DOCTORAL THESIS

PhD in Food Science, Technology and Management

Advanced non-destructive evaluation of fruit quality and safety using spectral information and artificial neural networks

Salvador Castillo Gironés

Supervisors:

Dr. Juan Gómez Sanchis

Dr. Sandra Munera Picazo

Dr. Sergio Cubero García

Valencia, September 2025

S · S

EX LIBRIS

Those who journey far and seek to learn gather a wealth of sights and wisdom, for the world reveals its secrets to the curious spirit. Yet, it is kindness that gives true meaning to all discoveries.

To all who have lifted me with kindness, faith, and support, your belief has been the light guiding my path. My gratitude is as immense as your generosity.

INDEX

ABSTRACT 11

RESUMEN 12

RESUM 13

PREFACE 15

Research framework 16

Dissemination of the results 16

INTRODUCTION 22

1. Spectral Imaging and Raman Spectroscopy 24

1.1 Hyperspectral Imaging 24

1.1.1 Principles and Technologies 25

1.1.2 Types of Hyperspectral Imaging 26

1.1.2.1. Point scanning. 27

1.1.2.2. Area scanning. 27

1.1.2.3. Line scanning. 27

1.1.2.4. Snapshot. 27

1.1.3 Calibration: Ensuring Data Accuracy 29

1.2 Raman Spectroscopy 30

1.2.1 Principles and Comparison with Infrared Spectroscopy 31

1.2.2 Laser Selection in Raman Spectroscopy 33

1.2.3 Types of Raman Spectroscopy 34

1.1.3.1. Traditional Raman. 34

1.1.3.2. Raman Microscopy. 35

1.1.3.3. Raman Imaging. 35

1.1.3.4. Surface-Enhanced Raman. 35

1.2.4 Calibration: Ensuring Data Accuracy 36

2. ANN-Based Predictive Modeling for Fruit Quality and Safety 38

Abstract 40

2.2. Type of data 44

2.3. Data preprocessing 45

2.4. Artificial Neural Networks 49

2.4.1. Multi-Layer Perceptron (MLP) 51

2.4.2. Deep Belief Networks (DBNs) 51

2.4.3. Recurrent Neural Networks (RNNs) 52

2.4.4. Radial Basis Function Network (RBFN) 53

2.4.5. Autoencoders (AEs) 54
2.4.6. Convolutional Neural Networks (CNNs) 56
2.4.7. Generative models 58
2.4.7.1. Generative Adversarial Networks (GANS) 58
2.4.7.2. Diffusion Models 59
2.4.8. Miscellaneous models 59
2.4.8.1. Transformers 59
2.4.8.2. Capsule Neural Networks (CNs) 60
2.4.8.3. Self-Organising Maps (SOMs) 62
2.5. Applications of ANNs in quality and safety 62
2.6. Conclusions and future trends 67
Acknowledgements: 68
References 69
OVERVIEW OF THE RESEARCH 86
AIMS 90
General aims 91
Specific aims: 91
CHAPTER I. Advanced evaluation of strawberry quality, consumer preference, and cultivar discrimination through spectral imaging and neural networks. 94
Abstract 96
1. Introduction 97
2. Material and methods 98
2.1. Fruit samples and experimental design 98
2.2. Spectral image acquisition 99
2.3. Destructive quality measurements 100
2.4. Affective sensory evaluation 100
2.5. Image processing and data extraction 100
2.6. Data analysis 101
2.6.1. Quality parameters and consumer acceptance 101
2.6.2. Model creation and optimisation 102
3. Results and discussion 105
3.1. Quality parameters and affective sensory evaluation 105
3.2. Spectral analysis 108
3.3. Prediction of quality parameters 110
3.4. Prediction of consumer acceptance 111
3.5. Cultivar discrimination 112

4. Conclusions 115
Acknowledgements 115
References 116
CHAPTER II. Firmness classification in 'Rojo Brillante' persimmon fruit after cold storage using hyperspectral imaging and deep learning with variational autoencoders. 121
Abstract 123
1. Introduction 124
2. Material and methods 126
2.1 Fruit samples 126
2.2 Hyperspectral imaging acquisition and processing 127
2.3 Firmness analysis and curves processing 129
2.4 Firmness clustering 132
2.5 Firmness classification using multispectral images 132
3. Results and discussion 134
3.1 Firmness Clustering 134
3.2 Spectral dimensionality reduction 137
3.3 Latent variable representation 139
3.4 Firmness classification 139
4. Conclusions 141
Acknowledgements 141
References 142
CHAPTER III. Detection of subsurface bruises in plums using spectral imaging and deep learning with wavelength selection. 147
Abstract 149
1. Introduction 150
2. Materials and Methods 151
2.1 Plant material and bruising 151
2.2 Hyperspectral image acquisition and processing 152
2.3. Data analysis 154
2.3.1 Wavelength selection 154
2.3.2 Deep Learning Models 154
2.4 GradCAM heatmap visualization 156
3. Results 156
3.1. Important wavelengths 156
3.2. Bruise detection models 158
3.3. GradCam analysis 161

4. Discussion ... 164
5. Conclusion ... 165
Acknowledgements ... 165
References ... 166
CHAPTER IV. Raman spectroscopy for multi-label identification of common apple pesticide mixtures using CNNs and Gradient-weighted Class Activation Mapping. ... 172
Abstract ... 174
1. Introduction ... 175
2. Materials and methods ... 177
2.1. Pesticides and sample preparation ... 177
2.2. Raman spectra acquisition ... 181
2.3. Data processing and analysis ... 182
2.4. Multilabel classification model creation ... 183
3. Results ... 185
3.1. Pesticide spectra characterisation ... 185
3.2. Spectral analysis ... 188
3.3. Multilabel classification of pesticides ... 192
4. Conclusions ... 197
Acknowledgements ... 198
References ... 199
GENERAL DISCUSSION ... 204
CONCLUSIONS ... 209

ABSTRACT

Ensuring high internal quality and safety in fruit products is a major challenge for the agri-food industry, as traditional assessment methods are often destructive and inefficient. This doctoral thesis aimed to explore the use of two non-destructive optical techniques: hyperspectral imaging (HSI) and Raman spectroscopy, combined with advanced neural network models, to predict internal quality attributes, consumer acceptance, and food safety in strawberries, persimmons, plums, and apples.

The research involved four main studies. First, HSI and neural networks were applied to 17 strawberry cultivars to predict key quality parameters such as sugar content, acidity, and texture, as well as overall consumer acceptance and cultivar identification to prevent food fraud. Second, autoencoder-based deep learning models were used to classify 'Rojo Brillante' persimmon HSI into texture groups relevant for packing house sorting. Third, three convolutional neural networks (CNNs) were trained and compared to detect invisible mechanical bruises in 'Presenta' plums, aiming to reduce postharvest losses. Fourth, Raman spectroscopy combined with CNNs enabled the detection and identification of multiple pesticide active compounds in complex pesticide mixtures commonly used in apple orchards.

The results demonstrated that HSI and neural networks provided accurate predictions for strawberry quality (TSS R^2 = 0.85, TA R^2 = 0.81), mean cultivar consumer acceptance (R^2 = 0.78), and cultivar discrimination (F1 = 84 %). In persimmons, autoencoders achieved a 99 % reduction in data dimensionality and 81.3 % accuracy in texture classification. CNNs detected early-stage bruises in plums with an F1 score of 90 % and key wavelengths were identified. Raman spectroscopy with deep learning models identified the presence of specific pesticides in complex mixtures with an F1 of up to 100 % in some cases.

These findings highlight the practical value of combining advanced optical sensing and neural networks for rapid, non-destructive assessment of fruit quality and safety. This approach offers significant benefits for consumers, producers, and food safety authorities by improving product quality, reducing waste, and ensuring compliance with safety standards.

RESUMEN

Garantizar una buena calidad interna e inocuidad en los productos frutícolas es un gran desafío para la industria agroalimentaria, ya que los métodos tradicionales de evaluación suelen ser destructivos e ineficientes. Esta tesis doctoral tuvo como objetivo explorar el uso de dos técnicas ópticas no destructivas: imagen hiperespectral (HSI) y espectroscopía Raman, combinadas con modelos avanzados de redes neuronales, para predecir atributos de calidad interna, aceptación del consumidor y seguridad alimentaria en fresas, caquis, ciruelas y manzanas.

La investigación incluyó cuatro estudios principales. Primero, se aplicaron HSI y redes neuronales a 17 cultivares de fresa para predecir parámetros clave de calidad como contenido de azúcar, acidez y textura, así como la aceptación general del consumidor e identificación de cultivares para prevenir fraudes alimentarios. Segundo, se usaron modelos de aprendizaje profundo basados en autoencoders para clasificar la imagen hiperespectral de caquis 'Rojo Brillante' en grupos de textura relevantes para la clasificación en centrales hortofrutícolas. Tercero, se entrenaron y compararon tres redes neuronales convolucionales (CNN) para detectar golpes invisibles en ciruelas 'Presenta', con el objetivo de reducir pérdidas poscosecha. Cuarto, la espectroscopía Raman combinada con CNN permitió la detección e identificación de múltiples compuestos activos de productos fitosanitarios en mezclas complejas comúnmente usadas en huertos de manzana.

Los resultados demostraron que HSI y redes neuronales proporcionaron buenas predicciones para la calidad de la fresa (TSS R^2 = 0.85, TA R^2 = 0.81), aceptación media del consumidor por cultivar (R^2 = 0.78) y discriminación de cultivares (F1 = 84 %). En caquis, los autoencoders lograron una reducción del 99 % en la dimensionalidad de los datos y un 81.3 % de precisión en la clasificación de textura. Las CNN detectaron golpes en etapa temprana en ciruelas con una puntuación F1 del 90 % y la identificación de longitudes de onda clave. La espectroscopía Raman con CNNs identificó la presencia de fitosanitarios en mezclas complejas con hasta un F1 de 100 % en algunos casos.

Estos resultados destacan el valor práctico de combinar sensores ópticos avanzados y redes neuronales para una evaluación rápida y no destructiva de la calidad y seguridad de la fruta. Este enfoque ofrece beneficios significativos para consumidores, productores y autoridades de seguridad alimentaria al mejorar la calidad del producto, reducir el desperdicio y garantizar el cumplimiento de las normativas de seguridad.

RESUM

Assegurar una bona qualitat interna e inocuitat en fruites és un repte important per a la indústria agroalimentària, ja que els mètodes tradicionals d'avaluació sovint són destructius i ineficients. Aquesta tesi doctoral ha tingut com a objectiu explorar l'ús de dues tècniques òptiques no destructives: la imatge hiperespectral (HSI) i l'espectroscòpia Raman, combinades amb models avançats de xarxes neuronals, per predir atributs de qualitat interna, acceptació del consumidor i seguretat alimentària en maduixes, caquis, prunes i pomes.

La investigació va incloure quatre estudis principals. Primer, es van aplicar la imatge hiperespectral i les xarxes neuronals a 17 cultius de maduixa per predir paràmetres clau de qualitat com el contingut de sucre, l'acidesa i la textura, així com l'acceptació global del consumidor i la identificació dels cultius per prevenir el frau alimentari. En segon lloc, es van utilitzar models d'aprenentatge profund basats en autoencoders per classificar la imatge hiperespectral de caqui 'Rojo Brillante' en grups de textura rellevants per a la classificació a les centrals de confecció. Tercer, es van entrenar i comparar tres xarxes neuronals convolucionals (CNN) per detectar cops invisibles en prunes 'Presenta', amb l'objectiu de reduir les pèrdues postcollita. Quart, l'espectroscòpia Raman combinada amb CNN va permetre la detecció i identificació de múltiples compostos actius de plaguicides en mescles complexes de plaguicides habituals als pomerars.

Els resultats van demostrar que la imatge hiperespectral i les xarxes neuronals van oferir prediccions precises per a la qualitat de la maduixa (TSS R^2 = 0,85, TA R^2 = 0,81), l'acceptació mitjana del consumidor per cultivar (R^2 = 0,78) i la discriminació dels cultius (F1 = 84 %). En caquis, els autoencoders van aconseguir una reducció del 99 % en la dimensionalitat de les dades i un 81,3 % d'exactitud en la classificació de la textura. Les CNN van detectar cops en fase inicial en les prunes amb una puntuació F1 del 90 % i es van identificar longituds d'ona clau. L'espectroscòpia Raman amb models d'aprenentatge profund va identificar la presencia de plaguicidas en mescles complexes amb un F1 de fins al 100% en alguns casos.

Aquests resultats posen de manifest el valor pràctic de combinar tècniques òptiques avançades i xarxes neuronals per avaluar, de manera ràpida i no destructiva, la qualitat i seguretat dels fruits. Aquest enfocament ofereix beneficis significatius per als consumidors, productors i autoritats de seguretat alimentària, millorant la qualitat del producte, reduint el malbaratament i assegurant el compliment de les normes de seguretat.

PREFACE

Research framework

This doctoral thesis is the result of the research work accomplished by the author during the period 2021-2025 as a member of the Centro de Agroingeniería, which belongs to the Instituto Valenciano de Investigaciones Agrarias (IVIA) with an FPI-INIA grant (PRE2020-094491), partially supported by the European Regional Development Funds (FEDER). This thesis is part of two research projects funded by the INIA with the support of FEDER funds, in which I have participated:

- PID2019-107347RR-C31. Inspección no destructiva y predicción de la calidad interna y propiedades de las frutas mediante espectroscopia VIS/NIR y modelos basados en aprendizaje profundo. This doctoral thesis was possible thanks to the scholarship associated to this project with the number PRE2020-094491.
- PID2023-150192OR-C31. Automatización de la inspección de la calidad interna y seguridad de frutas en línea, utilizando espectroscopia VIS/NIR e inteligencia artificial.

Dissemination of the results

Publications in international journals of JCR and SJR

Published

- Castillo-Girones, S., Arnould, Q., Gómez-Sanchis, J., Blasco, J., Pigeon, O., Baeten, V., & Fernández Pierna, J. A. (2025). Raman spectroscopy for multi-label identification of common apple pesticide mixtures using CNNs and gradient-weighted class activation mapping. Food Control, 178, 111460. https://doi.org/10.1016/J.FOODCONT.2025.111460
- Salvador Castillo-Gironés, Jos Ruizendaal, Xiomara Salas-Valderrama, Sandra Munera, Jose Blasco, Gerrit Polder. (2025). Advanced evaluation of strawberry quality, consumer preference, and cultivar discrimination through spectral imaging and neural networks. Food Control, 175, 111339. https://doi.org/https://doi.org/10.1016/j.foodcont.2025.111339
- Salvador Castillo-Gironés, Sandra Munera, Marcelino Martínez-Sober, Jose Blasco, Sergio Cubero, Juan Gómez-Sanchís. (2025). Artificial Neural Networks in Agriculture, the core of artificial intelligence: What, When, and Why. Computers and Electronics in Agriculture, 230, 109938. https://doi.org/10.1016/J.COMPAG.2025.109938
- Salvador Castillo-Gironés, Remi Van Belleghem, Niels Wouters, Sandra Munera, Jose Blasco, Wouter Saeys. (2024). Detection of subsurface bruises in plums using spectral imaging and deep learning with wavelength selection. Postharvest Biology and Technology, 207, 112615. https://doi.org/https://doi.org/10.1016/j.postharvbio.2023.112615

- Salvador Castillo-Gironés, Sandra Munera., Alejandro Rodríguez, Nuria Aleixos., Sergio Cubero, Marina López-Chuliá, Nerea Martínez-Onandi, Jose Blasco. (2023). Using Vis/NIR hyperspectral imaging and wavelength selection for accurate postharvest discrimination of loquat cultivars of similar appearance. Acta Horticulturae 1396, 215-220. DOI: 10.17660/ActaHortic.2024.1396.29

<u>Submitted - Under review</u>

- Salvador Castillo-Gironés, Ángel González, Sandra Munera, Marcelino Martínez-Sober, Sergio Cubero, Nariane Quaresma Vilhena, Iván Blanco-Álvarez, Juan Gómez-Sanchís. Firmness classification in 'Rojo Brillante' persimmon fruit after cold storage using hyperspectral imaging and deep learning with variational autoencoders. LWT.

Publications in national and international conferences

<u>Oral communications</u>

- Salvador Castillo-Gironés, Marina López-Chuliá, Sergio Cubero, Sandra Munera, Alejandro Rodríguez, Juan Gómez-Sanchis, Nuria Aleixos, Jose Blasco. Assessing and Predicting the Evolution of Persimmon Fruit Flesh Texture During Cold Storage Using Hyperspectral Imaging. IX International Postharvest Symposium. Rotorua, New Zealand. 11th - 15th November 2024
- Salvador Castillo-Gironés, Remi Van Belleghem, Marina López-Chuliá, Niels Wouters, Sandra Munera, Jose Blasco, Wouter Saeys. Detection of invisible damages in plums with deep learning. Final SensorFINT International Conference. Córdoba, Spain. 29th - 31st May 2024
- Salvador Castillo-Gironés, Marina López Chuliá, Sergio Cubero, Sandra Munera, Alejandro Rodríguez, Juan Gómez-Sanchís, Nuria Aleixos, Jose Blasco. Evolución de la textura del caqui almacenado en frío a dos temperaturas mediante imagen hiperespectral. XII Congreso Ibérico y X Congreso Iberoamericano de Ciencias y Técnicas del Frío. Elche, Spain. 26th - 28th June 2024
- Salvador Castillo-Gironés, Sergio Cubero, Sandra Munera, Alejandro Rodríguez, Juan Gómez-Sanchís, Nuria Aleixos, Jose Blasco. Determinación de la textura en caqui `Rojo brillante' mediante imagen hiperspectral Vis-NIR. XII Congreso Ibérico de Agroingeniería. Sevilla, Spain. 4th - 6th September 2023.
- Salvador Castillo-Gironés, Sergio Cubero, Marina López-Chuliá, Sandra Munera, Alejandro Rodríguez, Juan Gómez-Sanchís, Nuria Aleixos, Jose Blasco. Uso de imagen hiperespectral para la discriminación en postcosecha de variedades similares de níspero. XII Congreso Ibérico de Agroingeniería. Sevilla, Spain. 4th - 6th September 2023.
- Salvador Castillo-Gironés, Sandra Munera, Alejandro Rodríguez, Nuria Aleixos, Sergio Cubero, Marina López-Chuliá, Nerea Martínez-Onandi, Jose Blasco. Using Vis/NIR hyperspectral imaging and wavelength selection for accurate postharvest discrimination

of similar loquat cultivars. VII International Conference Postharvest Unlimited & XII International Symposium on Postharvest Quality of Ornamental Plants. Wageningen, Netherlands. 14th - 17th May 2023.

