutional neural network. *Remote Sens Environ*, Volumen 257, p. 112353. <https://doi.org/10.1016/J.RSE.2021.112353>.
- Qi, C. y otros, 2023. In-field classification of the asymptomatic biotrophic phase of potato late blight based on deep learning and proximal hyperspectral imaging. *Comput Electron Agric*, Volumen 205, p. 107585. <https://doi.org/10.1016/J.COMPAG.2022.107585>.
- Ramos, P., Prieto, F., Montoya, E. & Oliveros, C., 2017. Automatic fruit count on coffee branches using computer vision. *Comput. Electron. Agric.*, Volumen 137, pp. 9-22.
- Ram, T., Wiesman, Z., Parmet, I. & Edan, Y., 2010. Olive oil content prediction models based on image processing. *Biosyst Eng*, Volumen 105, p. 221–232.
- Rehman, T. y otros, 2020. Predictive spectral analysis using an end-to-end deep model from hyperspectral images for high-throughput plant phenotyping. *Comput Electron Agric*, Volumen 177, p. 105713. <https://doi.org/10.1016/J.COMPAG.2020.105713>.
- Ren, Z., Rao, Z. & Ji, H., 2018. Identification of Different Concentrations Pesticide Residues of Dimethoate on Spinach Leaves by Hyperspectral Image Technology. *IFAC-PapersOnLine*, Volumen 51, pp. 758-763.
- Ripley, B., 1996. *Pattern Recognition and Neural Networks*. s.l.:Cambridge University Press.

- Rouse, J. y otros, 1974. *Monitoring the vernal advancement and retrogradation (Greenwave effect) of natural vegetation*, Greenbelt, MD, USA: NASA/GSFCT Type III Final Report.
- Russell, S. & Norvig, P., 2010. *Artificial Intelligence. A Modern Approach. Third Edition*. s.l.:Prentice Hall.
- Sabine, C. y otros, 2014. Potential of hyperspectral imagery for the spatial assessment of soil erosion stages in agricultural semi-arid Spain at different scales. En: *IEEE Geoscience and Remote Sensing Symposium, IEEE*. s.l.:s.n., p. 2918–2921..
- Sánchez-Salmerón, A. & Ricolfe-Viala, C., 2016. Sistema de captura de imágenes. En: E. Alegre, G. Pajares & A. Escalera, eds. *Conceptos y métodos en Visión por Computador*. . s.l.:Grupo de Visión del Comité Español de Automática, p. 11.
- Sawyer, E. y otros, 2023. Phenotyping grapevine red blotch virus and grapevine leafroll-associated viruses before and after symptom expression through machine-learning analysis of hyperspectral images. *Front Plant Sci*, Volumen 14, p. 1117869. <https://doi.org/10.3389/FPLS.2023.1117869/BIBTEX>.
- Selci, S., 2019. The future of hyperspectral imaging. *Journal of Imaging*, 5 (11), p. 84.
- Sheela, K. & Deepa, S., 2013. Review on methods to fix number of hidden neurons in neural networks. *Math. Probl. Eng.*, p. 1–11. <https://doi.org/10.1155/2013/425740>.
- Shi, Y. y otros, 2022. *Remote Sensing*, Volumen 14, p. 396. <https://doi.org/10.3390/RS14020396>.
- Shuai, L. y otros, 2024. A research review on deep learning combined with hyperspectral Imaging in multiscale agricultural sensing. *Comput. Electron. Agric.*, Volumen 217, p. 108577.
- Simonyan, K. & Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. *In International Conference on Learning Representations (ICLR)*.
- Sindhu, S. & Manickavasagan, A., 2023. Nondestructive testing methods for pesticide residue in food commodities: A review. *Comprehensive Reviews in Food Science and Food Safety*, 22(2), pp. 1226-1256..
- Sirmaçek, B., Botteghi, N. & Plaza, S., 2019. Sequential image processing methods for improving semantic video segmentation algorithms. *ArXiv*, Volumen abs/1910.13348..
- Strachan, I., Pattey, E. & Boisvert, J., 2002. Impact of nitrogen and environmental conditions on corn as detected by hyperspectral reflectance. *Remote Sens. Environ*, 80(2), p. 213–224.
- Su, Z. y otros, 2021. Application of Hyperspectral Imaging for Maturity and Soluble Solids Content Determination of Strawberry With Deep Learning Approaches. *Front Plant Sci*, Volumen 12, p. 736334. <https://doi.org/10.3389/FPLS.2021.736334/BIBTEX>.
- Szegedy, C. y otros, 2015. Going deeper with convolutions.. En: *Computer Vision: Algorithms and Applications. 28th IEEE conference on computer vision and pattern recognition (CVPR)*. s.l.:Springer, p. 957 pp..