- Salvador Castillo-Gironés, Sandra Munera, Alejandro Rodríguez, Nuria Aleixos, Juan Gómez-Sanchís, Sergio Cubero, Rodolfo Canet, Jose Blasco. Discrimination of similar cultivars of loquat fruit using hyperspectral imaging. XX CIGR World Congress 2022. Kyoto, Japan. 5th - 10th December 2022
- Salvador Castillo-Gironés, Ángel González, Sergio Cubero, Sandra Munera, Alejandro Rodríguez, Nariane Quaresma Vilhena, Alejandra Salvador, Juan Gómez-Sanchís, Nuria Aleixos, Jose Blasco. Autoencoder-based neural network and Partial Least Square Discriminant Analysis to classify 'Rojo Brillante' persimmon in texture categories from hyperspectral images. XX CIGR World Congress 2022. Kyoto, Japan. 5th - 10th December 2022
- Salvador Castillo-Gironés, Alejandro Rodríguez, Sandra Munera, Alejandra Salvador, Juan Gómez-Sanchís, Nuria Aleixos, Jose Blasco. Use of hyperspectral imaging to classify 'Rojo Brillante' persimmon in three texture classes before and after storage. 1st SensorFINT International Conference. Izola, Slovenia. 10th - 12th May 2022

Posters

- Salvador Castillo-Girones, Marina Lopez-Chulia, Sergio Cubero, Sandra Munera, Alejandro Rodríguez, Juan Gómez-Sanchís, José Blasco. Loquat internal quality prediction and variety classification with hyperspectral imaging. 22nd International Conference on Near Infrared Spectroscopy. Rome, Italy, 8-12 June 2025
- Salvador Castillo-Girones, Raquel Martínez-Peña, Sara Álvarez, Sergio Vélez. Hyperspectral imaging for predicting fat (oleic acid) and protein content of Spanish pistachios. 22nd International Conference on Near Infrared Spectroscopy. Rome, Italy, 8-12 June 2025
- Xueping Yang, Lorenzo Serva, Kuikui Ni, Fuyu Yang, Paolo Berzaghi, Salvador Castillo-Girones. CNN-based methods vs. KNN-weighted PLSR: a comparative analysis for protein and ADF prediction across multi-product feed. 22nd International Conference on Near Infrared Spectroscopy. Rome, Italy, 8-12 June 2025
- Salvador Castillo-Girones, Marina Lopez Chulia, Sergio Cubero, Sandra Munera, Alejandro Rodríguez, Juan Gómez-Sanchís, Alejandra Salvador, Nariane Vilhena, Ana Moreno, Jose Blasco. Non-destructive texture prediction of persimmon texture using hyperspectral imaging. 15th International FRUTIC Symposium 2025. Rimini, Italy. 7-8 May 2025
- Marina Lopez-Chulia, Salvador Castillo-Girones, Alejandra Salvador, Pau Talens, Jose Blasco. Prediction of Acidity and Soluble Solid Content Changes In Loquat Fruit During Cold Storage through Non-Destructive Spectral Methods. 15th International FRUTIC Symposium 2025. Rimini, Italy. 7-8 May 2025
- Salvador Castillo-Gironés, Marina López-Chuliá, Salvador Castillo, Daniel Valero, María Serrano, Sergio Cubero, Jose Blasco. Internal quality and colour prediction of blood

oranges using NIR hyperspectral imaging. 9th International Conference in Spectral Imaging. Bilbao, Spain. 6th - 10th July 2024

- Iván Blanco-Álvarez, Marcelino Martínez-Sober, Sandra Munera, Salvador Castillo-Gironés, Sergio Cubero, Ana Quiñones, Nuria Aleixos, Jose Blasco-Ivars, Juan Gómez-Sanchis. Estimation of nutritional levels on citrus leaves using spectral imaging and machine learning techniques. 8th International Conference in Spectral Imaging. Bilbao, Spain. 6th - 10th July 2024
- Salvador Castillo Gironés, Ricardo Felipe Lima de Souza, Sandra Munera, Alejandro Rodríguez, Sergio Cubero, Juan Gómez-Sanchís, Nerea Martínez-Onandi, Marina López-Chuliá, Jose Blasco. Early detection of Penicillium Digitatum using hyperspectral imaging and deep learning. 21st International Conference on Near Infrared Spectroscopy (NIRS 2023). Innsbruck, Austria. 20th - 24th August 2023
- Salvador Castillo-Gironés, Remi Van Belleghem, Sandra Munera, Juan Gómez-Sanchís, Jose Blasco, Wouter Saeys. Detection of hidden bruises in plums using hyperspectral imaging and a 3D convolutional neural network. 7th International Symposium on Applications of Modelling as an Innovative Technology in the Horticultural Supply Chain (Model-IT 2023). Potsdam, Germany. 11th - 14th June 2023
- Salvador Castillo-Gironés, Alejandro Rodríguez-Ortega, Sandra Munera, Nuria Aleixos, Sergio Cubero, Jose Blasco. Hyperspectral imaging and a deep neural network for detection of damages in 'rojo brillante' persimmon. 8th International Conference in Spectral Imaging (IASIM 2022). Esbjerg, Denmark. 3rd - 6th July 2022
- Sandra Munera, Vicente Serna-Escolano, Salvador Castillo-Gironés, Pedro J. Zapata, Jose Blasco, Alicia Dobón-Suárez, Sergio Cubero. Non-destructive evaluation of quality, cultivar and origin of lemons using visible and near-infrared spectroscopy. 14th International FRUTIC Symposium. Valencia, Spain. 29th June - 1st July 2022
- Alejandro Rodríguez-Ortega, Sandra Munera, Salvador Castillo-Gironés, Nuria Aleixos, Sergio Cubero, Jose Blasco. Deep neural network using to automatic detection of hiding bruise in 'rojo brillante' persimmon fruit using hyperspectral imaging. 14th International FRUTIC Symposium. Valencia, Spain. 29th June - 1st July 2022
- Sandra Munera, Salvador Castillo-Gironés, Juan Gómez-Sanchís, Nuria Aleixos, Sergio Cubero, and Jose Blasco. Discriminación de defectos internos y externos en níspero 'Algerie' usando imagen hiperespectral Vis/NIR y técnicas de aprendizaje automático. XIII Simposio Nacional y XI Ibérico Sobre Maduración y Postcosecha (POST 2022). Zaragoza, Spain. 15th - 17th June 2022
- Alejandro Rodríguez-Ortega, Sandra Munera, Salvador Castillo-Gironés, Jose Blasco, Sergio Cubero, Nuria Aleixos. A non-destructive method to measure the light penetration depth and optical properties of "Rojo Brillante" persimmons. 1st SensorFINT International Conference. Izola, Slovenia. 10th - 12th May 2022

Predoctoral stays at foreign institutions

- Research stay at the Wrocław University of Environmental and Life Sciences (Wrocław, Poland). 1st – 30th April 2025. Different drying techniques were applied to pumpkin snacks, and non-destructive evaluation was performed using hyperspectral imaging and X-ray tomography. This research stay was possible thanks to the PROM Project – Short-Term Academic Exchange by the Polish National Agency for Academic Exchange (NAWA).
- Research stay at the Walloon Agricultural Research Centre (Gembloux, Belgium). 1st – 31st March 2024. Identification of pesticide-active compounds in complex pest mixtures commonly applied in apples using Raman spectroscopy. March 2024. This research stay was possible thanks to the PRE2020-094491 associated with the project PID2019-107347RR-C31.
- Research stay at Wageningen University and Research (Wageningen, Netherlands). 17th April – 17th July 2023. Prediction of internal quality, variety identification and consumer acceptance of strawberries using hyperspectral imaging. April-June 2023. This research stay was possible thanks to the PRE2020-094491 associated with the project PID2019-107347RR-C31.
- Research stay at KU Leuven (Leuven, Belgium). 1st – 30th September 2022. Invisible damages in plums using hyperspectral images. September 2022. This research stay was possible thanks to the EU COST ACTION CA19145 Short-Time Scientific Mission (STSM).

INTRODUCTION

The postharvest industry is a crucial component of the fruit industry, particularly in Spain and across Europe, where it plays a critical role in ensuring the quality, safety, and marketability of fresh fruits. This sector is vital for its contribution to food security and its significant economic impact. In Spain, the agricultural sector comprised 8.9 % of the national GDP in 2022 (MAPA, 2025), emphasising its important economic role. Spain is also one of Europe's leading producers of fruits and vegetables, with annual outputs exceeding 13 million tonnes of fresh fruit and nearly 10 million tonnes of fresh vegetables (Koch & Mangelberger, 2025). However, the industry faces significant challenges due to post-harvest losses, which amount to approximately 30 % of fruits and vegetables globally (Bisht & Singh, 2024). In Spain, postharvest losses total hundreds of millions of euros each year. For example, persimmons, a key crop in the Valencia region, can experience an 18.7 % postharvest loss due to bruising and fungal infections (Fernandez-Zamudio et al., 2020). This underscores the need for advanced technologies to mitigate such losses while maintaining high-quality standards.

Accurately predicting internal quality parameters such as texture, sugar content, and acidity is essential for various stakeholders across the agricultural value chain. For breeders, assessing these attributes non-destructively facilitates the selection of cultivars with optimised traits that are both commercially viable and appealing to consumers. Quality measurements are often destructive, which limits the ability to monitor how these traits develop over time and under various storage conditions. For consumers, reliable internal quality prediction ensures that fruits meet expectations for flavour, nutritional value, and safety. Texture influences eating not only pleasure but also nutrient retention. Sugar-acid ratios directly affect consumer taste perception, with imbalances leading to consumer rejection and wasted purchases (Crisosto et al., 2005).

In the agricultural industry, quality estimation provides both economic and operational advantages. Traditional destructive testing methods can lead to financial losses by sacrificing sellable products for testing purposes. Additionally, sampling errors can result in the misgrading of entire batches.

Besides, detecting pesticide residues is crucial for protecting public health, as the incorrect or illegal use of chemicals can lead to harmful residues that are associated with chronic health risks. Adhering to EU regulations regarding maximum residue limits and identifying prohibited pesticides is not just a regulatory formality; it is essential for preventing exposure to toxic substances and maintaining consumer trust in food safety standards. Furthermore, detecting pesticide residues is crucial to avoid penalties for non-compliance, including shipment rejections, legal liabilities, and exclusion from profitable markets such as the EU.

Non-destructive technologies have emerged as transformative tools in this context, offering innovative solutions to assess fruit quality, detect defects, and ensure food safety without damaging the product while enabling whole-batch analysis. Among these technologies, hyperspectral imaging has proven particularly effective for assessing fruit quality and reducing food waste. Hyperspectral cameras capture detailed spectral data across hundreds of wavelengths, providing insights into internal quality attributes such as ripeness, firmness, sugar content, and the presence of defects like bruises or fungal infections (Wang et al., 2023). This capability allows producers to sort fruits more efficiently, ensuring that only high-quality products reach consumers while reducing waste from rejected shipments.

Raman spectroscopy is a non-destructive technology that has gained importance in ensuring food safety. This technique can accurately identify chemical residues on surfaces, making it invaluable for detecting pesticide contamination without altering the sample. In addition to identifying pesticides, Raman spectroscopy effectively detects mycotoxins and pathogens (Huang et al., 2025) that can pose risks to human health. These capabilities make Raman spectroscopy an essential tool for ensuring compliance with legislation while ensuring consumer safety and trust.

Thus, integrating these non-destructive technologies into postharvest operations has substantial economic and environmental benefits. By reducing food waste and improving sorting accuracy, these systems help producers maximise their yields and revenues while minimising their carbon footprint. Moreover, these technologies contribute to sustainability by eliminating the need for destructive testing methods that generate waste.

Looking ahead, Artificial Neural Networks (ANNs) are increasingly being applied to enhance the quality and safety of fruits and vegetables, offering significant advancements over traditional methods (Castillo-Girones et al., 2025). These systems are particularly valuable in non-destructive quality assessment, identifying non-linear relationships in complex data such as hyperspectral imaging or Raman spectroscopy. By automating feature extraction and learning from large datasets, ANNs achieve higher accuracy and efficiency than conventional machine learning models. The adaptability of ANNs significantly increases their usefulness (Attri et al., 2023).

1. Spectral Imaging and Raman Spectroscopy

1.1 Hyperspectral Imaging

Hyperspectral imaging (HSI) is a technology built upon two foundational scientific disciplines: spectroscopy and imaging. Spectroscopy focuses on the interaction between light and matter, analysing how materials absorb, reflect, or emit light at various wavelengths. This analysis reveals unique spectral fingerprints that provide insights into a material's composition and properties. For instance, spectroscopy can identify specific chemical compounds or detect subtle differences in material characteristics based on their spectral behaviour. Imaging, on the other hand, captures spatial information about objects or scenes by recording light intensity at specific wavelengths. Unlike conventional imaging systems, which typically operate within the visible spectrum and record data in three broad bands (Red, Green and Blue) or a limited number of discrete spectral bands, as in multispectral imaging, HSI merges these two approaches (spectroscopy and imaging) into a single system capable of capturing both spatial and spectral information simultaneously. This process generates a three-dimensional dataset, commonly called a hyperspectral data cube (D. W. Sun, 2010).

The hyperspectral data cube is the foundation for this technology's analytical power. Its spatial dimensions (x, y) provide a detailed representation of the physical layout of the imaged scene. In contrast, its spectral dimension (λ) contains reflectance or radiance values across hundreds of contiguous wavelength bands for each pixel. This combination allows hyperspectral imaging to deliver both spatial context and material-specific information in a single dataset. In practice, this means that HSI can simultaneously map the spatial distribution of objects or features within a scene while providing detailed insights into their material composition. By recording the full spectrum of light at each pixel in an image, HSI a transformative tool across a wide range of fields and is very useful for fruit quality and safety assessment (Pathmanaban et al., 2019; Qin et al., 2013), as in the case of the present work.

1.1.1 Principles and Technologies

HSI systems are specialised instruments that disperse incoming light into its constituent wavelengths and record this information for each pixel. These cameras are composed of several key components that work together to construct the hyperspectral data cube, which is a three-dimensional dataset that provides detailed insights into the composition and properties of materials.

At the core of a hyperspectral camera is its optical system, which focuses light from the scene onto the sensor. The optics must be carefully designed to handle a wide range of wavelengths, often ranging from the visible to the near-infrared or even further into the electromagnetic spectrum. This requires high-quality lenses with broadband anti-reflective coatings to maximise light transmission while minimising distortions and losses. The optical system must also be robust enough to maintain performance under varying environmental conditions, such as temperature fluctuations, which can affect refractive indices and alignment (Amigo & Grassi, 2019).

The slit is another essential component of most hyperspectral cameras. It limits the field of view to a narrow line or point, ensuring that only a specific portion of the scene is analysed at any given time. By isolating this small section, the slit allows for precise spectral measurements while reducing stray light and noise that could interfere with data quality. The width of the slit is carefully calibrated to balance spatial resolution with spectral resolution; a narrower slit improves spectral precision but may reduce spatial detail. Once light passes through the slit, it encounters a dispersive element, such as a diffraction grating, a prism or tunable filters, depending on the HIS technique, which separates it into its individual spectral components. Diffraction gratings are commonly used for their ability to disperse light across a wide range of wavelengths efficiently. These gratings operate by utilising interference patterns that occur when light interacts with finely spaced grooves on their surfaces. Alternatively, some systems can use prisms because they refract light based on its wavelength. However, prisms are generally less common due to their bulkier design and lower efficiency at extreme wavelengths. Tunable filters allow only light transmission at specific wavelengths at any given time (Kerekes & Schott, 2006; Lodhi et al., 2019).

The dispersed light is finally captured by a detector array, which records the intensity of each wavelength for every spatial position within the slit's field of view. Detector arrays are typically composed of charge-coupled devices (CCDs) or complementary metal-oxide-semiconductor (CMOS) sensors, both sensitive to light across specific spectral ranges, for applications requiring sensitivity beyond the visible spectrum, such as in the near-infrared (NIR) or shortwave infrared (SWIR), specialised detectors made from materials like indium gallium arsenide (InGaAs) are employed (Okubo, 2021).

The illumination or lighting system is a crucial yet often overlooked component of hyperspectral imaging systems. Proper illumination is essential for ensuring accurate and reliable data collection, as it directly influences reflectance measurements, image quality, and overall system performance. A well-designed lighting setup ensures uniform illumination across the scene, vital for avoiding inconsistencies in reflectance values that could compromise analytical results. To ensure accurate spectral data acquisition, the lighting system must cover the full range of wavelengths the hyperspectral camera requires. For example, halogen-tungsten lamps are commonly used due to their broad spectral output, but other sources may be chosen based on specific application needs. Uniformity in lighting eliminates dark or overly saturated regions in images, which can distort reflectance values and hinder material identification or prediction accuracy. Additionally, appropriate lighting intensity prevents underexposure or overexposure of certain spectral bands. Overexposure can lead to saturation in detector responses, while underexposure can result in poor signal-to-noise ratios, reducing the reliability of spectral signatures. These factors are particularly important when working with materials that exhibit subtle spectral differences or when high precision is required (Amigo, 2019).

In dynamic scenarios such as agricultural monitoring or industrial inspection, consistent lighting also minimises artefacts caused by shadows or varying illumination angles. This consistency enhances the system's predictive capabilities by ensuring that reflectance measurements accurately correspond to material properties rather than external lighting variations.

1.1.2 Types of Hyperspectral Imaging

The construction of hyperspectral data cubes depends on the imaging technique used. There are four primary techniques: point scanning, area scanning, line scanning, and snapshot imaging. Each method has distinct advantages and limitations, affecting how spatial and spectral information is captured (Figure 1). The choice of the technique is often dictated by the application requirements, such as the need for speed, resolution, or adaptability to dynamic environments.

1.1.2.1. Point scanning.

Point scanning (Figure 1A) is one of the earliest and most precise methods, known as the whiskbroom technique. In this approach, a single-point spectrometer captures spectral data one pixel at a time. The system systematically scans across the scene, pixel by pixel, to construct the hyperspectral data cube. This method provides exceptional spectral resolution because each pixel is measured individually with high precision. Its main drawback is its slow acquisition speed, which makes it unsuitable for situations that require rapid data collection or monitoring of dynamic scenes. Consequently, point scanning is primarily used in controlled laboratory environments where time constraints are less critical and high spectral fidelity is essential (Amigo & Grassi, 2019; D. W. Sun, 2010).

1.1.2.2. Area scanning.

Another approach is area scanning (Figure 1B), also known as spectral scanning. This method captures an entire two-dimensional spatial image at a single wavelength during each exposure. The hyperspectral data cube is constructed over time by sequentially changing wavelengths across the desired spectral range. Area scanning excels in capturing detailed spatial information for static scenes with minimal or absent motion. However, its slower acquisition speed renders it less effective for dynamic environments where objects or conditions change rapidly during data collection. This limitation restricts its use to specific applications like laboratory analysis or imaging of stationary objects (Grahn & Geladi, 2007; Kerekes & Schott, 2006).

1.1.2.3. Line scanning.

Line scanning (Figure 1C) is a more widely adopted technique for practical applications, commonly called pushbroom. Instead of capturing individual pixels, this approach records an entire line of spatial data at once using a slit and dispersive optics. The sensor or scene moves relative to each other, often by mounting the sensor on platforms like aircraft, drones, or conveyor belts. This setup progressively builds the hyperspectral data cube line by line. Line scanning effectively balances speed and resolution, making it the most popular method for applications like agricultural monitoring, environmental studies, and industrial inspection. Its ability to capture high-resolution data while maintaining reasonable acquisition speeds makes it ideal for large-scale or dynamic environments where efficiency is crucial (Lodhi et al., 2019; D. W. Sun, 2010).

1.1.2.4. Snapshot.

The most advanced and fastest method is snapshot imaging (Figure 1D), which captures the entire hyperspectral data cube in a single exposure. This technique relies on innovative sensor designs to simultaneously acquire both spatial and spectral information across all dimensions of the data cube. Snapshot imaging is particularly suited for real-time applications where rapid data acquisition is essential, such as medical diagnostics, real-time surveillance, or high-speed industrial processes. However, this method often sacrifices spectral resolution compared to point or line scanning techniques due to the complexity of integrating all dimensions into a single exposure (Grahn & Geladi, 2007; Lodhi et al., 2019).

Each imaging technique involves a balance between speed, resolution, and suitability for various scenarios. Area scanning delivers detailed spatial information but struggles with dynamic scenes due to its slower wavelength-by-wavelength acquisition process. Point scanning offers unparalleled spectral precision but at the cost of time efficiency. Line scanning provides a practical balance that suits many field applications requiring moderate speed and resolution. Finally, snapshot imaging prioritises speed and real-time capability but compromises spectral detail in exchange for its rapid performance. Thus, selecting an appropriate imaging technique depends on the specific demands of the application. For example, agricultural monitoring often benefits from line scanning because it can cover large areas efficiently while maintaining sufficient resolution to detect subtle variations in crop health. Conversely, medical diagnostics or industrial quality control may favour snapshot imaging for its ability to provide instantaneous results without sacrificing operational speed.

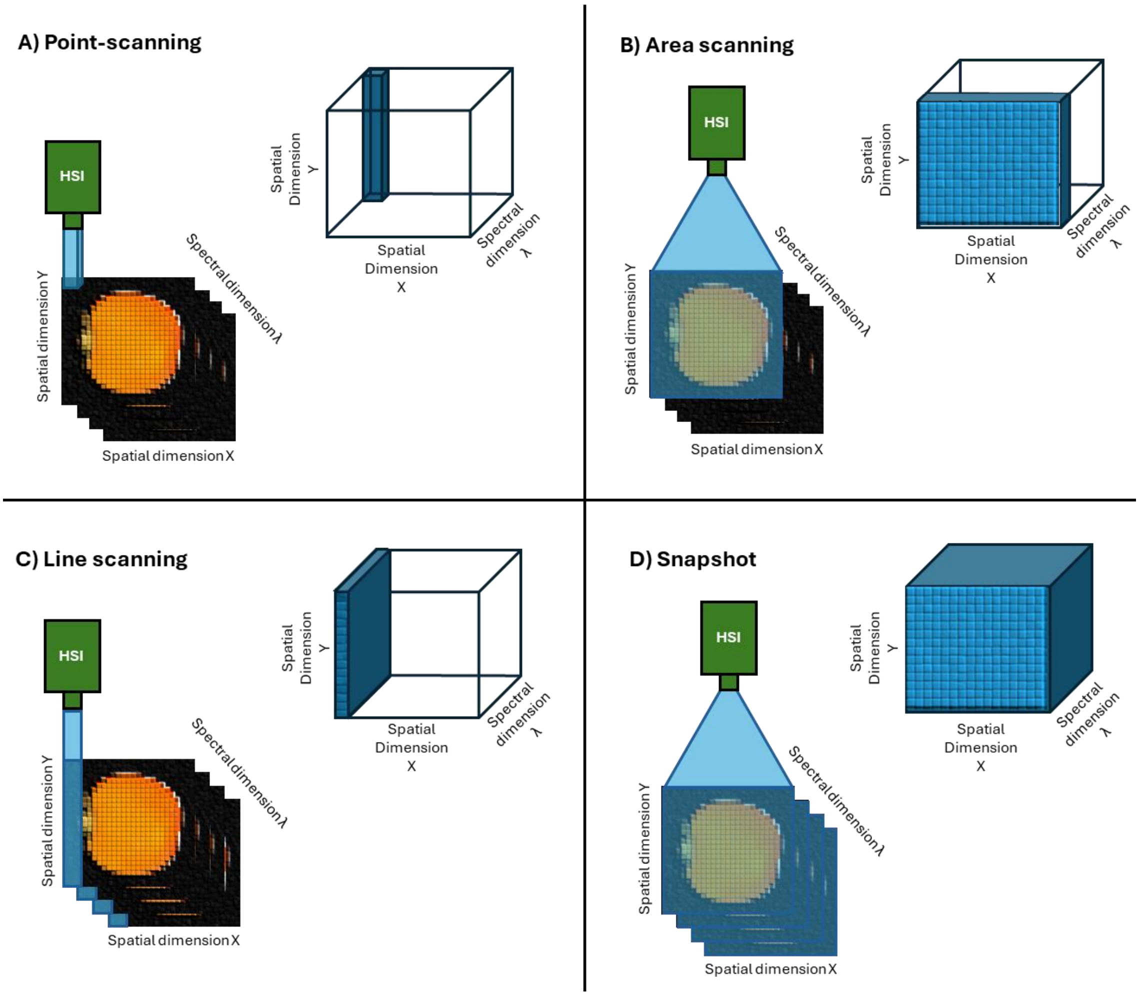


Figure 1. Main hyperspectral imaging techniques

1.1.3 Calibration: Ensuring Data Accuracy

Calibration is a critical process in hyperspectral imaging that ensures the accuracy and reliability of the data collected. Hyperspectral imaging systems capture raw measurements that are influenced by various factors, including sensor characteristics, environmental conditions, and lighting variations. Without proper calibration, these measurements may be corrupted by noise, sensor artefacts, or inconsistencies in illumination, leading to inaccurate representations of physical properties. Calibration transforms raw data into standardised and corrected values, enabling precise analyses and comparisons across different datasets. Calibration in hyperspectral imaging can be categorised into three types: radiometric, spectral, and geometric calibration. Each type addresses specific challenges related to the imaging process (Pillay et al., 2019).

Radiometric calibration is fundamental to converting raw sensor outputs into physical units such as radiance (the amount of light emitted or reflected by a surface) or reflectance (the proportion of light reflected relative to the incident light). This process involves two key steps: dark reference correction and white reference normalisation. Dark reference images are used to account for sensor noise, unwanted signals generated by the detector even in the absence of light. These dark images are typically captured by covering the sensor completely or using a shutter mechanism to block incoming light. Subtracting this noise from the raw data ensures that only meaningful signals are retained. White reference images are obtained by capturing light reflected from a standard white reference material with a known and uniform reflectance across all wavelengths. White reference normalisation corrects for variations in illumination intensity and spectral distribution, ensuring that reflectance values are independent of lighting conditions. By dividing the raw data by the white reference signal, the system compensates for uneven illumination and provides accurate reflectance measurements (Amigo, 2019; D. W. Sun, 2010). Each image taken must undergo radiometric calibration as described. Equation 1 details the method for correcting images at every wavelength:

$$Im_{Refl} = \rho^{Reference}(\lambda)\frac{Im_{Raw(\lambda)} - Im_{Dark(\lambda)}}{Im_{White(\lambda)} - Im_{Dark(\lambda)}}$$

Equation 1. HSI radiometric calibration

Where $\rho^{Reference}(\lambda)$ is the reflectance value of the standard surface at a specific wavelength, $Im_{Dark(\lambda)}$ and $Im_{White(\lambda)}$ are the dark and white refeences at that specific wavelength, and $Im_{Raw(\lambda)}$ is the raw image at that wavelength.

Spectral calibration is the process of aligning the recorded wavelengths of the hyperspectral sensor with their true or reference values to ensure that each recorded spectrum accurately represents the true properties of the material being imaged. This process is essential because hyperspectral sensors often exhibit slight deviations in their wavelength assignments due to manufacturing tolerances or environmental factors such as temperature changes. Spectral calibration typically involves using a light source with known spectral features (e.g., emission lines from a gas lamp or LEDs that emit light within a specific spectral band) to verify and adjust the wavelength mapping of each pixel. Accurate spectral calibration removes errors like wavelength shifts or misalignments, preventing incorrect material identification (Geladi et al., 2004).

Geometric calibration in hyperspectral imaging ensures that the spatial dimensions (x, y) of the hyperspectral data cube accurately correspond to the real-world geometry of the scene being imaged. It corrects for distortions caused by the sensor, optics, or motion during data acquisition, aligning the spatial information with true physical coordinates. This process is essential for accurate spatial analysis and mapping in applications like agriculture or remote sensing (Špiclin et al., 2010).

Thus, calibration ensures consistency across different sensors, imaging setups, and acquisition conditions. This consistency is crucial when comparing datasets collected at different times or under varying environmental circumstances. Furthermore, calibrated hyperspectral data enhances the reliability of downstream analyses such as material classification, anomaly detection, or quantitative modelling.

1.2 Raman Spectroscopy

Raman spectroscopy is a powerful and versatile analytical technique that provides molecular-level insights into the composition, structure, and dynamics of materials. Named after Sir Chandrasekhara Venkata Raman, who discovered the Raman effect in 1928 and was awarded the Nobel Prize in Physics in 1930, this method has become indispensable across various scientific fields. Raman spectroscopy relies on the interaction of light with matter to detect vibrational, rotational, and other low-frequency modes of molecules. This capability allows it to generate detailed "molecular fingerprints" unique to each substance, making it an essential tool for material identification and characterisation. Raman spectroscopy's greatest strength is its ability to provide precise molecular information without requiring extensive sample preparation or destructive testing. It can analyse solids, liquids, and gases directly and is particularly well-suited for studying aqueous samples because water produces only weak Raman signals that do not interfere with measurements (Ferraro et al., 2003).

The fundamental principle of Raman spectroscopy is based on the phenomenon of inelastic scattering of light, known as Raman scattering. When a monochromatic light source, such as a laser, interacts with a sample, most photons are scattered elastically, meaning their energy remains unchanged. This is called Rayleigh scattering. However, a very small fraction of photons (about one in ten million) experience inelastic scattering, where their energy changes due to interactions with the vibrational or rotational movements of the molecules in the sample. The amount of energy change corresponds to the difference between the incident photon's energy and the energy associated with these molecular vibrations or rotations. The collection of these energy shifts forms what is known as the Raman spectrum. The Raman effect explains how this inelastic scattering occurs. When an incident photon interacts with a molecule, it temporarily excites the molecule's electron cloud to a virtual energy state, which is a short-lived and unstable state. As the molecule returns to its original ground state, it emits scattered light with a slightly different energy than the incoming photon. If the scattered photon loses energy because the molecule absorbs it and transitions to a higher vibrational state, this is called Stokes scattering. Conversely, if the scattered photon gains energy because the molecule is already in an excited state and transitions to a lower vibrational state, this is called anti-Stokes scattering (Dietzek et al., 2018; Moore, 2014).

1.2.1 Principles and Comparison with Infrared Spectroscopy

A typical Raman spectrometer comprises different components (Figure 2) that work together to capture and analyse Raman spectra: the laser source, the optical system, the spectrograph and a computer/device to process the data. The laser source provides monochromatic light to excite the sample. Common laser wavelengths include 532 nm (visible), 785 nm (near-infrared), and 1064 nm (infrared), with each wavelength having specific advantages depending on factors such as fluorescence interference and spectral resolution requirements. Shorter wavelengths generally produce stronger Raman signals but are more prone to fluorescence interference, while longer wavelengths reduce fluorescence but may yield weaker signals (Moore, 2014).

The optical system is designed to focus the laser beam onto the sample and efficiently collect the scattered light. This system uses carefully aligned lenses and mirrors to capture as much light as possible while minimising losses. The scattered light then passes through filters, which block intense Rayleigh-scattered light while allowing Raman-shifted light to pass through for analysis. The filtered light is then directed into a spectrograph, where a diffraction grating separates it into its constituent wavelengths. The diffraction grating works by dispersing the light based on its wavelength, much like how a prism splits white light into a rainbow. This scattered light is captured by a detector, typically a charge-coupled device (CCD), which converts the optical signal into digital data. A computer processes the spectral data to generate a Raman spectrum, which is a plot of intensity versus wavelength shift relative to the incident laser frequency. This Raman spectrum contains peaks at specific positions corresponding to the molecules' vibrational modes in the sample. Each peak provides information about a particular molecular vibration, which is directly related to the chemical bonds and structure of the sample (Agarwal & Atalla, 2020; Moore, 2014).

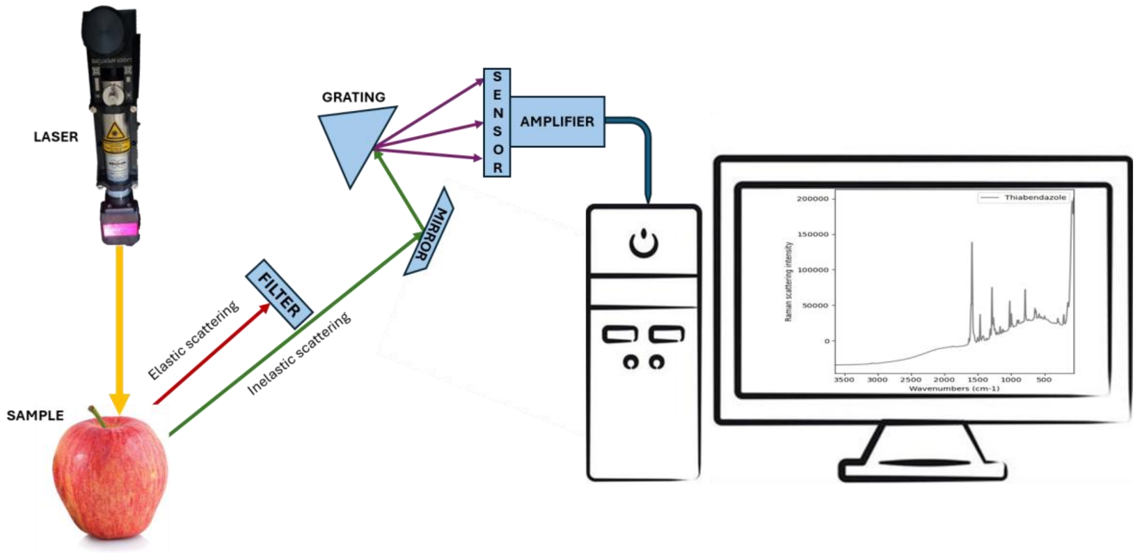


Figure 2. Raman spectroscopy principles

Unlike infrared (IR) spectroscopy, Raman spectroscopy detects changes in molecular polarizability instead of relying on changes in dipole moments. This nature allows Raman spectroscopy to identify vibrational modes that may not be visible in IR spectra. Both are complementary vibrational spectroscopic techniques that provide molecular-level insights into the structure and composition of materials. While both analyse molecular vibrations to generate unique spectral fingerprints, their fundamental principles and applications differ, making each suited to specific analytical needs (Abbas et al., 2020).

The key difference lies in their mechanisms. IR spectroscopy measures light absorption caused by changes in molecular dipole moments, making it highly sensitive to polar bonds such as O-H and C=O. On the other hand, Raman spectroscopy relies on inelastic scattering of light, detecting changes in molecular polarizability. This makes it particularly effective for non-polar bonds like C=C or symmetric vibrations that produce weak IR signals. One of Raman's significant advantages is its compatibility with aqueous samples. Water strongly absorbs infrared light, interfering with IR spectra, whereas it produces weak Raman scattering, allowing precise analysis of samples in aqueous environments. The two techniques also differ in their applications. IR excels at identifying functional groups in organic compounds and polymers due to its sensitivity to polar bonds and detailed fingerprint regions. Raman complements this by providing strong signals for symmetric and polarisable bonds, making it ideal for analysing inorganic compounds, crystalline structures, and complex mixtures. Despite these differences, Raman and IR spectroscopy are complementary. By combining both methods, researchers can comprehensively understand molecular structures and interactions (Abbas et al., 2020; Dietzek et al., 2018).

1.2.2 Laser Selection in Raman Spectroscopy

The selection of the laser wavelength in Raman spectroscopy is a critical decision that directly impacts the quality and interpretability of the resulting Raman spectra. Since the Raman effect relies on the interaction between monochromatic light and molecular vibrations, the wavelength of the laser determines not only the intensity of the Raman scattering but also the potential for interference from fluorescence and other background signals. Understanding these trade-offs is essential for optimising spectral acquisition and ensuring accurate analysis. Shorter wavelength lasers, such as those in the visible spectrum (532 nm laser is commonly used), are often chosen for their ability to produce higher Raman scattering intensity. This is because the efficiency of Raman scattering increases with shorter wavelengths, as it is inversely proportional to the fourth power of the wavelength. As a result, shorter wavelengths can generate stronger signals, which enhances the sensitivity of detection and allows for better resolution of fine spectral features. However, shorter wavelengths are more likely to induce fluorescence in certain samples. Fluorescence occurs when molecules absorb high-energy photons and re-emit light at longer wavelengths, creating a background signal that can overpower or mask weaker Raman signals. This is particularly problematic when analysing organic or biological samples, which often contain fluorescent compounds (Eckhardt, 1966; Panczer et al., 2012).

To mitigate fluorescence interference in Raman spectroscopy, lasers with longer wavelengths, such as those operating at 785 nm and 1064 nm, are frequently used. These wavelengths fall within the NIR region and are less energetic than shorter wavelengths like 532 nm. These wavelengths are less energetic and less likely to excite electronic transitions that lead to fluorescence. By reducing fluorescence background, longer wavelength lasers enable more precise detection of Raman peaks, especially in samples with complex matrices or high levels of intrinsic fluorescence. However, longer wavelengths produce weaker Raman signals due to their lower scattering cross-sections. This can result in reduced sensitivity and may require longer acquisition times or more sensitive detectors to compensate for the diminished signal intensity. However, while 785 nm and 1064 nm lasers effectively minimise fluorescence, each has distinct characteristics that make it more suitable for specific applications (Panczer et al., 2012).

The 785 nm laser is one of the most commonly used wavelengths in Raman spectroscopy because it balances fluorescence suppression, signal strength, and cost-effectiveness. It significantly reduces fluorescence compared to visible lasers, like 532 nm, while producing a relatively strong Raman signal. This is because Raman scattering efficiency decreases with increasing wavelength. As a result, the 785 nm laser generates stronger signals than the 1064 nm laser, allowing for faster acquisition times and better sensitivity. Additionally, the Raman signals produced by a 532 and 785 nm laser fall within the detection range of silicon-based CCDs, which are highly sensitive and widely available. It is particularly advantageous in applications where fluorescence is present but not overwhelming, such as pharmaceutical analysis, polymer characterisation, and general material identification. On the other hand, the 1064 nm laser is specifically advantageous for samples with severe fluorescence interference. Its longer wavelength minimises or eliminates fluorescence entirely, making it ideal for highly fluorescent materials such as dyes, pigments, or dark-coloured organic compounds. This wavelength is also less likely to cause sample heating or degradation due to its lower photon energy, which is beneficial for heat-sensitive samples like biological tissues or polymers. However, it produces a weaker Raman signal due to the reduced scattering efficiency at longer wavelengths. To compensate, higher laser powers or longer acquisition times are often required. Additionally, detecting Raman signals at 1064 nm requires specialised detectors like InGaAs arrays because traditional CCDs are ineffective beyond approximately 1100 nm. These detectors are more expensive and have higher intrinsic noise than CCDs (Griffiths, 2001; Kerr et al., 2015).

Ultraviolet (UV) lasers with wavelengths below 300 nm may be used for specific applications requiring extremely high resolution or minimal fluorescence interference. UV lasers offer enhanced spatial resolution due to their short wavelengths and reduced fluorescence because many materials do not fluoresce under UV excitation. Additionally, UV excitation can access higher-energy vibrational modes that are not easily observed with visible or NIR lasers. However, their high energy can damage sensitive samples, particularly biological tissues or delicate materials, through photodegradation or thermal effects. As a result, their use is generally limited to specialised applications where these risks can be carefully managed (Griffiths, 2001).

In summary, the optimal choice of laser wavelength ultimately depends on several factors specific to the sample and experimental aims. For example, if a sample is highly fluorescent, a longer wavelength laser, such as 785 nm or 1064 nm, may be necessary to suppress background noise and improve spectral clarity. Conversely, a shorter wavelength laser, like 532 nm, may be preferred if maximum signal intensity and resolution are required for non-fluorescent samples.

1.2.3 Types of Raman Spectroscopy

Raman spectroscopy has developed significantly over the years, resulting in various techniques designed to address specific scientific and analytical challenges. Each method aims to overcome the limitations of traditional Raman spectroscopy or to take advantage of distinct properties for targeted applications. The key types of Raman spectroscopy used in agriculture include spontaneous Raman spectroscopy, Raman microscopy, and Raman imaging. Besides, to enhance Raman signals, surface-enhanced Raman spectroscopy (SERS) can be combined with any of these techniques (Jones et al., 2019).

1.1.3.1. Traditional Raman.

Spontaneous Raman Spectroscopy, or traditional Raman spectroscopy, is the foundational method that measures standard Stokes or anti-Stokes scattering. It is widely used for general material characterisation and provides a molecular fingerprint of substances. While effective, its sensitivity is limited due to the weak nature of Raman scattering, which has driven the development of more advanced techniques and variations (Jones et al., 2019; Kazlagić & Omanović-Mikličanin, 2020).

1.1.3.2. Raman Microscopy.

Raman Microscopy (µ-Raman) combines Raman spectroscopy with optical microscopy to achieve a high spatial resolution. This integration allows researchers to precisely analyse small samples or specific regions within heterogeneous materials. In the agronomic field, Raman microscopy is particularly valuable for studying plant tissues at the cellular level, identifying chemical compositions in specific regions of seeds, leaves, or roots, and detecting stress markers or nutrient distributions in plants. It is also useful for analysing soil microstructures and localised chemical changes in agricultural samples (Das & Agrawal, 2011; Gierlinger & Schwanninger, 2007).

1.1.3.3. Raman Imaging.

Raman Imaging extends the capabilities of Raman microscopy by producing spatially resolved chemical maps. Acquiring spectra at multiple points across a sample generates detailed images that reveal the distribution of molecular species. This technique is indispensable for mapping the spatial distribution of nutrients, pesticides, or contaminants in plants and soils in agronomy. It can provide insights into plant health by visualising biochemical changes during growth or stress conditions and can be used to study the composition of agricultural products such as grains or fruits to assess quality and detect adulteration (Salzer & Siesler, 2014).

1.1.3.4. Surface-Enhanced Raman.

To address the inherent weakness of standard Raman signals, Surface-Enhanced Raman Spectroscopy (SERS) utilises nanostructured metal surfaces, typically gold or silver nanoparticles, to amplify the intensity of Raman scattering by factors as high as 10^{10} to 10^{15}. This increase in sensitivity enables the detection of trace analytes down to single-molecule levels, making SERS a powerful tool in applications such as biosensing, environmental monitoring, and forensic science (Das & Agrawal, 2011).

The observed enhancement in SERS results from two main mechanisms: electromagnetic enhancement and chemical enhancement. Electromagnetic enhancement is caused by localised surface plasmon resonances (LSPRs), which occur when light interacts with the nanostructured metal surface. LSPRs are collective oscillations of free electrons on the metal surface that are excited by specific wavelengths of light. These oscillations create intense electric fields, or "hot spots," near the surface of the metal. The enhanced electric fields amplify the incident light and the scattered Raman signal, significantly increasing signal strength. The magnitude of this enhancement depends on factors such as the size, shape, and arrangement of the metallic nanostructures. Chemical enhancement, on the other hand, involves charge transfer between the analyte molecules and the metal surface. This interaction modifies the polarizability of the molecules during vibration, boosting the signal intensity. While electromagnetic enhancement is generally considered the dominant mechanism, both processes often work together to achieve remarkable amplification (Han et al., 2022).

Ag and Au nanoparticles are ideal for SERS because they exhibit LSPR when illuminated with light at specific wavelengths. On the one hand, silver nanoparticles are known for their powerful plasmonic effects, making them highly effective at producing intense hot spots that amplify Raman signals by several orders of magnitude. However, silver is prone to oxidation, which can limit its performance over time or in certain environments. On the other hand, while slightly less effective than silver in terms of plasmonic strength, gold nanoparticles offer superior chemical stability and resistance to oxidation. Silver nanoparticles are often chosen for applications requiring maximum sensitivity, such as detecting trace pollutants or hazardous chemicals. Gold nanoparticles, on the other hand, are preferred for biological and biomedical applications where long-term stability and biocompatibility are essential. Additionally, both silver and gold nanoparticles can be engineered with precise control over their size, shape, and aggregation state to optimise their plasmonic properties for specific excitation wavelengths or target analytes (M. Fan et al., 2012; Langer et al., 2020).

Thus, Raman spectroscopy offers a range of techniques suited to different applications. Spontaneous Raman spectroscopy is ideal for general material characterisation and obtaining molecular fingerprints. Raman microscopy is best for analysing small or specific regions within heterogeneous samples, and Raman imaging excels in creating spatially resolved chemical maps. Besides, SERS should be used when detecting trace analytes or overcoming weak Raman signals, particularly in highly sensitive applications.

1.2.4 Calibration: Ensuring Data Accuracy

Calibration is a critical aspect of Raman spectroscopy, ensuring the accuracy, reliability, and reproducibility of spectral data. Without proper calibration, measurements can be affected by environmental factors, instrument ageing, or misalignments in the optical system, leading to errors in spectral interpretation. Calibration aligns the instrument's performance with known standards, enabling precise and consistent data collection across different instruments and conditions. There are three primary types of calibration in Raman spectroscopy: wavelength calibration, intensity calibration, and spatial calibration (Diz et al., 2022).