- Szeliski, R., 2010. *Computer Vision: Algorithms and Applications*. s.l.:Springer, 957 pp.
- Tan, J., 2004. Meat quality evaluation by computer vision. *Journal of food engineering*, 61(1), pp. 27-35.
- Teke, M. y otros, 2013. A short survey of hyperspectral remote sensing applications in agriculture. En: *6th International Conference on Recent Advances in Space Technologies (RAST), IEEE*. s.l.:s.n., p. 171–176.
- Terwilliger, B. & Robbins, J., 2020. *Small Unmanned Aircraft Systems Guide: Exploring Applications, Operations, and Integration*. New York: McGraw-Hill Education.
- Therios, I., 2009. Olives. *CABI*, Volumen 18.
- Thomas, S. y otros, 2018. Benefits of hyperspectral imaging for plant disease detection and plant protection: a technical perspective. *Journal of Plant Diseases and Protection*, Volumen 125, pp. 5–20. <https://doi.org/10.1007/S41348-017-0124-6/FIGURES/4>.
- Ting, K., 2011. Confusion Matrix. En: C. Sammut & G. Webb, edits. *Encyclopedia of Machine Learning*. Boston, MA: Springer, pp. https://doi.org/10.1007/978-0-387-30164-8_157.
- Torres, J., 2020. *Python Deep Learning. Introducción práctica con Keras y TensorFlow 2*. s.l.:Marcombo.
- Torres-Tello, J. & Ko, S., 2021. A novel approach to identify the spectral bands that predict moisture content in canola and wheat. *Biosyst Eng*, Volumen 210, p. 91–103. <https://doi.org/10.1016/J.BIOSYSTEMSENG.2021.08.004>.
- Tudi, M. y otros, 2021. Agriculture development, pesticide application and its impact on the environment. *International journal of environmental research and public health*, 3(1), p. 18.
- van der Meer, F. y otros, 2012. Multi- and hyperspectral geologic remote sensing: A review. *Int. J. Appl. Earth Obs. Geoinformation*, Volumen 14, pp. 112-128.
- Vanloot, P. y otros, 2014. Artificial vision and chemometrics analyses of olive stones for varietal identification of five French cultivars. *Comput Electron Agric.*, Volumen 102, p. 98–105.
- Vapnik, V., 1995. *The Nature of Statistical Learning Theory*. 2ª ed. Nueva York, USA: Springer Verlag.
- Velasco, J. G., 2009. *Energías renovables*. s.l.:Editorial Reverte.
- Wambugu, N. y otros, 2021. A hybrid deep convolutional neural network for accurate land cover classification. *Int. J. Appl. Earth Obs. Geoinformation*, Volumen 103, p. <https://doi.org/10.1016/J.JAG.2021.102515>.
- Wang, C. y otros, 2021. A review of deep learning used in the hyperspectral image analysis for agriculture. *Artificial Intelligence Review*, 54 (7), pp. 5205–5253. <https://doi.org/10.1007/S10462-021-10018-Y>.

- Wang, D. y otros, 2022b. A Review of Deep Learning in Multiscale Agricultural Sensing. *Remote Sensing*, Volumen 14, p. 559. <https://doi.org/10.3390/RS14030559>..
- Wang, D. y otros, 2022c. BO-DRNet: An Improved Deep Learning Model for Oil Spill Detection by Polarimetric Features from SAR Images. *Remote Sens.*, Volumen 14, p. 264.
- Wang, J., Bretz, M., Dewan, M. & Delavar, M., 2022a. *Sci Total Environ*, Volumen 822, p. 153559. doi: 10.1016/j.scitotenv.2022.153559. Epub 2022 Jan 31. PMID: 35114222..
- Wang, J., Wang, S., Liu, N. & Shang, F., 2021. A detection method of two carbamate pesticides residues on tomatoes utilizing excitation-emission matrix fluorescence technique. *Microchemical Journal*, Volumen 164, p. 105920.
- Wang, M. y otros, 2019. A feature selection approach for hyperspectral image based on modified ant lion optimizer. *Knowl. Based Syst.*, Volumen 168, pp. 39-48.
- Wan, L. y otros, 2022. Hyperspectral Sensing of Plant Diseases: Principle and Methods. *Agronomy*, Volumen 12, p. 1451. <https://doi.org/10.3390/AGRONOMY12061451>..
- Wan, S., Yeh, M. & Ma, H., 2021. An Innovative Intelligent System with Integrated CNN and SVM: Considering Various Crops through Hyperspectral Image Data. *ISPRS International Journal of Geo-Information*, Volumen 10, p. 242. <https://doi.org/10.3390/IJGI10040242>.
- Weiss, M., Jacob, F. & Duveiller, G., 2020. Remote sensing for agricultural applications: A meta-review. *Remote Sensing of Environment*, 236(111402).
- Wei, Y. y otros, 2023. Classification for GM and Non-GM Maize Kernels Based on NIR Spectra and Deep Learning. *Processes*, Volumen 11, p. 486. <https://doi.org/10.3390/PR11020486>.
- Wieme, J. y otros, 2022. Application of hyperspectral imaging systems and artificial intelligence for quality assessment of fruit, vegetables and mushrooms: A review. *Biosystems engineering*, Volumen 222, pp. 156-176.
- Wu, N. y otros, 2022. Deep convolution neural network with weighted loss to detect rice seeds vigor based on hyperspectral imaging under the sample-imbalanced condition. *Comput Electron Agric*, Volumen 106850. <https://doi.org/10.1016/J.COMPAG.2022.106850>, p. 196.
- Xu, H., 2006. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. . *International Journal of Remote Sensing*, 27(14), p. 3025–3033.
- Xu, Y., Du, B., Zhang, F. & Zhang, L., 2018. Hyperspectral image classification via a random patches network. *ISPRS Journal of Photogrammetry and Remote Sensing*, Volumen 142, p. 344–357. <https://doi.org/10.1016/J.ISPRSJPRS.2018.05.014>.
- Yang, Q., 1996. Apple stem and calyx identification with machine vision. *Journal of Agricultural Engineering Research*, Volumen 63, pp. 229-236.

Yang, W. y otros, 2021. Estimation of corn yield based on hyperspectral imagery and convolutional neural network. *Comput Electron Agric*, Volumen 184, p. 106092. <https://doi.org/10.1016/J.COMPAG.2021.106092>.

Yang, W. y otros, 2021. Estimation of corn yield based on hyperspectral imagery and convolutional neural network. *Comput Electron Agric*, Volumen 184, p. 106092. <https://doi.org/10.1016/J.COMPAG.2021.106092>.

Ye, W. y otros, 2022. Detection of Pesticide Residue Level in Grape Using Hyperspectral Imaging with Machine Learning. *Foods*, Volumen 11, p. 1609. <https://doi.org/10.3390/FOODS11111609>.

Yoosefzadeh-Najafabadi, M., Tulpan, D. & Eskandari, M., 2021. Using hybrid artificial intelligence and evolutionary optimization algorithms for estimating soybean yield and fresh biomass using hyperspectral vegetation indices. *Remote Sens*, Volumen 13, p. 2555. <https://doi.org/10.3390/RS13132555/S1>.