Wavelength calibration is fundamental to aligning spectral peaks with their correct positions on the Raman shift axis. This process assigns precise wavenumber values (in cm^{-1}) by using reference materials with well-defined Raman peaks, such as silicon (520.45 cm^{-1}) or polystyrene. This calibration typically involves measuring the spectrum of a standard material and comparing the observed peak positions with their known values. Any discrepancies are corrected by applying a polynomial function that maps the detector's positions to their corresponding wavenumbers. This ensures that each Raman peak accurately reflects the vibrational energy of the sample molecules. Wavelength calibration is essential for maintaining consistency across instruments and for enabling reliable comparisons with spectral libraries (Bocklitz et al., 2015; Jakubek & Fries, 2020).

Intensity calibration normalises the spectral intensity across different instruments or experimental conditions. Variations in detector response, laser power, or optical components can lead to discrepancies in peak intensities, which may affect quantitative analyses or relative peak comparisons. Intensity calibration uses a reference material with a known spectral intensity profile, such as NIST-certified luminescent glass standards, to correct for variations. The instrument's response is standardised by applying a correction factor based on the reference material's spectrum, ensuring that measured intensities are proportional to the actual photon counts from the sample. This type of calibration is important for quantitative applications, such as determining concentrations of analytes or comparing spectra collected on different systems (Fryling et al., 1993).

Spatial calibration aligns spectral data with spatial coordinates in imaging applications. In techniques like Raman microscopy or Raman imaging, where spatially resolved chemical maps are generated, it is crucial to ensure that each spectrum corresponds accurately to its physical location on the sample. Spatial calibration is the process of verifying and adjusting the alignment of optical systems, such as lenses, mirrors, and gratings, to eliminate distortions or offsets that may misrepresent spatial relationships within a sample. This ensures high-resolution imaging and accurate mapping of molecular distributions. Calibration standards, such as patterned samples or fluorescent markers, are often used to verify the spatial accuracy of the system and the positions of these known features are compared to their corresponding spectral data to identify any misalignments. Adjustments are then made to correct these errors, ensuring each spectrum corresponds precisely to its physical location on the sample. In Raman systems with imaging capabilities, the aim is to align the Raman data with video or optical images captured to ensure that the spatial coordinates of the Raman measurements correspond exactly to the visual image of the sample (Diz et al., 2022; Mosca et al., 2021).

Regular calibration is necessary to maintain instrument performance over time. Environmental factors like temperature fluctuations can cause slight shifts in laser wavelength or grating angles, leading to the misalignment of spectral peaks. Many Raman systems now feature built-in calibration routines that allow users to calibrate wavelength and intensity quickly and accurately with minimal manual intervention. Calibration is essential for harmonising data from different Raman systems. Since no two instruments, even those of the same model, produce identical raw spectra due to differences in design and detector response, proper calibration ensures that measurements are comparable across devices. This harmonisation is crucial for developing strong spectral libraries, conducting inter-laboratory studies, and implementing global models for industrial quality control.

2. ANN-Based Predictive Modeling for Fruit Quality and Safety

The decision to conduct a review of artificial neural networks (ANNs) in agriculture forms a foundational and strategic component of this thesis. Artificial neural networks have emerged as indispensable tools for extracting meaningful insights from complex, high-dimensional datasets as the agri-food industry undergoes a profound transformation driven by advances in sensing technologies and data analytics. The integration of ANNs with non-destructive analytical equipment such as hyperspectral imaging and Raman spectroscopy, as explored throughout this research, has enabled the development of powerful models for quality assessment, safety monitoring, and process optimisation across various fruit products.

This review served several important purposes. First, it offered the theoretical and practical context needed for selecting, adapting, and implementing various neural network architectures customised to the unique challenges of agricultural data. Understanding the underlying principles, strengths, and limitations of each technique was essential not only for optimising model performance but also for ensuring that the chosen approaches were scientifically valid and aligned with the specific aims of each case study. This thesis reviews recent advances and current trends in the application of ANNs in agriculture, positioning its methodological contributions within the broader context of ongoing research and development. This approach facilitates the identification of best practices and state-of-the-art solutions while also highlighting gaps and opportunities for further innovation.

The review process served as an important learning opportunity. This understanding informed experimental design, helped to choose relevant performance metrics, and encouraged the adoption of interpretability techniques. As the field of agricultural data science continues to grow, the importance of methodological transparency and reproducibility is crucial.

Artificial Neural Networks in Agriculture, the core of artificial intelligence: What, When, and Why

Salvador Castillo-Girones[1], Sandra Munera[2], Marcelino Martínez-Sober[3], José Blasco[1], Sergio Cubero[1], Juan Gómez-Sanchis[3*]

[1]Centro de Agroingeniería, Instituto Valenciano de Investigaciones Agrarias (IVIA), CV-315, km 10,7,

46113, Moncada (Valencia) Spain; castillo_salgirb@gva.es, blasco_josiva@gva.es, cubero_ser@gva.es

[2]Departamento de Ingeniería Gráfica. Universitat Politècnica de València. Camino de Vera, s/n, 46022 Valencia, Spain. sanmupi@upv.es

[3]IDAL, Departamento de Ingeniería Electrónica, Universidad de Valencia, Av. de la Universidad, S/N, 46100 Burjassot (Valencia) Spain. Juan.Gomez-Sanchis@uv.es, marcelino.martinez@uv.es

Computers and Electronics in Agriculture, 230, 109938.
https://doi.org/10.1016/J.COMPAG.2025.109938

Abstract

Artificial Neural Networks (ANNs) based models have emerged as a powerful tool for solving complex nonlinear problems in agriculture. These models simulate the human nervous system's structure, allowing them to learn hierarchical features from the data and solve nonlinear problems efficiently. Despite requiring a large amount of training data, ANNs with shallow architectures demonstrate superior performance in extracting relevant features and establishing accurate models, instilling confidence in their effectiveness compared to conventional machine learning methods. The versatility of ANNs enables their application in various agricultural domains, including precision agriculture, species classification, phenotyping, and food quality and safety assessment. ANNs combined with image analysis have proven valuable in disease detection, plant phenotyping, and fruit quality evaluation. The use of deep learning in agriculture has experienced exponential growth, as evident from the increasing number of publications in recent years. This article overviews recent advancements in applying ANNs in agriculture. It delves into the fundamental principles behind various types of agricultural data and ANN models, discussing their benefits and challenges. The article offers valuable insights into the proper use and functioning of each neural network, data processing for improved model outcomes, and the diverse applications of ANNs in the agricultural sector. It aims to equip readers with practical information on data utilisation, model selection based on data type, functionality, and current research applications.

Keywords: Deep Learning; Neural Networks; Applications; Agricultural Products; Quality

2.1. Introduction

Artificial Intelligence (AI) has transformed numerous industries, providing advanced solutions to complex problems through various techniques and algorithms. Among these techniques, artificial neural networks stand out for their ability to learn and improve, inspired by the structure and functioning of the human brain. The idea behind ANN-based models is to create algorithms that simulate the structure of the human nervous system, which is made up of connected neurons (Nayak et al., 2020). These systems use nodes or artificial neurons interconnected in a layered structure that resembles the human brain, obtaining non-linear relationships from the data more efficiently than other methods (Arul, 2021) to solve complex problems in artificial intelligence. Despite the increased interest in AI by academia, industry, and public institutions, there is no universally accepted definition of AI. Various perspectives have attempted to define AI in terms of human intelligence. Nevertheless, several fundamental elements consistently emerge as essential components of AI, encompassing environmental perception, acquisition and comprehension of input as data, decision-making, execution of actions, autonomous task performance, and attaining predefined goals (S et al., 2020). Among the AI definitions highlights the proposed by the High-Level Expert Group on Artificial Intelligence (HLEG) that states, "AI systems are systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best actions to take to achieve the given goal" (HLEG, 2019).

Deep learning (DL) models are highlighted among the supervised AI-based algorithms (Chollet, 2021). DL is a machine learning subfield based on learning features related to different abstraction levels in a particular problem (Hinton et al., 2006). The DL paradigm uses artificial neural network (ANN)--based models to solve problems. It is expected that DL is associated with ANNs with many hidden layers, but it is not the number of layers that define DL. The underlying concept of deep learning is the hierarchical processing of data (Bengio, 2009), where ANNs are used to obtain increasingly meaningful representations of the data through layered learning.

AI is being applied across many industrial and social sectors and activities. Among them, agriculture presents enormous potential for the development of AI systems. Agriculture has been driven by empirical knowledge passed down from generation to generation. This traditional character, deeply rooted in agricultural practices, has evolved over recent decades thanks to technological advancements following the Industrial Revolution. The 20th century introduced modern machinery and efficient irrigation systems that boosted productivity through fertilizers and pesticides. However, the current state of agricultural technology reveals that while some crops have embraced digitised and sensor-based solutions, conventional techniques still dominate others. In this sense, the emergence of AI represents a new paradigm that redefines how we interpret data, make decisions, and manage resources. Nevertheless, this transition has obstacles for a traditional and relatively less skilled sector compared to industries such as healthcare or manufacturing (Sharma et al., 2021). In industrial environments, processes tend to be predictable.

On the other hand, agriculture operates in a context of high uncertainty, where decisions must be made based on uncontrollable and diverse variables. In this regard, the agricultural sector requires a balance between intuition, grounded in experience with crop management, and data-driven analysis by AI systems that are crucial to ensure social, economic and environmental sustainability in a changing and globalised world. The adaptability of ANNs is particularly relevant to agriculture because of their ability to process large volumes of heterogeneous data, model non-linear relationships, and adapt to dynamic conditions. Agriculture generates diverse data types, ranging from soil composition and weather patterns to satellite imagery and crop health statistics that are often complex, high-dimensional, and noisy. ANNs excel at extracting meaningful insights from such data, making them indispensable for solving agricultural problems.

The relevance of ANNs to agriculture is further underscored by the wide range of applications they enable. Among different ANN applications, yield prediction (Gutiérrez et al., 2019), soil management, precision spraying, crop identification (Xiao et al., 2018), plant phenology (Q. Yang et al., 2020), plant phenotyping (P. Song et al., 2021), or early pest and disease detection (Polder et al., 2019). In postharvest, these technologies are applied to assess fruit quality (Jahanbakhshi et al., 2020) or prevent fraud (Osako et al., 2020) as principal topics.

Depending on the application, different ANN-based models can be used. Among them, based on the scientific literature, the most used ANNs for agricultural applications are Multilayer Perceptron (MLP), Recurrent Neural Networks (RNN), Radial Basis Function Networks (RBFN), Autoencoders (AE), Deep Belief Networks (DBN), Convolutional Neural Networks (CNN), Generative Adversarial Networks (GAN) and Capsule networks (CN) (Elgendy, 2020). However, in many cases, these techniques are not used correctly. Among the main problems, the incorrect use of nomenclature stands out. Sometimes, the definition of ANN is overused. For example, it is common to assimilate an MLP or other specific techniques to an ANN when an MLP is a particular type of ANN. Another example is the interpretation of a CNN as an evolution of ANNs when it is just another family within ANNs. Currently, public libraries offer various ANN-based techniques that are relatively simple to use. This often leads to the libraries using the default hyperparameters as black boxes fed with an Xtrain and returning a model without adjusting them to the type of data or the problem being addressed. However, for correct and optimal use, parameters that influence the model's reliability must be adjusted, such as the network architecture, the type of activation function or the learning rate. The implication of default hyperparameters is that the model's performance may not be optimal, leading to inaccurate predictions. A typical case is the incorrect use of a continuous output activation function in classification problems.

Another typical error is the use of cross-validation (CV) techniques, resembling generically CV to the k-fold method. K-fold is a widely used CV technique in which the set of samples is divided into k subsets, and the model is trained with k-1 subsets and validated with the remaining one.

This process is repeated k times, with a different test subset selected in each iteration. Once the iterations are completed, the accuracy or the error are calculated for each of the produced models, and the final accuracy and error are obtained as an average of the k-trained models. Thus, this technique avoids overfitting by selecting the optimal set of hyperparameters in those cases with a relatively reduced set of samples. However, its use is so widespread that the k-fold may seem like the only cross-validation technique when other techniques exist to better suit the problem under study (Seraj et al., 2023).

A common generalisation is using MAE or RMSE as measures of the performance of regression models and the accuracy in classification problems. Their use is correct, but only give one metric. Moreover, in binary problems, metrics such as accuracy, sensitivity, specificity, recall, or F1-score are expected to be found, which provides insights into the model's performance. However, it may lack robustness when dealing with imbalanced datasets and may not be suitable for specific situations where one class holds greater importance. For example, in a task involving the detection of cancer patients, it is preferable to classify a healthy patient as having cancer than to misclassify an ill patient as healthy. The potential consequence of not considering the class imbalance issue is that the model may be biased towards the majority class, leading to poor performance in detecting the minority class. In this case, it would be advisable to use other metrics, such as the area under the ROC curve. It would be advisable to perform an exploratory analysis of the data before feeding it into the algorithm to check for imbalances and, if necessary, employ techniques to correct biases, such as over- or under-sampling techniques or adjusting model parameters to adjust the class weights (Naidu et al., 2023).

Furthermore, little attention is paid to feature engineering. One of the significant advantages of neural models is their ability to extract complex relationships in data. ANNs extract the relationships between the data to produce an output regardless of the input data. While this is true, it is not entirely accurate. To obtain the best results, it is essential to perform sound feature engineering, either extracting the most important features, retaining the most information or selecting those that allow the problem to be modelled with greater simplicity. Finally, a common mistake might be using complex techniques without attempting simpler ones first. However, the problem can be solved using more straightforward classical methods that are easier to interpret. For example, if a problem can be solved with a linear model, it is unnecessary to use a neural one. Furthermore, by observing the parameters of the linear model, some information about the data could be obtained, such as which variables are most important.

This article presents an overview of recent advancements in applying ANNs in agriculture. It explores the fundamental principles behind various types of agricultural data and ANN models, discussing their benefits and challenges. Addressing these topics comprehensively provides a solid foundation for understanding why neural networks are particularly suited for this sector compared to others.

2.2. Type of data

Artificial detection in agriculture uses sensors to collect and analyse data about the agricultural ecosystem. It is the basis for generating knowledge and creating tools to help farmers make informed decisions about crop management, such as determining when to plant, water, fertilise or harvest. Various types of sensors are used in agriculture. Some of them provide one-dimensional (1D) signals, such as weather sensors to monitor temperature, humidity, rainfall, wind speed, soil properties like salinity, acidity, moisture content, nutrient concentration, spectra information, etcetera, and others that capture n-dimensional (nD) data such as those that capture images for remote or proximal sensing (colour, multi or hyperspectral, thermal, time of flight, stereo cameras).

The data these sensors capture must be analysed efficiently to obtain information helpful for farmers. The field is an open environment that is constantly evolving, where meteorological events can rapidly alter the data acquisition conditions. The characteristics of these data depend on multiple factors such as the crop, variety, management, soil, or weather, among others. In addition, these data are sometimes derived from heterogeneous sources, which must be assembled, processed, and converted into useful information for the farmer. Therefore, data analysis techniques must effectively combine and process large amounts of data to quickly adapt to changing conditions and make predictions capable of improving operational decisions. The ANNs are presented as valuable technologies for analysing the problems associated with the field due to their flexibility and adaptability. However, the data structure and the result to be obtained must be considered when using one or another type of ANN.

The election of a particular type or family must be based on the intrinsic structure of the data. Each ANN type uses different input data. Most kinds of ANN can be fed with various data structures and provide good results, but, in our opinion, some algorithms perform better with 1D data due to their internal structure. In contrast, others are specially developed for n-dimensional (nD) data. Hence, only choosing the most appropriate algorithm with the optimal hyperparameters can lead to the best result. For example, in problems in which the information is encoded in a 1D array, the appropriate ANNs could be based on MLP, DBN or RBFN since they have a structure in which, generally, the input neurons of the network correspond to a predictor variable or attribute, sampled at a particular time.

On the other hand, for nD problems (i.e. multichannel images), the most appropriate ANN architectures could be based on CNN, CN or GAN since they have multidimensional input, processing or output layers in their structure. Examples of such multidimensional structural units are flattening input layers aimed at encoding an image as a vector or convolution unit that extracts features from a region of an image. Due to their typology, other architectures, such as AEs or RNNs, are suitable for different data types. However, RNNs have recurring layers that make them appropriate for handling time series of data.

Examples of ANNs that process signals or 1D data are widely found in the literature. Spectrometric data was used together with a DBN by Lu et al. (2018) to determine the rice germination, Xiong et al. (2021) used an RBFN to analyse potassium content in fresh lettuces using an RBFN, while Pourdarbani et al. (2020) estimated apple ripening training an MLP. RNNs are particularly appropriate for processing temporal data, such as climate data. Ghose et al.(2018) forecast groundwater levels, and Le et al. (2017) drought projections using this type of recursive ANN.

Images are one of the primary data types in agriculture. They can have two dimensions if they are grayscale images or three dimensions data if they are colour, multispectral or hyperspectral images. Multispectral images are composed of several electromagnetic spectrum bands. In contrast, Colour images are a specific type of multispectral images which consist of 3 channels corresponding to the red, green and blue (RGB) electromagnetic spectrum bands. Hyperspectral images comprise several, even hundreds, spectral bands in the electromagnetic spectrum. Some examples are using RGB images and a CNN used by Dong et al. (2021) for strawberry disease identification. Also, RGB images and depth information were used by Yang et al.(2020) for citrus detection and by Zhang et al. (2018) for mango detection, in both cases using a CNN. Hyperspectral images comprise monochromatic images captured in consecutive bands, providing spectral and spatial information. Examples of the use of hyperspectral images processed through ANNs are given by Park et al. (2023) for blueberry firmness estimation through a CNN, by Fan et al. (2019) for rice infestation detection by applying an AE, by Agarwal et al. (2020) for detecting Trigoderma Granarium in grains using CNs, or by Wang et al., (2019) for tomato spotted Wilt virus detection applying GANs. Besides, AEs and GANS, due to their structure, are appropriate for generating new images and data augmentation. For instance, Abbas et al. (2021) employed GANs to create new samples as a preprocessing step before training a CNN to classify tomato leaf diseases. Furthermore, AEs have been used for feature extraction in different domains. Xin et al. (2020) utilised AEs to extract meaningful features for predicting cadmium residues in lettuce leaves. Similarly, Huang et al. (2021) harnessed AEs for feature extraction to predict apple-soluble solid content.

2.3. Data preprocessing

ANNs can extract (in the hidden layers of the net) important features by learning from the input data. This allows, especially in the case of CNNs, to work with raw data or images as input without preprocessing or a feature engineering stage. Nevertheless, this ability has a drawback: the number of samples to train these algorithms is very high, so the training time is also very long compared to other machine learning methods. Thus, pre-processing steps facilitate the training effort and performance of any machine learning method (Chollet, 2021), which is crucial for other traditional Machine Learning methods to obtain reliable results (Rinnan et al., 2009).

Various preprocessing techniques can be utilised depending on the nature of the data and the type of correction required. Several preprocessing steps can be applied with 1D data, such as those captured by sensors providing signals in agriculture. When analysing data collected from various sources, machine learning relies on data normalisation to account for the differing units of measurement. Data normalisation aims to standardise the values in numerical columns to a standard scale, thus enhancing the stability and effectiveness of machine learning models. This is important because data measured in different ranges can disproportionately impact the model's results (Aurélien Géron, 2019; Müller & Guido, 2016). In some data coming from sensor signals, such as spectral information, scatter correction is required since the interaction of the light with the tissue disperses the light in many directions, causing undesired variations in spectra (Dawson & Acton, 2018). In these cases, multiplicative scatter correction (MSC) and standard average variance (SNV) are the common preprocessing techniques. In MSC, each spectrum is regressed by least squares against the mean spectrum, whereas, in SNV, the spectra are mean-centred and divided by their standard deviation. The signal captured by sensors can be affected by several factors, such as electronic noise, mechanical vibrations, sensor ageing or power supply variations. These factors cause high-frequency peaks in the signals that are convenient to smooth, improve the signal-to-noise ratio, and make it easier to identify underlying patterns. Savitzky-Golay (SG) is one of the primary methods for smoothing the signal where successive sub-sets of adjacent data points are fitted with a low-degree polynomial using the linear least squares. Spectral derivatives are widely used to remove both additive and multiplicative effects in signals like spectra, where they are commonly used for separating peaks of overlapping bands (Fearn et al., 2009).

Preprocessing nD data requires techniques different from those used for signals. In the case of multi (including colour) or hyperspectral images, which are probably the most common nD data captured by agricultural sensors, spatial noise due to variations in the sensor or the illumination, geometric corrections, or dimensionality reduction are among the most required preprocessing.

The correction of hyperspectral images is typically done to reduce the influence of the particular sensitivity of the sensor or the illumination. The most used correction involves obtaining a relative reflectance through a dark reference (to remove the impact of electronic noise due to dark currents in the sensor) and a calibrated white reference of a known reflectance. This results in improved image quality, spectrally and spatially corrected, allowing accurate data analysis (Grahn & Geladi, 2007).

Among preprocessing steps, dimensionality reduction is aimed at removing redundant features and noisy and irrelevant data to improve learning feature accuracy and reduce the training time (Velliangiri et al., 2019). Improving the learning process and reducing the training computational requirements is critical since this step is usually complex, requires powerful computing units and requires a long time to build the models. Moreover, working with high-dimensional data can lead to overfitting and decrease model interpretability. Reducing dimensionality while retaining the relevant information can reveal hidden patterns, relationships, and underlying structures in the data, which helps to understand the intrinsic characteristics of data better. These techniques can be divided into feature selection and feature extraction (Velliangiri et al., 2019). The choice of dimensionality reduction technique must depend on the data's specific characteristics, the analysis's goals, and the computational resources available. Velasquez et al. (2024) use six dimensionality reduction methods to obtain a subset of wavelengths from hyperspectral images to detect anthracnose early in mango.

Feature selection methods aim to choose a subset of the most influential features from the original dataset concerning the target variable. It helps improve model performance, removing irrelevant features and providing interpretability by simplifying models. These methods are the election in problems aimed at obtaining a reduced number of essential variables, for instance, to select multispectral data from hyperspectral data to build faster models or to construct a camera using only the key wavelengths. Feature selection approaches can be split into filter, wrapper and embedded methods (Jović et al., 2015). Filter methods (Sánchez-Maroño et al., 2007) are used to evaluate features independently of the machine learning model. These methods use different approaches to assess the importance of each feature for prediction. Some standard techniques include correlation-based methods that select features with high correlation with the target variable; mutual information that quantifies the amount of information shared between a feature and the target variable, choosing those with high mutual information; and statistical tests, like chi-squared or ANOVA, that assess the statistical significance of a feature concerning the target variable. Wrapper methods evaluate the performance of a subset of features iteratively using a specific machine learning algorithm. Each iteration adds or removes a variable and compares the performance with the previous iteration, determining if the variable positively or negatively influences the model. Examples include forward selection, backward elimination, and recursive feature elimination (RFE). Forward Selection starts with no features and adds them individually, selecting the one that most improves model performance. Backward Elimination begins with all features and removes them one by one, excluding the one that has the most negligible impact on model performance. Finally, recursive feature elimination (RFE) iteratively trains the model and removes the least essential features until the desired performance level is achieved (El Aboudi & Benhlima, 2016). Generally, wrappers are computationally more demanding than filter methods. Embedded methods embed feature selection within the model-building process, considering the importance of the feature the model is trained (Lal et al., 2006). A popular embedded algorithm is LASSO, which uses a penalty to shrink many regression coefficients to zero, selecting the features with non-zero regression coefficients (R & Rohini, 2016).

Feature extraction methods transform the original data into a lower-dimensional representation. These methods are beneficial for reducing dimensionality and finding underlying patterns or relationships within the data by creating new features containing the most crucial information. Commonly used feature extraction methods are Partial Least Squares (PLS), Linear Discriminant Analysis (LDA), Principal Components Analysis (PCA), autoencoders, and manifold learning (Dai et al., 2020). PLS and LDA are used for supervised learning, whereas PCA, autoencoders, and manifold learning are employed as unsupervised feature extraction methods (Guyon et al., 2006). PLS is used for regression (PLS-R) and classification tasks (PLS-DA) and finds latent variables (linear combinations of the original features) that maximise the covariance between the features and the target variable (Cohen et al., 2013). LDA is a supervised learning algorithm used for classification. It finds linear combinations of features that maximise between-class variance and minimise within-class variance to improve class separation (Haenlein & Kaplan, 2004). PCA identifies principal components, linear combinations of the original features, that maximise the variance. It is an unsupervised technique widely used for data compression and visualisation that reduces data dimensionality while preserving as much variance as possible (Garcia et al., 2022). Autoencoders are ANNs that can be used for non-linear dimensionality reduction by learning compact data representations (Bank et al., 2021). Manifolds help visualise and understand complex data patterns by revealing the lower-dimensional "manifold" where data points reside. It is used for dimensionality reduction, mainly when dealing with non-linear relationships, and is also commonly used for data visualisation. Some manifold learning techniques are Isomap, which builds a graph representing pairwise distances between data points (Balasubramanian & Schwartz, 2002), and Locally Linear Embedding, which preserves the local linear relationships between data points (Roweis & Saul, 2000).

Finally, generating new synthetic samples can be considered a preprocessing step, which is required in cases of limited datasets or when it is not easy to obtain original samples. Sometimes, obtaining actual samples in the field or when complex and expensive laboratory analyses are required is challenging. In many machine learning problems, especially those dealing with images, a large amount of data is needed to train effective models. In practice, obtaining sufficient labelled data is often challenging, particularly in agricultural processes, where collecting field data or conducting the costly laboratory analyses needed for reference labelling remains difficult (Elgendy et al., 2020). Traditionally, chemometric-based methods have been successfully applied in these cases with relatively few samples. Data augmentation techniques allow the size of the training data set to be artificially increased, which can improve model performance by having more samples and introducing variability in the training data. This can help prevent overfitting to the model since the model should not overly depend on specific characteristics in the original data samples. This process can be done in two main ways: generating new data samples from existing ones by applying varied and controlled transformations or using generative ANNs to create new synthetic data from the original ones. Augmentation techniques apply one or several transformations to the original data, such as image rotation, translation, or clipping (Blazhko et al., 2021). In contrast, generative techniques such as Variational Autoencoders (VAE) and Generative Adversarial Networks (GAN) (Lalitha & Latha, 2022) can create new samples based on information provided by the original data set. VAE starts with an encoder layer that transforms the input into a latent feature space. Subsequently, a decoder processes each point in the latent space, which generates a reconstruction of the original data from this point. The VAEs are trained to minimise a global loss function, which has two terms: one that measures the similarity between the original input and the reconstruction and another known as Kullback-Leibler divergence, which measures the difference between the latent distribution generated by the encoder and a standard Gaussian distribution (Huang et al., 2021). GANs start with random samples, generate new data that resembles the actual data, evaluate their similarity, and learn to distinguish them. This process is repeated iteratively, seeking to generate data that is increasingly similar to the actual data until they reach a threshold of similarity (A. Abbas et al., 2021).

2.4. Artificial Neural Networks

Neural networks consist of layers that contain interconnected neurons. A layer of an ANN is a set of neurons whose inputs come from a previous layer (or from the input data in the case of the first layer) and whose outputs are the input of a later layer. ANNs are formed of at least three layers, as shown in Figure 1: an input layer which receives the input data, one or more hidden layers that serve as the network's learning and abstraction mechanism, extracting meaningful features, and an output layer that makes the prediction. Depending on the complexity and data structure of the problem, the user determines the number of layers, the number of neurons in each layer, and the type of layers based on the data and the expected outcome. ANNs with more layers result in higher-level and more complex features.

There are three main types of layers: fully connected, recurrent, and convolutional. Fully connected layers feature dense structures where each neuron connects to every neuron in the next layer. They are used to learn complex patterns from the data. Recurrent layers of neurons are designed for sequential or time series data, incorporating information from past data to influence the current input and output. Convolutional layers are specialised for image processing, capturing spatial hierarchies by analysing information within small, overlapping regions (Elgendy et al., 2020).

ANNs use forward propagation to predict data, as shown in Figure 1. This means that data entering the input layer is processed through the entire network, passing from one layer and neuron to the next until reaching the output layer, where the prediction is made. The input of each neuron (node) is weighed, and then a bias value may be added to compute the output of a neuron, allowing the network to adapt and learn patterns. The value of the following neuron is obtained by summing the output of neurons connected to it. This value is then passed to a non-linear activation function (Chollet, 2021). Activation functions determine whether the neuron should be activated based on the input, thus introducing non-linearity into the neural network. This nonlinearity is essential for the network to learn and capture complex patterns within the data effectively. Several activation functions are commonly used, including the sigmoid function, hyperbolic tangent (Tanh), rectified linear activation (ReLU), and leaky ReLU function. The sigmoid function produces outputs between 0 and 1, while Tanh produces outputs between -1 and 1. ReLU is the most used and outputs a positive value for positive inputs and zero for non-positive values. Like ReLU, Leaky ReLU allows for negative values, providing additional flexibility (Arul, 2021).

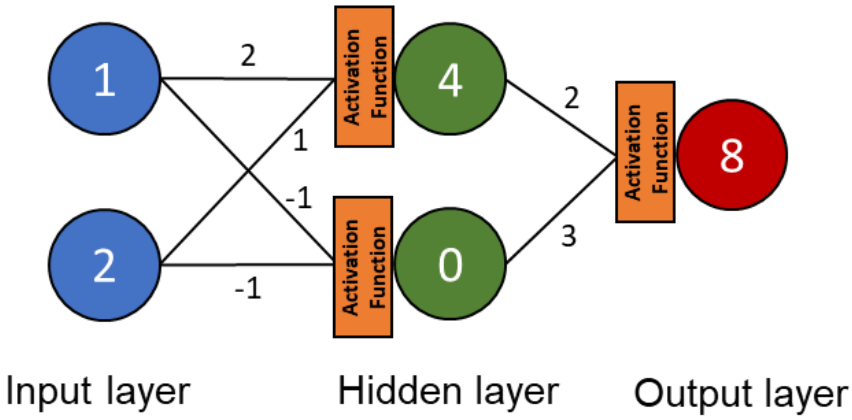


Figure 2. Forward propagation. Example using the ReLU activation function. Mumbers are model weights.

During the training process, the weights of the nodes in the network are adjusted through an iterative process called backpropagation until the error between the expected and obtained values is minimal. The error is measured through a loss or objective function. There are several types of loss functions, which are chosen based on the type of data being predicted. For instance, Mean Squared Error (MSE) is commonly used for regression problems, Binary Crossentropy is used for two-class classification, and Categorical Crossentropy is used for multiclass classification. Gradient descent is used to optimise this process, where the loss function guides the optimisation process by quantifying the model's performance. During each iteration, the gradient/slope of the loss is calculated. The aim is to find the point where the slope of the loss function is 0, the bottom of the function, and where the error between the desired value and the obtained value is 0. The slope of each point in the loss function is calculated with the chain rule (the cost function derivative). If the slope is positive, it gets multiplied by the learning rate, leading to an update in the weight, as shown in Equation 1.

$$w' = w - \lambda * \nabla L(w)$$

Equation 2. Neuron weight update during model training.

Where w is the weight of the neuron, w' is the new weight of the neuron, λ is the learning rate (defined by the user), and ∇L(w) is the gradient of the loss with respect to the weights. The user defines the learning rate and the step size taken during optimisation. It determines how much the model's parameters should be adjusted based on the calculated gradient. A higher learning rate means larger steps but could overshoot the optimal values, while a lower learning rate takes smaller steps but might slow down convergence. Finding the right balance is crucial for effective model training (Elgendy et al., 2020).

ANNs need to be trained. Training is the process of teaching a neural network to perform a task. In principle, neural networks learn by processing several large sets of labelled or unlabeled data. The training strategy depends on publicly labelled data availability, the task's uniqueness, and the task's similarity. There are three ways to train ANNs: from scratch, transfer learning, and feature extraction. Training a neural network from scratch implies starting the training process without prior information. This approach can be advantageous when working on a task without a pre-existing model. It is necessary for scenarios where the data is distinctive or the problem is new. Using this method, the model can learn task-specific patterns customised to the specific dataset. Transfer learning is a technique that uses a pre-trained model on a related task and applies this knowledge to a similar task. It proves helpful when there is a small number of labelled data for the target task. Transfer learning saves time and resources by leveraging the information gained from a related task. It is effective when the source and target tasks share standard features. Finally, feature extraction uses a pre-trained model to extract essential features from data. It is convenient when the focus is on the high-level representation of data rather than the detailed information. This method allows for using a pre-trained model's ability to capture complex features of the problem without retraining the whole model (Chollet, 2021).

2.4.1. Multi-Layer Perceptron (MLP)

An MLP is one of the most basic forms of ANN. It comprises three or more layers, including an input layer, at least one fully connected hidden layer and an output layer. Each layer has different nodes or neurons; every node is connected to another node in the next layer (C. Zhu, 2021). MLPs are commonly used in agrifood applications due to their simplicity yet effectiveness. For instance, they have been implemented to predict kiwifruit firmness (Torkashvand et al., 2017), sugarcane yield prediction (P. Wang et al., 2021) or in nutritional status assessment of olive crops (Noguera et al., 2021). Velasquez et al. (2024) used multiple methods, including MLP, to classify asymptomatic mangoes as either healthy or diseased (infected by anthracnose). Among the techniques evaluated, MLP demonstrated superior performance in accurately distinguishing between the two categories.

2.4.2. Deep Belief Networks (DBNs)

Geoffrey Hinton proposed DBNs in 2006, which are commonly used for image generation and recognition. They combine the power of deep architectures and probabilistic graphical models in tasks with limited labelled data because they excel in unsupervised learning, automatically capturing complex patterns in data through probabilistic modelling. In DBNs, hidden layers consist of stacks of Restricted Boltzmann Machines (RBMs) that transform input data into a latent space. RBMs have fully connected connections between their layers, but nodes within each RBM are conditionally independent, as shown in Figure 2. This allows RBMs to learn complex patterns in the data hierarchy efficiently (Hinton et al., 2006).

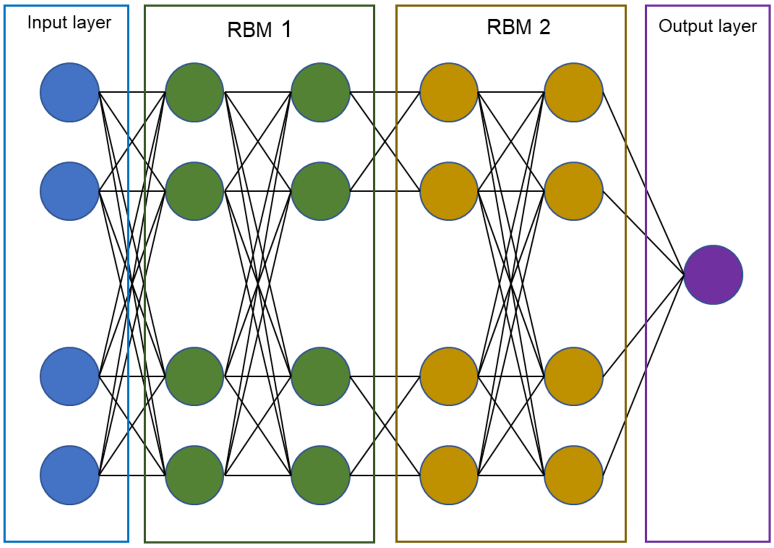


Figure 2. Deep Belief Network

RBMs are suitable for tasks like image and speech recognition, image generation and feature extraction due to their hierarchical feature learning and generative capabilities (Hinton et al., 2006). DBNs are not commonly used in the agriculture and food industries. However, some studies have applied them to eliminate variable correlation before using SVM to predict the transpiration rate of strawberry leaves (Shuaishuai et al., 2018), for feature extraction for pesticide residue measurement (Wu et al., 2019) or for detecting rice germination rates (W. Lu et al., 2018). Besides, some studies have shown that DBNs can produce better results than traditional artificial neural networks (ANNs), such as MLPs, in modelling the spatiotemporal distribution of soil moisture (X. Song et al., 2016).

2.4.3. Recurrent Neural Networks (RNNs)

Recurrent Neural Networks (RNNs) are specifically designed for tasks that involve sequences and time dependencies. They can retain past information, adjust to sequences of varying lengths, and effectively capture contextual information. This makes them ideal for natural language processing and predicting time series applications. On the other hand, other ANNs like MLPs are structured to approximate functions and face difficulties when modelling dynamic systems, as well as systems that involve temporal components with sequential data (Hochreiter & Schmidhuber, 1997)

They work by sequentially processing data and maintaining the memory of past information through the hidden layers. RNNs consist of three main layers: the Input layer, which receives data; the hidden layers, which capture information from their input and retain the memory of past layers; and the output layer, which produces the final output based on the data processed by the hidden layers as shown in Figure 3 (C. Wang et al., 2021)

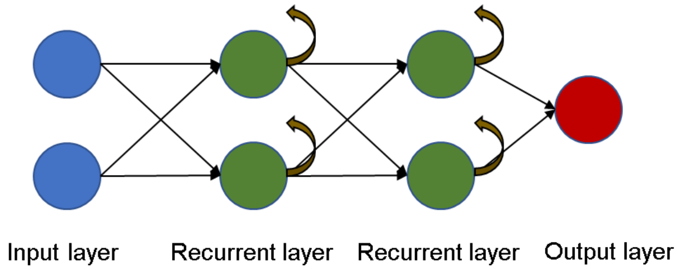


Figure 3. Recurrent Neural Network

However, in basic RNNs, repeated multiplication of weights with a value less than one can cause values to shrink exponentially, leading to very small weights over many time steps. This phenomenon is known as "vanishing gradients, which pose challenges in training. Therefore, Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) architectures were developed to solve that problem and are the most common RNN models. LSTMs are designed to remember important information from past layers, making them great for understanding sequences. They use a gate to decide what to store, update, or forget in their memory. This helps the network capture and retain meaningful patterns over extended sequences (Hochreiter & Schmidhuber, 1997)

GRUs decide what information to remember or forget while processing sequences. They have a simplified system with only two gates: an update gate and a reset gate. This makes GRUs computationally less intensive compared to LSTMs (Jing et al., 2019)

Recurrent Neural Networks (RNNs) are ideal for tasks that involve sequences. RNNs can effectively remember and utilise information from previous steps, making them well-suited for applications like natural language processing and time series prediction. It was challenging to find articles related to agriculture that used RNNs. However, one such study was conducted by Ghose et al. (2018) for groundwater level forecasting, where they used different types of sequential data such as rainfall, temperature, humidity, evapotranspiration and runoff as inputs, and groundwater depth as the output. Another study by Le et al., 2017) projected droughts in California using sequential data such as soil moisture, precipitation, evapotranspiration, moisture recharge, and run-off from 1896 to 2006, which was later validated with the period from 2006 to 2015.

2.4.4. Radial Basis Function Network (RBFN)

Broomhead and Lowe (1988) introduced Radial Basis Function Networks with the primary objective of curve fitting. RBFNs offer two key advantages over MLPs: firstly, they are connected to a well-established fitting technique while still leveraging the advantages of neural networks. Secondly, RBFNs can yield good results with a straightforward model in practical applications due to their ability to blend linear dependence on variable weights and capture nonlinear relationships. RBFNs have a simple structure (Figure 4) and a fast training process with tolerance to input noise (Hashemi Fath et al., 2020). A radial basis function uses a Gaussian function (Fernández-Ruiz et al., 2010), transforming the input into another form that feeds the network to get linear separability. This property allows the network to separate data points with a straight line or plane. This is achieved by transforming the input data into a higher-dimensional space where the data points become linearly separable (Nayak et al., 2020).

These ANNs consist of three layers: an input layer, a hidden layer known as the RBF layer, and an output layer consisting of neurons that perform linear combinations. Each RBF neuron retains a prototype (an instance of the training set). Each neuron computes the Euclidean distance between the input and its respective prototype during classification. The hidden layer relies on a radial basis function as the activation function for approximation. They are used in function approximation, classification and prediction. In a study focused on detecting powdery mildew stages in squash, an RBFN was employed for classification. The study aimed to develop a system that accurately identifies the different stages of powdery mildew in squash plants, a common fungal disease affecting squash production. The RBFN achieved accuracy exceeding 92 % across seven classes and distinguishing between diseased and healthy samples (Abdulridha et al., 2020).

Additionally, RBFNs have been used as an alternative to MLPs for regression tasks, such as predicting lycopene content in tomato juice (Fernández-Ruiz et al., 2010). Another application involving RBFNs for regression was to estimate potassium levels in fresh lettuces, demonstrating their ability to outperform widely used chemometric models like PLS regression (Xiong et al., 2021). The superiority of RBFNs over PLS was further evidenced in a study on rice germination rate by Lu et al. (W. Lu et al., 2018).

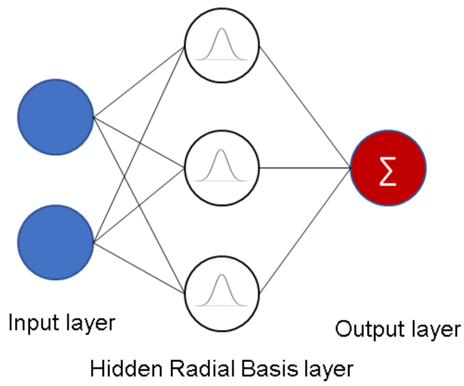


Figure 4. Radial Basis Function Network

2.4.5. Autoencoders (AEs)

Autoencoders (AEs) are a type of neural network for unsupervised learning. They can work with data of different dimensions, ranging from 1-D to n-D, but are mainly used for image data. Autoencoders have two parts: an encoder and a decoder, as shown in Figure 5. The encoder compresses input data by extracting features and creating a compact representation called a bottleneck. The decoder then reconstructs the original data from this bottleneck. Depending on the input data, fully connected layers with 1-D data or convolutional layers in the case of images are used (Bank et al., 2021).