Zagolski, F. y otros, 1996. Forest canopy chemistry with high spectral resolution remote sensing. *International Journal of Remote Sensing*, 17(6), pp. 1107-1128.

Zarco-Tejada, P., Hubbard, N. & Loudjani, P., 2014. *Precision Agriculture: An opportunity for EU farmers—potential support with the CAP 2014–2020*, Belgium: EPRS: European Parliamentary Research Service.

Zhang, C. & Wang, H., 2019. Application of High-Resolution Remote Sensing in Precision Agriculture. *Remote Sensing*, 11(6), p. 650.

Zhang, C. y otros, 2020c. Application of near-infrared hyperspectral imaging for variety identification of coated maize kernels with deep learning. *Infrared Phys Technol*, Volumen 111, p. 103550. <https://doi.org/10.1016/J.INFRARED.2020.103550>.

Zhang, C. y otros, 2022. End-to-End Fusion of Hyperspectral and Chlorophyll Fluorescence Imaging to Identify Rice Stresses. *Plant Phenomics*, Volumen <https://doi.org/10.34133/2022/9851096>.

Zhang, D. y otros, 2020b. Integrating spectral and image data to detect Fusarium head blight of wheat. *Comput Electron Agric*, Volumen 175, p. 105588. <https://doi.org/10.1016/J.COMPAG.2020.105588>.

Zhang, W., Jiang, F. & Ou, J., 2011. Global pesticide consumption and pollution: with China as a focus. *Proceedings of the international academy of ecology and environmental sciences*, 1(2), p. 125.

Zhang, X. y otros, 2022b. Rapid and non-invasive surface-enhanced Raman spectroscopy (SERS) detection of chlorpyrifos in fruits using disposable paper-based substrates charged with gold nanoparticle/halloysite nanotube composites. *Microchimica Acta*, 189(5), p. 197.

Zhang, X., Chen, S., Jiang, H. & Shuai, Y., 2020a. Monitoring Crop Growth Using Satellite Remote Sensing: A Review. *Remote Sensing*, 12(1), p. 24.

- Zhang, X. y otros, 2019. A Deep Learning-Based Approach for Automated Yellow Rust Disease Detection from High-Resolution Hyperspectral UAV Images. *Remote Sensing*, Volumen 11, p. 1554. <https://doi.org/10.3390/RS11131554>.
- Zhang, Y. y otros, 2021. *Remote Sens Environ*, Volumen 26, p. 112724. <https://doi.org/10.1016/J.RSE.2021.112724>.
- Zhao, R. y otros, 2023. Improving chlorophyll content detection to suit maize dynamic growth effects by deep features of hyperspectral data. *Field Crops Res*, Volumen 297, p. 108929. <https://doi.org/10.1016/J.FCR.2023.108929>.
- Zhong, Z., Li, J., Luo, Z. & Chapman, M., 2018. Spectral-Spatial Residual Network for Hyperspectral Image Classification: A 3-D Deep Learning Framework. *IEEE Transactions on Geoscience and Remote Sensing*, Volumen 56, p. 847–858. <https://doi.org/10.1109/TGRS.2017.2755542>.
- Zhong, Z. y otros, 2017. Deep residual networks for hyperspectral image classification. En: *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. s.l.:s.n., p. <https://doi.org/10.1109/IGARSS.2017.8127330>.
- Zhu, J., Fang, L. & Ghamisi, P., 2018. Deformable convolutional neural networks for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, Volumen 15, p. 1254–1258. <https://doi.org/10.1109/LGRS.2018.2830403>.
- Zhu, J., Li, H., Rao, Z. & Ji, H., 2023. Identification of slightly sprouted wheat kernels using hyperspectral imaging technology and different deep convolutional neural networks. *Food Control*, Volumen 143, p. 109291. <https://doi.org/10.1016/J.FOODCONT.2022.109291>.
- Zhu, S. y otros, 2019. Identification of Soybean Varieties Using Hyperspectral Imaging Coupled with Convolutional Neural Network. *Sensors*, Volumen 19, p. 4065. <https://doi.org/10.3390/S19194065>.