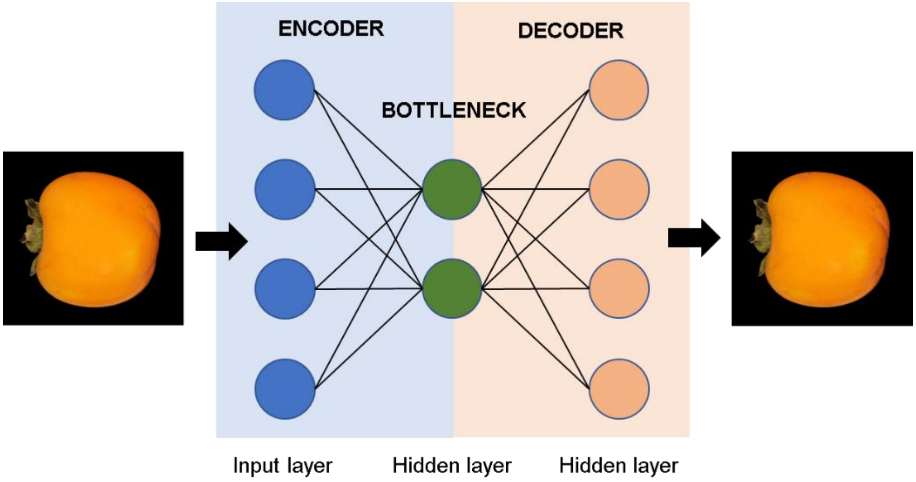


Figure 5. Autoencoder

There are three main types of autoencoders: Sparse Autoencoders (SpAEs), Stacked Autoencoders (StAEs) and Variational Autoencoders (VAEs). SpAEs learn compact representations of data while encouraging sparsity in the learned features. Sparsity means that only a few neurons are active at a time in the hidden layer, which can lead to more efficient and interpretable representations as the network focuses on capturing the most important features of the input data. StAEs involve combining multiple autoencoders, one on top of the other, so the output of one autoencoder becomes the input for the next. Each autoencoder in the stack learns to capture different data abstraction levels, allowing the network to capture complex patterns. In VAEs, the bottleneck vector is replaced by two vectors: one represents the standard deviation and the other the mean of the distribution. Apart from learning to represent data, they also learn the uncertainty in their representations, letting them generate new, diverse samples (Bank et al., 2023).

Autoencoders can be used for various tasks, such as data compression, image denoising, generating new images, or extracting features that can be utilised for classification, regression, or prediction. Many studies have employed autoencoders due to their capability of feature extraction, which can be later used for classification, such as for predicting soluble solids in apples (Huang et al., 2021), where a StAE was used for feature extraction before soluble solid prediction using an MLP. The same approach was used to extract the most essential spectral features to predict the nitrogen concentration in oilseed rape leaf (X. Yu et al., 2018). Other authors have AEs in combination with traditional machine learning methods, such as PLS or SVMs, to extract features. For instance, AEs were used to predict cadmium residues in lettuce leaves. (Xin et al., 2020).

2.4.6. Convolutional Neural Networks (CNNs)

CNNs are a specific type of ANN developed to overcome the limitations of traditional neural networks, such as MLPs, in processing and recognising visual data. The idea behind CNNs is to mimic the visual processing in the human brain and take advantage of the spatial hierarchies present in visual data. CNNs excel at capturing local patterns and hierarchical features in images, making them well-suited for tasks such as image classification, object detection, and image segmentation. However, CNNs require a large amount of data for training (O'Shea & Nash, 2015). A CNN consists of several layers, including an input layer that receives the input data, convolutional and max-pooling layers that extract and learn features from the input, fully connected layers, and an output layer where the final prediction is made. This is illustrated in Figure 6. The concept started with the Neocognitron model developed by Fukushima (1980), who laid the foundation for capturing hierarchical features in visual data using a form of convolutional layers.

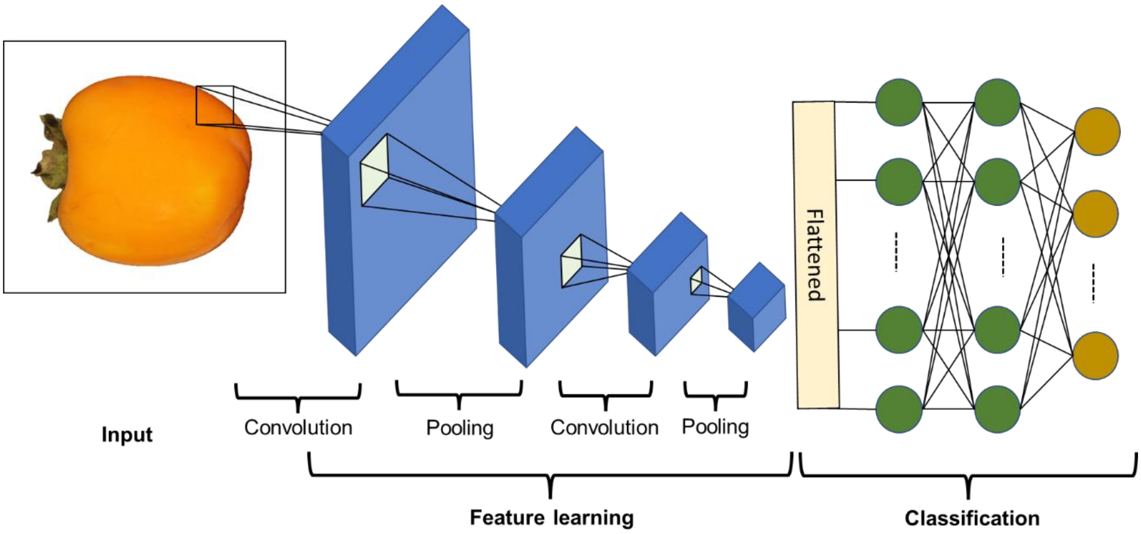


Figure 6. Convolutional Neural Network

The convolutional layer is responsible for extracting features from images and consists of a series of filters or kernels that can be trained. Each filter is applied to the raw pixel values, resulting in a product between the filter and the input pixel, as shown in Figure 7. This process produces an activation map known as a feature map. Each layer captures different features from the input images, starting from basic ones like lines and progressively becoming more complex and specific in the final layers (Elgendy et al., 2020). CNN models can employ 1D, 2D or 3D convolution filters, where 2D filters are the most used, especially in grayscale and RGB images, which are reshaped into two dimensions before adding them to a 2D CNN. In contrast, 1D is used for 1D signals and 3D filters are usually used for 3D structures such as spectral images (Liu et al., 2021).

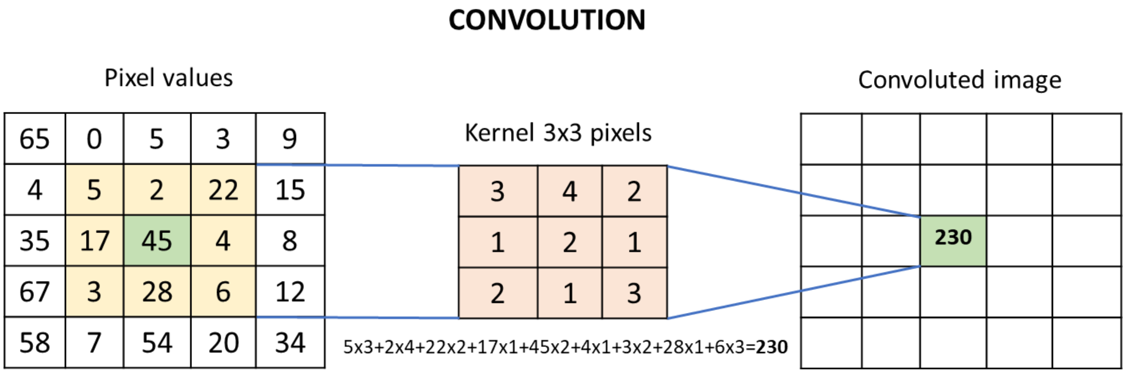


Figure 7. Example of the convolution process in a CNN

Pooling layers serve as sub-sampling layers within CNNs. They are crucial in reducing convolutional maps' size while ensuring the model's invariance to rotations and translations. A nonlinear activation function is applied, resulting in the pooling layer's output being determined by selecting the maximum activation value within each feature map of the input layer. Figure 8 illustrates this process with the ReLU activation function. By preserving the spatial relationships between features, convolution and pooling layers enable the network to learn local patterns effectively (Chollet, 2021). After feature extraction, the fully connected layers use the learned features to classify or predict. However, they require a 1D input. Therefore, flattening is performed to reshape the output of the last convolutional or pooling layer into a vector (Elgendy et al., 2020).

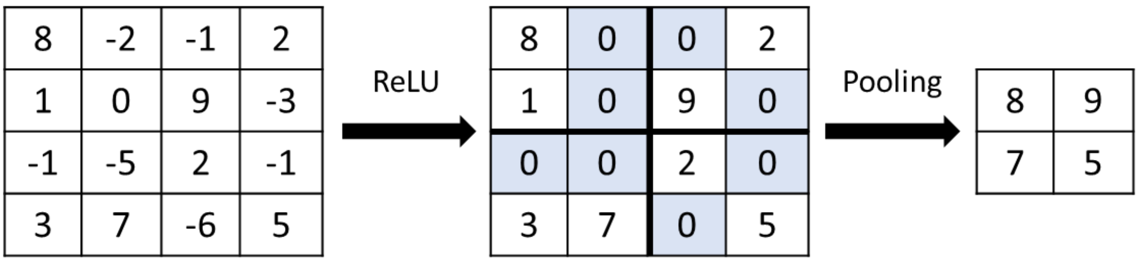


Figure 8. Example of the pooling process in a CNN

CNNs have significantly increased in computer vision and image processing applications. Different CNN architectures are employed depending on the aim. Pre-trained architectures like LeNet5, AlexNet, VGGNet, GoogLeNet, and ResNet are commonly used in classification. Region-based CNN, Fast-RCNN, Faster-RCNN, Mask R-CNN, and You Only Look Once (YOLO) are among the most popular for object detection. For image segmentation, SegNet, U-Net, and SegFast are some of the most used architectures (Santosh et al., 2022). The flexibility of CNN to learn and adapt to changing scenarios makes them extensively employed in agriculture, as in grapevine phenotyping (Milella et al., 2019), pest and disease identification as in Banana (Amara et al., 2017) or strawberry (Dong et al., 2021), plant cultivar and species identification (Javanmardi et al., 2021), fruit detection (F. Gao et al., 2022), or agricultural product quality and safety assessment (Bonora et al., 2022). CNNs proved to be better than other ANNs at processing images for detecting apple defects (S. Fan et al., 2020), subtle bruises on winter jujube (Feng et al., 2019), grading dragon fruits based on shape, size, weight, colour and diseases (Patil et al., 2021), sorting cherries based on regular and irregular shape (Momeny et al., 2020).

2.4.7. Generative models

2.4.7.1. Generative Adversarial Networks (GANS)

(Goodfellow et al. (2014) proposed Generative Adversarial Networks (GANs) to address the challenges of approximating intractable probabilistic computations and leveraging the benefits of piecewise linear units in generative models. In GANs, data is formed from random noise using traditional convolutional layers. They are composed of a generator and a discriminator model (Figure 9). The generator produces synthetic data from noise, and the discriminator distinguishes between real and synthetic data, ensuring that the generated data is close to the real one. Both the generator and discriminator are trained simultaneously and try to surpass each other iteratively.

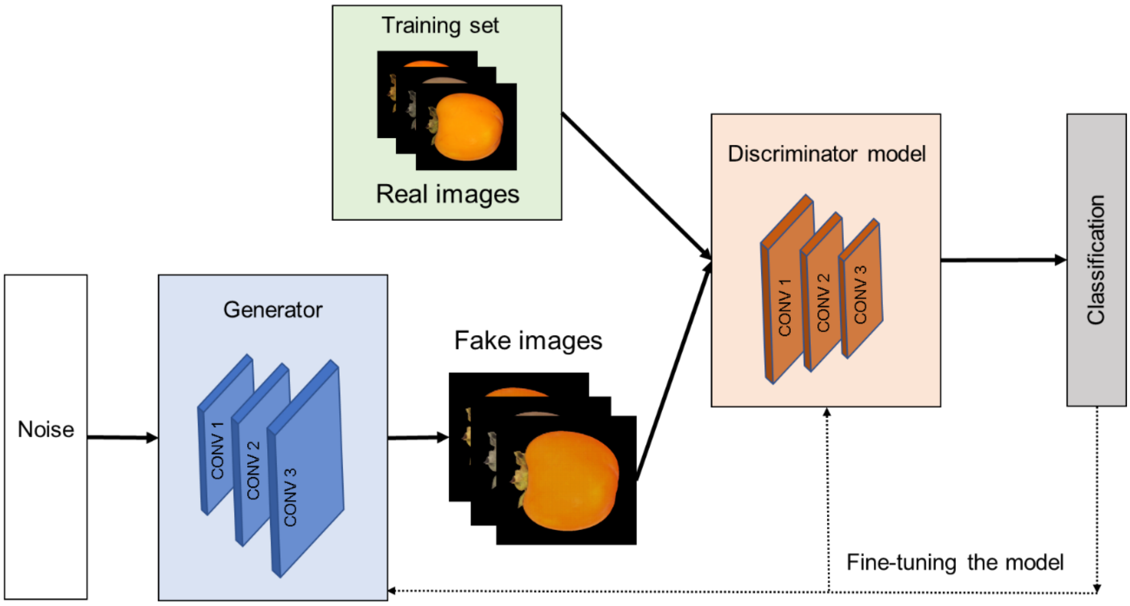


Figure 9. Generative adversarial Network

There are several different GAN implementations. CycleGAN, Conditional GANs, PGGA, or deep convolutional GANs are the most common. The idea of CycleGAN is a type of neural network that can transform images from one domain to another without paired examples for training. It uses two generators and discriminators. For instance, it can turn pictures of cars into pictures of motorbikes, even if it has never seen or learned that transformation before. It learns by trying to transform and then turning it back into a car, making sure it still looks like the original (K. Wang et al., 2017). Conditional GANs can be used to process input images. In some Conditional GANs, it is possible to provide them with sentences to generate specific outputs. PGGAs are

GANs with a progressively growing generator and a discriminator to generate from low-resolution images to high-resolution images. Finally, deep convolutional GANs are a robust variant of GANs that generate high-quality images (Navidan et al., 2021).

When working with agrifood products, obtaining the vast samples required to train ANNs or performing large amounts of expensive, destructive analysis is complex. Therefore, creating synthetic samples mimicking the real ones is a solution. GANs have been used for this purpose, as in Abbas et al. (2021), who generated new samples for data augmentation to train a CNN for tomato leave disease classification. They are also used for semantic segmentation, as shown by Chen et al., (2021), who segmented occluded tree branches using Pix2Pix GAN and compared the results with other CNNs, obtaining the best results using the GAN. Wang et al., (2019) detected the tomato spotted wilt virus using network analysis based on GANs, integrating the plant segmentation and spectrum classification.

2.4.7.2. Diffusion Models

Diffusion models were introduced in 2015 at the 32nd International Conference on Machine Learning. They are a generative model that models the gradual transformation of a simple probability distribution into the desired target distribution. A simple initial distribution (often Gaussian noise) is transformed step by step into complex data distribution (images) (Sohl-Dickstein et al., 2015). During the training process, the model learns to iteratively transform the initial distribution into samples that resemble the true data distribution. Once trained, it can generate new samples (images) by applying the learned diffusion process to a starting point, such as Gaussian noise. Diffusion models can also perform an inverse process, which means it can go from the complex data distribution back (images) to the simpler distribution. This property allows the model to be used for tasks such as denoising and inpainting. (Ho et al., 2020; Sohl-Dickstein et al., 2015). The application of these networks is relatively new in agriculture, and only one research study has been conducted on them. Chen et al., (2022) used a diffusion model for data augmentation applied to weed recognition enhancement using various pre-trained CNNs (VGG, Inception, DenseNet and ResNet).

2.4.8. Miscellaneous models

2.4.8.1. Transformers

Transformers were introduced by Vaswani et al. (2017). Its architecture was initially designed for text translation and natural language processing. However, in agriculture applications, they are mainly used to process images and are called vision transformers.

Vision transformers transform images by dividing them into smaller patches, flattening these patches into 1D vectors, and applying a self-attention mechanism. This mechanism helps the model capture local and global relationships between patches, learning spatial dependencies for practical image understanding. The process involves multiple layers and attention heads, and positional encodings convey information about the spatial arrangement of patches. If applied for image prediction, it is performed with an MLP at the end of the transformer for classification (Dosovitskiy et al., 2021). The basic structure of a vision transformer is shown in Figure 10.

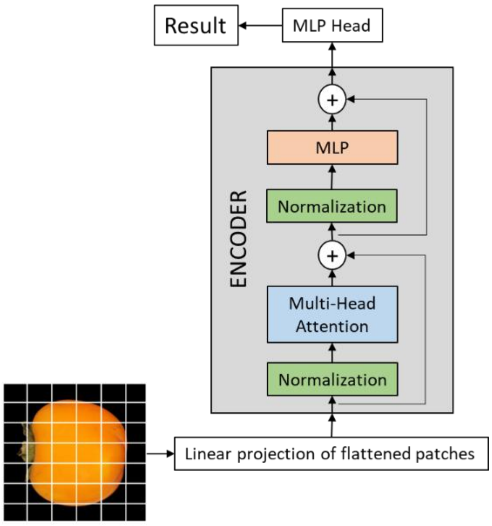


Figure 10. Vision transformer

These models are widely used in agriculture for image classification, as in L.-H. Li & Tanone. (2022), who classified several vegetables using a public dataset and obtained a 98 % accuracy. In addition, transformers have been shown to outperform other types of CNNs in image classification, such as in the classification of Cassava leaf diseases (Thai et al., 2021), diagnosing sugarcane leaf diseases (X. Li et al., 2022) and identifying plant diseases (S. Yu et al., 2023). Furthermore, transformers have been effective in segmentation tasks, such as ripeness estimation in tomatoes, where they outperformed a ResNet CNN (Shinoda et al., 2023). Transformers are handy for large-sized images, as shown in the land cover classification of multispectral LIDAR data (Y. Yu et al., 2022), crop type mapping (K. Li et al., 2022), and satellite images in deforestation monitoring (Kaselimi et al., 2022).

2.4.8.2. Capsule Neural Networks (CNs)

Capsule networks were proposed by Geoffrey Hinton in a paper called Learning to Parse Images to solve the loss of location information in neural networks (Hinton et al., 1999). These networks use capsules as the basic unit, which consists of a group of neurons that perform a specific task to combine and generate results. In the context of images, CNNs cannot identify the position of an object as they are only invariant to translation. On the contrary, capsule networks can identify translation and precisely predict the position of an object. Additionally, when reducing the size of convolutional maps, the pooling layers of CNNs lose valuable information required for prediction, while capsule networks do not have this disadvantage. Consequently, capsule networks require less data for training and offer more accurate predictions (mainly used in classification, object recognition, and segmentation) as they know the object's position. (Hinton et al., 1999; Phaye et al., 2018)

CN's structure consists of two main modules: the encoder and the decoder. In the encoder module, there are three essential layers. First is the convolution layer, which processes the input image. Next is the primary caps layer, housing capsules designed to capture fundamental features such as edges and corners. Each capsule within this layer employs convolutional kernels to enhance feature representation. The third layer, called digit caps, is specialised for recognising and encoding information about specific input digits. The decoder module's role is to reconstruct the original input image using fully connected layers that receive the encoder output as input, as shown in Figure 11. This reconstruction process aims to produce an image that closely resembles the input (Hinton et al., 1999)

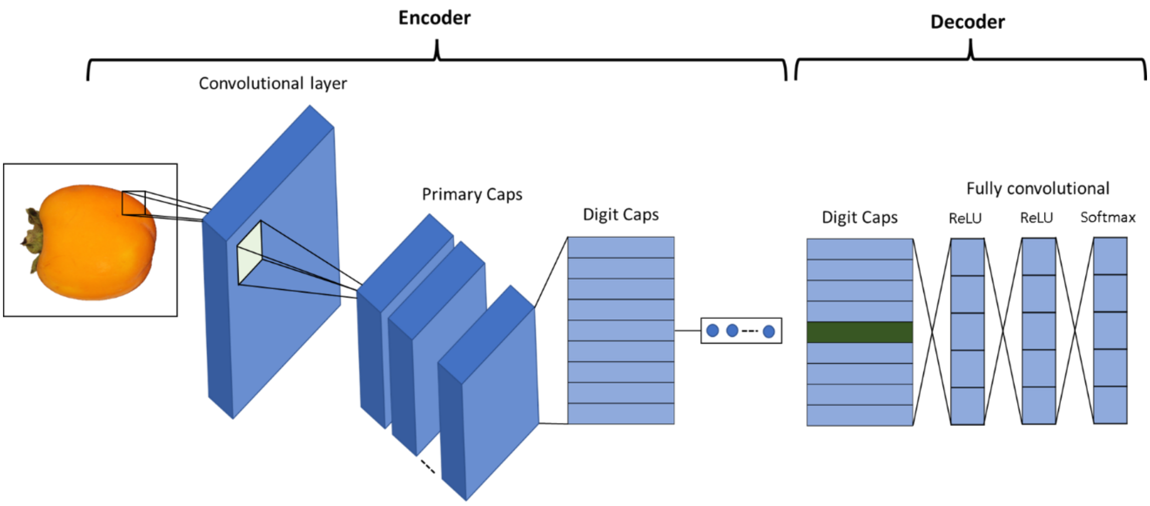


Figure 11. Capsule Neural Networks

CNs are typically used in image classification, object recognition, and segmentation. Some authors have compared capsule networks to other methods, such as CNNs, and have obtained the highest accuracies when using capsule networks, as in identifying carrot appearances (H. Zhu et al., 2021), in rice recognition (Y. Li, Qian, et al., 2019), identifying grassland coverage (Y. Sun et al., 2021) or identifying and diagnosing *Trogoderma granarium* and *Trogoderma variabile* pests in grains (M. Agarwal et al., 2020). However, the technique is not fully developed. Some experts are sceptical about its advantages, and several changes have been suggested to address its limitations (Kwabena Patrick et al., 2022). Classical capsule networks use shallow convolution, which makes them less effective in handling complex data for classification tasks. To overcome this issue, Dense and Diverse Capsule Networks have been developed (Phaye et al., 2018).

2.4.8.3. Self-Organising Maps (SOMs)

Self-organising maps (SOMs) were introduced by Teuvo Kohonen in 1982 and are designed explicitly for clustering and dimensionality reduction tasks. There are two types of SOMs: supervised and unsupervised (Kohonen, 1982).

SOMs are primarily used as an unsupervised learning technique for clustering, visualisation, and dimensionality reduction. They project high-dimensional data onto a lower-dimensional space, typically a 2D grid, while preserving the topological relationships of the data. The architecture of SOMs includes an input layer and an output layer organised as a grid of neurons. Unlike conventional neural network architectures like MLPs or CNNs, SOMs lack hidden layers and utilise competitive learning instead of error-correction-based methods. In contrast, supervised SOMs are tailored to handle labelled data for tasks such as classification or regression. They adjust the training process to incorporate label information, enabling them to perform supervised learning tasks effectively (Miljkovic, 2017)

Several authors have employed SOMs for cluster classification, demonstrating excellent results across various studies. For example, Yufeng Li et al. (2019) applied SOMs to analyse land use, irrigation patterns, and the impact of rice-wheat rotation on pond water quality. Similarly, Theanjumpol et al. (2019) used SOMs to estimate granulation levels in 'Sai Num Pung' tangerine fruit through near-infrared spectroscopy. In this study, SOMs outperformed other methods such as PLS, quadratic discriminant analysis (QDA), K-Nearest Neighbors (KNN), and LDA, achieving a 95 % accuracy in correct classification. Also, using near-infrared spectroscopy, Sringarm et al. (2024) tested LDA, QDA, and SOM for quantitative and qualitative evaluation of maltodextrin products in the industry. However, in this case, better results were obtained using PLS-DA.

2.5. Applications of ANNs in quality and safety

2.5.1. Quality assessment

Quality assessment is crucial in the agricultural industry, especially when measuring and predicting the internal and external quality of fruits and agricultural products. Ensuring that products meet consumer expectations and comply with market regulations is essential. High-quality products appeal to consumers, extend shelf life, reduce waste, and increase profitability. To meet industry standards, the agricultural sector must ensure that products comply with requirements for attributes like size, colour, firmness, sweetness, and the absence of defects such as bruises or rot. Quality standards ensure uniformity in what reaches the market, ultimately protecting consumers and sellers. For instance, Regulation (EU) 2023/2429 sets rules to standardise fruit and vegetable quality within the European Union. This law includes criteria for freshness, physical characteristics, and even permissible imperfections. Besides, internal quality (like sugar content or firmness) is just as important as external quality (appearance) since it directly influences taste, texture, and storability.

ANNs excel in quality prediction by learning complex relationships between external features (colour, size) and internal attributes (sugar content, firmness), providing a complete quality profile. 2D and 3D CNNs are commonly employed for quality assessment using image data due to their effectiveness in image analysis and feature extraction. Similarly, 1D CNNs and MLPs are utilised with spectroscopic or non-image data. Some articles utilizing artificial neural networks for quality prediction are presented in Table 1.

Maturity estimation is essential for harvesting fruit at optimal conditions since accurate estimation helps pick the product at the right time, reducing waste and improving market value, and at postharvest to ensure optimal quality, flavour, and shelf life. Due to the complexity of crop maturation and fruit ripening, the ability of ANNs to handle complex data makes them more suitable than other machine learning models. Some studies estimated maturity in the field, such as Wendel et al. (2018), who estimated the maturity of three different mango cultivars on various dates and climate conditions. A robot platform with HSI cameras was used to predict dry matter. Dry matter is closely related to mango maturity and obtained an R^2 value of 0.64 using a CNN, showing higher results than a PLS model. Alternatively, Gao et al. (2020) estimated the ripeness of strawberries using 244 hyperspectral images from the greenhouse and the laboratory. AlexNet CNN proved to be more accurate than an SVM model, with a 98 % accuracy.

In postharvest, it is important to classify the maturity of the product. This helps determine whether the product needs to be rejected, sold immediately because it is fully ripened, or if it can be stored for some time. In winter jujube, Z. Lu et al. (2021) created a grading robot to classify jujubes into three categories based on their maturity using different CNNs: Modified YOLOv3, SSD, and Faster-R-CNN. The modified YOLOv3 achieved 97% accuracy using 147 sets of images, each containing 35 jujube fruits.

Besides, estimating quality parameters such as proteins, oil, moisture, or starch content helps farmers and companies ensure that their products meet market and legal standards. Moisture content was predicted by F. Zeng et al. (2022), who used hyperspectral images and low-field magnetic resonance to predict salted sea cucumber moisture on 510 samples to detect fraud. Its proposed CNN obtained an R^2 of 0.99. Liang et al. (2024) used a portable NIR spectrometer, allowing in-site inspection on 122 samples to estimate crude protein in corn feed. He trained with PLS, SVM and a CNN: CropNet model, with which he obtained the most accurate results, an R^2 of 0.89. Other authors, such as Zhang et al. (2022), predicted other components, such as oil content. In this case, oil content in the maise kernel using his and compared different models and model combinations: PLS, SVM, GAN+PLS and GAN+SVM. The most precise results were obtained using a GAN (Nongda616) for data augmentation and an SVM for classification, obtaining an R^2 of 0.80. These results emphasise the usefulness of GANs for data augmentation in model creation.

Alternatively, in fruits, factors such as soluble solids, acidity, and firmness are essential for consumers since they play a key role in ensuring consumer acceptance based on sensory quality. In Fuji apples, Bai et al. (2019) used spectroscopy and predicted soluble solids with an MLP, PLS, and an RF model, obtaining an R^2 of 0.99 with the MLP. 208 samples from three geographical regions were used, allowing the model to eliminate the effects of geographical origin. In Kyoho grape, Xu et al. (2022) used 240 HSI and PLS, SVM, AE+PLS, and AE+SVM models. The AE+SVM model obtained an R^2 of 0.92.

Furthermore, classifying agricultural products based on internal and external defects is essential for ensuring product quality. Defects like blemishes or discolouration can make a product visually unappealing, leading to rejection by consumers or markets. However, defects such as mechanical bruising, internal rotting, or tissue damage are often hidden and only become visible later, potentially affecting the fruit's shelf life and quality. Consumers quickly reject fruits and vegetables with external imperfections, even if the internal quality remains intact. ANNs can identify subtle defects and bruises that may be overlooked by human inspectors or other machine-learning models, leading to more accurate grading and sorting.

Mechanical bruises, for example, are often invisible in the early stages but can cause internal tissue damage, which may eventually lead to rotting. As the internal damage worsens, it can spread to the outer layers, becoming an external defect. When this happens, the product may no longer be suitable for sale, leading to waste. The key to preventing this is early detection. HSI can be essential for this purpose, as demonstrated by Castillo-Girones et al. (2024), in the case of detecting three different levels of invisible bruising in plums. Three different CNN models were used: ResNet50, HSCNN, and a customised CNN. ResNet was the most reliable, with an overall F1 = 90 %. J. Lu et al. (2024) classified the fruit with different defects, including oil spots, mechanical damage, dry scars and black spots. A customised model based on YOLOv5 achieved a 99% average precision. The model was trained using 330 images of fruits with standard surfaces and surfaces with defects. In addition, different CNNs were used, such as SqueezeNet, ShuffleNet, AlexNet, GoogLeNet, MobileNetV2, ResNet50, and VGG1, to detect and sort surface defects in soybean seeds into six classes. Among these models, MobileNetV2 performed the best, achieving 98% accuracy and processing 222 seeds per minute, making it the fastest option. Similarly, Cao et al. (2021) automatically graded zizania fruit based on its quality (high and defective quality based on size and defects) using different CNNs: LightNet, AlexNet, VGG11, ResNet18, MobileNet, MobileNetV2 and ShuffleNet V2. LightNet, a light model, was the best and fastest, with 99 % accuracy. The dataset of 4,900 RGB images used in this experiment was publicly released, which can aid other researchers and instil confidence in their work.

Unlike HSI, which only penetrates a few millimetres (Sun., 2010), X-ray images can provide a deeper look inside the fruit, making them more effective for identifying deeper internal issues. For example, in the study of mango seed spoilage locations, Ansah et al. (2023) analysed 98 X-ray images of mangoes using a VGG16 CNN model, achieving an accuracy of 97%. A Grad-CAM was performed to identify the regions in the images where the model was directing its attention, thus confirming its effectiveness and focus on the seeds. Similarly, in detecting pear disorders, Tempelaere et al. (2023) analysed 80 X-ray images to find internal disorders and segment them using a U-Net, obtaining an IOU of 0.88. In avocado, Matsui et al. (2023) detected internal fruit rot with a 98 % detection accuracy and 3.15 RMSE for segmentation using 286 images.

The quality of agricultural products is highly influenced by crop management. Farmers must assess the nutritional status of plants at the field level to improve crop health, increase yield and product quality, and significantly decrease chemical contamination in the soil. Non-destructive systems, along with ANNs, are becoming a promising solution to the challenge of avoiding time-consuming and costly ionomic analyses. For example, Noguera et al. (2021) in olive leaves, N, P, and K were predicted using multispectral UAV images (5 wavelengths) in two orchards covering 150.7 ha and 6177 images. MLP, PLS, SVM and Gaussian process regression were used, and MLP showed the best results. Similarly, Yu et al. (2018) predicted N content in oilseed rape leaves using hyperspectral images on 192 samples from plants with four different fertilisation levels. A stack AE was used to extract the most important features and an MLP to predict, achieving an R^2 of 0.90. Xiong et al. (2021) used transmission spectroscopy in 211 lettuce leaves from 2 different varieties for K prediction. RBFN and PLS models were trained, and the best results were obtained using RBFN with an R^2 of 0.86.

2.5.2. Toxic compounds and residues detection

Food safety is vital for consumers, who increasingly demand products free of pesticides. Additionally, toxic compounds in agricultural products at postharvest are prohibited due to their health risks, which can lead authorities to withdraw such items from the market. Optical sensors and ANNs are needed to manage these intricate tasks in this context. However, in situations such as identifying components with very low concentrations, ANNs face difficulty providing accurate predictions. Some articles utilizing artificial neural networks for safety prediction are presented in Table 1.

Using HSI or spectroscopy and CNNs has enabled the detection of pesticide residues and toxic biological compounds in agricultural products. For example, Jiang et al. (2019) detected four different pesticides in apples (chlorpyrifos, mycelium and mixtures of those two at 100 ppm) using 72 hyperspectral images per pesticide. They predicted using KNN, SVM, and an AlexNet CNN, which obtained a 99 % accuracy. Zhu et al. (2022) detected aflatoxin B1 in peanuts using the spectra of every pixel (197,265 spectra) and a customised CNN obtaining a 99 % accuracy. Wu et al. (2019) employed NIR spectroscopy to detect two pesticides (fenvalerate and triazoline at 0.24 and 0.25 mg/L) in 240 lettuce leaves. DBN was used for data extraction, and SVM, PLS-DA, and KNN were used for classification. The combination of DBM-SVM was the best, with a 95 % accuracy. Besides, Raman spectroscopy is gaining attention for its ability to detect even small concentrations of toxic components and pesticides, as in rice chemical residues, where Weng et al. (2023) identified chlormequat chloride and acephate at 15.5 and 100 mg/L concentrations in rice with an R^2 of 0.98.

Table 1. Quality and safety assessment articles that use ANNs

Application	Data type	Topic	Models used	Results	Reference
Quality assessment	X-Ray	Mango seed spoilage location	CNN: VGG16	97 % accuracy	(Ansah et al., 2023)
	Spectroscopy	Estimation of crude protein in corn feed	PLS, SVM, CNN: CropNet, CropResNet	CropResNet R^2 = 0.89	(Liang et al., 2024)
	Spectroscopy	Apple soluble solid prediction	ANN, PLS, RF	Proposed ANN R^2 = 0.99	(Bai et al., 2019)
	HSI images	Maise kernel oil content prediction	CNN and PLSR	R^2 = 0.92	(L. Zhang, An, et al., 2022)
	HSI images	Kyoho grape firmness and pH prediction	PLS, SVM, AE+PLS, AE+SVM	AE+SVM R^2 = 0.92	(M. Xu et al., 2022)
	HSI images	Salted sea cucumber moisture prediction	CNN: proposed CNN	R^2 = 0.99	(Zeng et al., 2022)
	HSI images	Bruise detection in Plums	CNN: ResNet50, HSCNN, customised CNN	ResNet is the most reliable. F1 = 90 %	(Castillo-Girones et al., 2024)
	X-Ray images	Internal disorder segmentation in pear	CNN: U-Net	Intersection Over Union (IOU) of 0.88	(Tempelaere et al., 2023)
	X-Ray	Hass avocado internal fruit rot	CNN: U-Net	98 % Classification accuracy. RMSE segmentation = 3.15	(Matsui et al., 2023)
	HSI images	In-field strawberry ripeness estimation	CNN and SVM	CNN (AlexNet) 98 % accuracy	(Z. Gao et al., 2020)
	HSI images	Mango maturity estimation	CNN and PLS	CNN R^2 = 0.64	(Wendel et al., 2018)
	RGB images	Automated Zizania quality grading	CNN: LightNet, AlexNet, VGG11, ResNet18, MobileNet, MobileNetV2 and ShuffleNet V2	LightNet is the best and fastest, with 99 % accuracy	(Cao et al., 2021)
	RGB images	Design of a winter jujube grading robot	CNN: YOLOv3, SSD and a Faster-R-CNN	YOLOv3 97 % accuracy	(Z. Lu et al., 2021)
	RGB images	Citrus peel fruit defect detection	CNN: proposed model based on YOLOv5, Yolo5s, Yolo7tiny, Yolo8n	99 % average precision	(J. Lu et al., 2024)
	Multispectral images	Olive crop nutritional status assessment	MLP, PLS, SVM and Gaussian Process Regression (GPR)	MLP better results	(Noguera et al., 2021)
	HSI images	Oilseed rape nitrogen prediction	AE+MLP	R^2 = 0.90	(X. Yu et al., 2018)
	Spectroscopy	Fresh lettuce potassium prediction	RBFN, PLS	RBFN R^2 = 0.86	(Xiong et al., 2021)

Toxic compounds and residues detection	HSI images	Peanut aflatoxin detection	CNN	99 % accuracy	(H. Zhu et al., 2022)
	HSI images	Apple pesticide residue detection	KNN, SVM, CNN: AlexNet	99 % accuracy	(Jiang et al., 2019)
	Raman spectroscopy	Rice chemical residue identification	CNN: Squeezenet	0.98 coefficient of determination	(Weng et al., 2023)
	Spectroscopy	Lettuce leaves pesticide residue identification	DBN for data extraction and SVM, PLS-DA, and KNN for classification	DBM-SVM 95 % accuracy	(Wu et al., 2019)

2.6. Conclusions and future trends

ANNs have become key tools in agriculture, agricultural products and quality and safety assessment, addressing critical challenges and enhancing efficiency across various applications. CNNs are the dominant approach for image-based tasks due to their exceptional image analysis, feature extraction, and pattern recognition capabilities. They are particularly effective in quality assessment. For example, CNNs enable early detection of pests, diseases or damages by identifying subtle visual patterns in RGB or hyperspectral images. Besides, emerging techniques like Capsule Networks offer promising solutions to overcome CNN limitations by incorporating object position identification and requiring less training data.

Additionally, AEs, GANs and Diffusion Models can be very useful for data augmentation, generating synthetic datasets to address the scarcity of labelled data. These techniques enhance model performance in tasks like pest detection and weed recognition. AEs and DBNs further support image-based applications by extracting features from complex datasets.

For non-image data, such as spectroscopic or time-series inputs, neural networks like MLPs, 1D-CNNs, RNNs, and RBFNs are commonly employed. MLPs are versatile for predicting the nutrient content of crops or quality attributes like sugar levels in fruits, while RNNs excel in capturing temporal dependencies for monitoring crop growth cycles. RBFNs effectively model localised relationships for specific predictions such as nutrient levels. Emerging architectures like Transformers demonstrate potential for large-scale crop mapping and species classification using multispectral and spatiotemporal data. Their ability to process diverse data types positions them as powerful tools for precision agriculture. Meanwhile, SOMs aid in clustering and visualising agricultural datasets, offering insights into underlying patterns.

The application of ANNs is expected to expand with the integration of emerging technologies such as the Internet of Things, cloud platforms, and 5G networks, together with the cost reduction of hardware capable of running ANNs. These technologies enable real-time data collection from sensors, providing comprehensive datasets for ANN models. Hybrid approaches that combine ANNs with other machine learning techniques or advanced algorithms like reinforcement learning could further enhance predictive accuracy and decision-making capabilities. Additionally, advancements in explainable AI likely improve trust in ANN-driven systems by providing interpretable insights into their decisions.

Despite these advancements, significant challenges remain. Field variability, resulting from differences in soil types, weather conditions, or crop varieties, creates complexities that necessitate robust models capable of generalising across diverse environments. The limited availability of labelled datasets is a critical barrier for training models. Furthermore, there are infrastructure limitations such as insufficient internet bandwidth or computational resources to deploy ANN-based systems effectively. Another challenge is the cost of implementing ANN solutions at scale. Small farmers or companies may lack access to the technology or expertise required to adopt these innovations. Ethical considerations regarding data privacy and ownership need attention as agriculture becomes increasingly digitised.

To address these challenges, data collection strategies should be implemented to create repositories of diverse datasets while ensuring privacy and security. Investments in connectivity infrastructure are also essential to support solutions for real-time ANN applications. In addition, creating affordable hardware solutions would promote equitable access to these technologies. Implementing training programs on effectively using tools based on ANNsis is also important.

Acknowledgements:

This work was partially funded through the projects MICIU AEI PID2023-150192OR-C31, C32 and C-33 with the support of FEDER funds, project AEI TED2021-130117B-C31 and C33, by the project GVA-PROMETEO CIPROM/2021/014. Salvador Castillo thanks INIA for the FPI-INIA grant number PRE2020-094491, partially supported by European Union FSE funds. Sandra Munera expresses her thanks for the postdoctoral contract Juan de la Cierva - Formación (FJC2021-047786-I) co-funded by MCIN/AEI/10.13039/501100011033 and European Union NextGenerationEU/PRTR.

References

Abbas, A., Jain, S., Gour, M., & Vankudothu, S. (2021). Tomato plant disease detection using transfer learning with C-GAN synthetic images. *Computers and Electronics in Agriculture*, *187*, 106279. https://doi.org/https://doi.org/10.1016/j.compag.2021.106279

Abbas, O., Pissard, A., & Baeten, V. (2020). Near-infrared, mid-infrared, and Raman spectroscopy. *Chemical Analysis of Food: Techniques and Applications, Second Edition*, 77–134. https://doi.org/10.1016/B978-0-12-813266-1.00003-6

Agarwal, M., Al-Shuwaili, T., Nugaliyadde, A., Wang, P., Wong, K. W., & Ren, Y. (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. *Computers and Electronics in Agriculture*, *173*, 105438. https://doi.org/https://doi.org/10.1016/j.compag.2020.105438

Agarwal, U. P., & Atalla, R. H. (2020). Raman Spectroscopy. *Surface Analysis of Paper*, 152–181. https://doi.org/10.1201/9780429279997-8

Amara, J., Bouaziz, B., & Algergawy, A. (2017). A Deep Learning-based Approach for Banana Leaf Diseases Classification. *BTW*.

Amigo, J. M. (2019). Chapter 1.1 - Hyperspectral and multispectral imaging: setting the scene. In J. M. Amigo (Ed.), *Hyperspectral Imaging* (Vol. 32, pp. 3–16). Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-444-63977-6.00001-8

Amigo, J. M., & Grassi, S. (2019). Configuration of hyperspectral and multispectral imaging systems. *Data Handling in Science and Technology*, *32*, 17–34. https://doi.org/10.1016/B978-0-444-63977-6.00002-X

Ansah, F. A., Amo-Boateng, M., Siabi, E. K., & Bordoh, P. K. (2023). Location of seed spoilage in mango fruit using X-ray imaging and convolutional neural networks. *Scientific African*, *20*, e01649. https://doi.org/https://doi.org/10.1016/j.sciaf.2023.e01649

Arul, V. H. (2021). 5 - Deep learning methods for data classification. In D. Binu & B. R. Rajakumar (Eds.), *Artificial Intelligence in Data Mining* (pp. 87–108). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-820601-0.00001-X

Attri, I., Awasthi, L. K., Sharma, T. P., & Rathee, P. (2023). A review of deep learning techniques used in agriculture. *Ecological Informatics*, *77*, 102217. https://doi.org/10.1016/J.ECOINF.2023.102217

Aurélien Géron. (2019). Hands-on machine learning with Scikit-Learn, Keras and TensorFlow: concepts, tools, and techniques to build intelligent systems. *O'Reilly Media*, 851. https://www.oreilly.com/library/view/hands-on-machine-learning/9781492032632/

Bai, Y., Xiong, Y., Huang, J., Zhou, J., & Zhang, B. (2019). Accurate prediction of soluble solid content of apples from multiple geographical regions by combining deep learning with spectral fingerprint features. *Postharvest Biology and Technology*, *156*, 110943. https://doi.org/https://doi.org/10.1016/j.postharvbio.2019.110943

Balasubramanian, M., & Schwartz, E. (2002). Schwartz, E.L.: The Isomap Algorithm and Topological Stability. Science 295, 5552. *Science (New York, N.Y.)*, *295*, 7. https://doi.org/10.1126/science.295.5552.7a

Bank, D., Koenigstein, N., & Giryes, R. (2021). *Autoencoders*.

Bank, D., Koenigstein, N., & Giryes, R. (2023). Autoencoders. *Machine Learning for Data Science Handbook*, 353–374. https://doi.org/10.1007/978-3-031-24628-9_16

Bengio, Y. (2009). Learning Deep Architectures for AI. *Foundations and Trends® in Machine Learning*, *2*(1), 1–127. https://doi.org/10.1561/2200000006

Bisht, A., & Singh, S. P. (2024). Postharvest Losses and Management of Horticultural Produce: A Review. *Journal of Scientific Research and Reports*, *30*(3), 305–320. https://doi.org/10.9734/JSRR/2024/V30I31881

Blazhko, U., Shapaval, V., Kovalev, V., & Kohler, A. (2021). Comparison of augmentation and pre-processing for deep learning and chemometric classification of infrared spectra. *Chemometrics and Intelligent Laboratory Systems*, *215*, 104367. https://doi.org/https://doi.org/10.1016/j.chemolab.2021.104367

Bocklitz, T. W., Dörfer, T., Heinke, R., Schmitt, M., & Popp, J. (2015). Spectrometer calibration protocol for Raman spectra recorded with different excitation wavelengths. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *149*, 544–549. https://doi.org/10.1016/J.SAA.2015.04.079

Bonora, A., Bortolotti, G., Bresilla, K., Grappadelli, L. C., & Manfrini, L. (2022). Reprint of: A convolutional neural network approach to detecting fruit physiological disorders and maturity in 'Abbé Fétel' pears. *Biosystems Engineering*, *223*, 224–232. https://doi.org/https://doi.org/10.1016/j.biosystemseng.2022.10.005

Broomhead, D. S., & Lowe, D. (1988). Multivariable Functional Interpolation and Adaptive Networks. *Complex Syst.*, *2*.

Butac, M. (2020). *Plum Breeding*. https://doi.org/10.5772/intechopen.92432

Cao, J., Sun, T., Zhang, W., Zhong, M., Huang, B., Zhou, G., & Chai, X. (2021). An automated zizania quality grading method based on deep classification model. *Computers and Electronics in Agriculture*, *183*, 106004. https://doi.org/https://doi.org/10.1016/j.compag.2021.106004

Castillo-Girones, S., Munera, S., Martínez-Sober, M., Blasco, J., Cubero, S., & Gómez-Sanchis, J. (2025). Artificial Neural Networks in Agriculture, the core of artificial intelligence: What, When, and Why. *Computers and Electronics in Agriculture*, *230*, 109938. https://doi.org/10.1016/J.COMPAG.2025.109938

Castillo-Girones, S., Van Belleghem, R., Wouters, N., Munera, S., Blasco, J., & Saeys, W. (2024). Detection of subsurface bruises in plums using spectral imaging and deep learning with wavelength selection. *Postharvest Biology and Technology*, *207*, 112615. https://doi.org/https://doi.org/10.1016/j.postharvbio.2023.112615

Chen, D., Qi, X., Zheng, Y., Lu, Y., & Li, Z. (2022). *Deep Data Augmentation for Weed Recognition Enhancement: A Diffusion Probabilistic Model and Transfer Learning Based Approach*. https://doi.org/10.48550/arxiv.2210.09509

Chen, Z., Ting, D., Newbury, R., & Chen, C. (2021). Semantic segmentation for partially occluded apple trees based on deep learning. *Computers and Electronics in Agriculture*, *181*, 105952. https://doi.org/https://doi.org/10.1016/j.compag.2020.105952

Chollet, F. (2018). *Deep Learning with Python*. Manning Publications Co.

Chollet, F. (2021a). Deep Learning with Python, Second Edition. *Deep Learning with Python*. https://www.oreilly.com/library/view/deep-learning-with/9781617296864/

Chollet, F. (2021b). Deep Learning with Python, Second Edition. *Deep Learning with Python*. https://www.manning.com/books/deep-learning-with-python-second-edition

Chollet, F. (2021c). Deep Learning with Python, Second Edition. *Deep Learning with Python*. https://www.manning.com/books/deep-learning-with-python-second-edition

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences, Third Edition. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences, Third Edition*, 1–704. https://doi.org/10.4324/9780203774441/APPLIED-MULTIPLE-REGRESSION-CORRELATION-ANALYSIS-BEHAVIORAL-SCIENCES-JACOB-COHEN-PATRICIA-COHEN-STEPHEN-WEST-LEONA-AIKEN

Crisosto, C. H., Crisosto, G. M., & Garner, D. (2005). Understanding tree fruit consumer acceptance. *Acta Horticulturae*, *682*, 865–870. https://doi.org/10.17660/ACTAHORTIC.2005.682.112

Dai, T., Hao, Z., Wang, Z., -, al, Ali, S. K., Sadoon -, S. A., Chen, H., Liu, Y., & Pan, X. (2020). Feature Extraction Methods: A Review. *Journal of Physics: Conference Series*, *1591*(1), 012028. https://doi.org/10.1088/1742-6596/1591/1/012028

Das, R. S., & Agrawal, Y. K. (2011). Raman spectroscopy: Recent advancements, techniques and applications. *Vibrational Spectroscopy*, *57*(2), 163–176. https://doi.org/10.1016/J.VIBSPEC.2011.08.003

Dawson, P. L., & Acton, J. C. (2018). 22 - Impact of proteins on food color. In R. Y. Yada (Ed.), *Proteins in Food Processing (Second Edition)* (Second Edition, pp. 599–638). Woodhead Publishing. https://doi.org/https://doi.org/10.1016/B978-0-08-100722-8.00023-1

Dietzek, B., Cialla, D., Schmitt, M., & Popp, J. (2018). Introduction to the Fundamentals of Raman Spectroscopy. *Springer Series in Surface Sciences*, *66*, 47–68. https://doi.org/10.1007/978-3-319-75380-5_3

Diz, E. L., Portela, R., Bañares, M. A., Charitidis, C., Ntziouni, A., Li, X., Thomson, J., & Xiarchos, I. (2022). Review of Existing Standards, Guides, and Practices for Raman Spectroscopy. *Applied Spectroscopy, Vol. 76, Issue 7, Pp. 747-772*, *76*(7), 747–772. https://doi.org/10.1364/AS.76.000747

Dong, C., Zhang, Z., Yue, J., & Zhou, L. (2021). Automatic recognition of strawberry diseases and pests using convolutional neural network. *Smart Agricultural Technology*, *1*, 100009. https://doi.org/https://doi.org/10.1016/j.atech.2021.100009

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). *AN IMAGE IS WORTH 16X16 WORDS: TRANSFORMERS FOR IMAGE RECOGNITION AT SCALE*. https://doi.org/https://doi.org/10.48550/arXiv.2010.11929

Eckhardt, G. (1966). Selection of Raman Laser Materials. *IEEE Journal of Quantum Electronics*, *QE-2*(1), 1–8. https://doi.org/10.1109/JQE.1966.1073738

El Aboudi, N., & Benhlima, L. (2016). Review on wrapper feature selection approaches. *2016 International Conference on Engineering & MIS (ICEMIS)*, 1–5. https://doi.org/10.1109/ICEMIS.2016.7745366

Elgendy, M. (2020). Deep Learning for Vision Systems. *Manning Publications Co.*, 480. https://www.manning.com/books/deep-learning-for-vision-systems

Elgendy, M., O'Reilly for Higher Education (Firm), & Safari, an O. M. Company. (2020). *Deep Learning for Vision Systems*. Manning Publications Co. https://books.google.com/books/about/Deep_Learning_for_Vision_Systems.html?hl=es&id=6gkLzAEACAAJ

Eurostat. (2023). https://ec.europa.eu/eurostat/databrowser/view/apro_cpsh1__custom_8070269/default/table?lang=en

Fan, M., Lai, F. J., Chou, H. L., Lu, W. T., Hwang, B. J., & Brolo, A. G. (2012). Surface-enhanced Raman scattering (SERS) from Au:Ag bimetallic nanoparticles : the effect of the molecular probe. *Chemical Science*, *4*(1), 509–515. https://doi.org/10.1039/C2SC21191B

Fan, S., Li, J., Zhang, Y., Tian, X., Wang, Q., He, X., Zhang, C., & Huang, W. (2020). On line detection of defective apples using computer vision system combined with deep learning methods. *Journal of Food Engineering*, *286*, 110102. https://doi.org/https://doi.org/10.1016/j.jfoodeng.2020.110102

Fan, Y., Zhang, C., Liu, Z., Qiu, Z., & He, Y. (2019). Cost-sensitive stacked sparse auto-encoder models to detect striped stem borer infestation on rice based on hyperspectral imaging. *Knowledge-Based Systems*, *168*, 49–58. https://doi.org/https://doi.org/10.1016/j.knosys.2019.01.003

FAOSTAT. (2023). https://www.fao.org/faostat/en/#data/QCL

Fearn, T., Riccioli, C., Garrido-Varo, A., & Guerrero-Ginel, J. E. (2009). On the geometry of SNV and MSC. *Chemometrics and Intelligent Laboratory Systems*, *96*(1), 22–26. https://doi.org/https://doi.org/10.1016/j.chemolab.2008.11.006

Feng, L., Zhu, S., Zhou, L., Zhao, Y., Bao, Y., Zhang, C., & He, Y. (2019). Detection of Subtle Bruises on Winter Jujube Using Hyperspectral Imaging With Pixel-Wise Deep Learning Method. *IEEE Access*, *7*, 64494–64505. https://doi.org/10.1109/ACCESS.2019.2917267

Fernández-Ruiz, V., Torrecilla, J. S., Cámara, M., Sánchez Mata, M. C., & Shoemaker, C. (2010). Radial basis network analysis of color parameters to estimate lycopene content on tomato fruits. *Talanta*, *83*(1), 9–13. https://doi.org/https://doi.org/10.1016/j.talanta.2010.08.020

Fernandez-Zamudio, M. A., Barco, H., & Schneider, F. (2020). Direct Measurement of Mass and Economic Harvest and Post-Harvest Losses in Spanish Persimmon Primary Production. *Agriculture 2020, Vol. 10, Page 581*, *10*(12), 581. https://doi.org/10.3390/AGRICULTURE10120581

Ferraro, J. R., Nakamoto, K., & Brown, C. W. (2003). Introductory Raman Spectroscopy: Second Edition. *Introductory Raman Spectroscopy: Second Edition*, 1–434. https://doi.org/10.1016/B978-0-12-254105-6.X5000-8

Fryling, M., Frank, C. J., & Mccreery, R. L. (1993). Intensity Calibration and Sensitivity Comparisons for CCD/Raman Spectrometers. *Http://Dx.Doi.Org/10.1366/0003702934066226*, *47*(12), 1965–1974. https://doi.org/10.1366/0003702934066226

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, *36*(4), 193–202. https://doi.org/10.1007/BF00344251

Gao, F., Fang, W., Sun, X., Wu, Z., Zhao, G., Li, G., Li, R., Fu, L., & Zhang, Q. (2022). A novel apple fruit detection and counting methodology based on deep learning and trunk tracking in modern orchard. *Computers and Electronics in Agriculture*, *197*, 107000. https://doi.org/https://doi.org/10.1016/j.compag.2022.107000

Gao, Z., Shao, Y., Xuan, G., Wang, Y., Liu, Y., & Han, X. (2020). Real-time hyperspectral imaging for the in-field estimation of strawberry ripeness with deep learning. *Artificial Intelligence in Agriculture*, *4*, 31–38. https://doi.org/https://doi.org/10.1016/j.aiia.2020.04.003

Garcia, J. B., Gornay, J., Lacoue-Negre, M., Garcia, S. M., Er-Rmyly, J., Bendoula, R., & Roger, J.-M. (2022). A novel methodology for determining effectiveness of preprocessing methods in reducing undesired spectral variability in near infrared spectra. *Journal of Near Infrared Spectroscopy*, *30*(2), 74–88. https://doi.org/10.1177/09670335211047959

Geladi, P., Burger, J., & Lestander, T. (2004). Hyperspectral imaging: calibration problems and solutions. *Chemometrics and Intelligent Laboratory Systems*, *72*(2), 209–217. https://doi.org/10.1016/J.CHEMOLAB.2004.01.023

Ghose, D., Das, U., & Roy, P. (2018). Modeling response of runoff and evapotranspiration for predicting water table depth in arid region using dynamic recurrent neural network. *Groundwater for Sustainable Development*, *6*, 263–269. https://doi.org/https://doi.org/10.1016/j.gsd.2018.01.007

Gierlinger, N., & Schwanninger, M. (2007). The potential of Raman microscopy and Raman imaging in plant research. *Spectroscopy*, *21*(2), 69–89.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Nets. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems* (Vol. 27). Curran Associates, Inc. https://proceedings.neurips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf

Grahn, H. F., & Geladi, P. (2007). Techniques and Applications of Hyperspectral Image Analysis. *Techniques and Applications of Hyperspectral Image Analysis*, 1–368. https://doi.org/10.1002/9780470010884

Griffiths, P. R. (2001). Introduction to Vibrational Spectroscopy. *Handbook of Vibrational Spectroscopy*. https://doi.org/10.1002/0470027320.S0102

Gutiérrez, S., Wendel, A., & Underwood, J. (2019). Ground based hyperspectral imaging for extensive mango yield estimation. *Computers and Electronics in Agriculture*, *157*, 126–135. https://doi.org/https://doi.org/10.1016/j.compag.2018.12.041

Guyon, I., Nikravesh, M., Gunn, S., & Zadeh, L. A. (Eds.). (2006). *Feature Extraction*. *207*. https://doi.org/10.1007/978-3-540-35488-8

Haenlein, M., & Kaplan, A. M. (2004). A Beginner's Guide to Partial Least Squares Analysis. *Understanding Statistics*, *3*(4), 283–297. https://doi.org/10.1207/S15328031US0304_4

Han, X. X., Rodriguez, R. S., Haynes, C. L., Ozaki, Y., & Zhao, B. (2022). Surface-enhanced Raman spectroscopy. *Nature Reviews Methods Primers 2022 1:1*, *1*(1), 1–17. https://doi.org/10.1038/s43586-021-00083-6

Hashemi Fath, A., Madanifar, F., & Abbasi, M. (2020). Implementation of multilayer perceptron (MLP) and radial basis function (RBF) neural networks to predict solution gas-oil ratio of crude oil systems. *Petroleum*, *6*(1), 80–91. https://doi.org/https://doi.org/10.1016/j.petlm.2018.12.002

Hinton, G. E., Ghahramani, Z., & Teh, Y. W. (1999). Learning to Parse Images. In S. Solla, T. Leen, & K. Müller (Eds.), *Advances in Neural Information Processing Systems* (Vol. 12). MIT Press. https://proceedings.neurips.cc/paper/1999/file/5a142a55461d5fef016acfb927fee0bd-Paper.pdf

Hinton, G. E., Osindero, S., & Teh, Y.-W. (2006). A Fast Learning Algorithm for Deep Belief Nets. *Neural Computation*, *18*(7), 1527–1554. https://doi.org/10.1162/neco.2006.18.7.1527

HLEG. (2019). *HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE A DEFINITION OF AI: MAIN CAPABILITIES AND DISCIPLINES Definition developed for the purpose of the AI HLEG's deliverables*. https://ec.europa.eu/digital-single-

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising Diffusion Probabilistic Models. *Advances in Neural Information Processing Systems*, *2020-December*. https://doi.org/10.48550/arxiv.2006.11239

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Huang, Y., Wang, H., Huang, H., Tan, Z., Hou, C., Zhuang, J., & Tang, Y. (2025). Raman Spectroscopy and Its Application in Fruit Quality Detection. *Agriculture 2025, Vol. 15, Page 195*, *15*(2), 195. https://doi.org/10.3390/AGRICULTURE15020195

Huang, Y., Wang, J., Li, N., Yang, J., & Ren, Z. (2021). Predicting soluble solids content in "Fuji" apples of different ripening stages based on multiple information fusion. *Pattern Recognition Letters*, *151*, 76–84. https://doi.org/https://doi.org/10.1016/j.patrec.2021.08.003

Introduction to Deep Learning - MATLAB. (n.d.). Retrieved October 22, 2021, from https://es.mathworks.com/videos/series/introduction-to-deep-learning.html?s_tid=wtest_ctent_PDF

Jahanbakhshi, A., Momeny, M., Mahmoudi, M., & Zhang, Y.-D. (2020). Classification of sour lemons based on apparent defects using stochastic pooling mechanism in deep convolutional neural networks. *Scientia Horticulturae*, *263*, 109133. https://doi.org/https://doi.org/10.1016/j.scienta.2019.109133

Jakubek, R. S., & Fries, M. D. (2020). Calibration of Raman wavenumber in large Raman images using a mercury-argon lamp. *Journal of Raman Spectroscopy*, *51*(7), 1172–1185. https://doi.org/10.1002/JRS.5887

Javanmardi, S., Miraei Ashtiani, S.-H., Verbeek, F. J., & Martynenko, A. (2021). Computer-vision classification of corn seed varieties using deep convolutional neural network. *Journal of Stored Products Research*, *92*, 101800. https://doi.org/https://doi.org/10.1016/j.jspr.2021.101800

Jiang, B., He, J., Yang, S., Fu, H., Li, T., Song, H., & He, D. (2019). Fusion of machine vision technology and AlexNet-CNNs deep learning network for the detection of postharvest apple pesticide residues. *Artificial Intelligence in Agriculture*, *1*, 1–8. https://doi.org/https://doi.org/10.1016/j.aiia.2019.02.001

Jing, L., Gulcehre, C., Peurifoy, J., Shen, Y., Tegmark, M., Soljacic, M., & Bengio, Y. (2019). Gated Orthogonal Recurrent Units: On Learning to Forget. *Neural Computation*, *31*(4), 765–783. https://doi.org/10.1162/neco_a_01174

Jones, R. R., Hooper, D. C., Zhang, L., Wolverson, D., & Valev, V. K. (2019). Raman Techniques: Fundamentals and Frontiers. *Nanoscale Research Letters 2019 14:1*, *14*(1), 1–34. https://doi.org/10.1186/S11671-019-3039-2

Jović, A., Brkić, K., & Bogunović, N. (2015). A review of feature selection methods with applications. *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 1200–1205. https://doi.org/10.1109/MIPRO.2015.7160458

Kaselimi, M., Voulodimos, A., Daskalopoulos, I., Doulamis, N., & Doulamis, A. (2022). A Vision Transformer Model for Convolution-Free Multilabel Classification of Satellite Imagery in Deforestation Monitoring. *IEEE Transactions on Neural Networks and Learning Systems*, 1–9. https://doi.org/10.1109/TNNLS.2022.3144791

Kazlagić, A., & Omanović-Mikličanin, E. (2020). Application of Raman Spectroscopy in Food Forensics: A Review. *IFMBE Proceedings*, *73*, 257–263. https://doi.org/10.1007/978-3-030-17971-7_40

Kerekes, J. P., & Schott, J. R. (2006). Hyperspectral Imaging Systems. *Hyperspectral Data Exploitation: Theory and Applications*, 17–45. https://doi.org/10.1002/9780470124628.CH2

Kerr, L. T., Byrne, H. J., & Hennelly, B. M. (2015). Optimal choice of sample substrate and laser wavelength for Raman spectroscopic analysis of biological specimen. *Analytical Methods*, *7*(12), 5041–5052. https://doi.org/10.1039/C5AY00327J

Khalid, S., Khalil, T., & Nasreen, S. (2014). A survey of feature selection and feature extraction techniques in machine learning. *2014 Science and Information Conference*, 372–378. https://doi.org/10.1109/SAI.2014.6918213

Koch, M., & Mangelberger, K. (2025). *Fruit Logistica. European statistics handbook*. www.AMI-informiert.de

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, *43*(1), 59–69. https://doi.org/10.1007/BF00337288

Kwabena Patrick, M., Felix Adekoya, A., Abra Mighty, A., & Edward, B. Y. (2022). Capsule Networks – A survey. *Journal of King Saud University - Computer and Information Sciences*, *34*(1), 1295–1310. https://doi.org/https://doi.org/10.1016/j.jksuci.2019.09.014

Lal, T. N., Chapelle, O., Weston, J., & Elisseeff, A. (2006). Embedded Methods. In I. Guyon, M. Nikravesh, S. Gunn, & L. A. Zadeh (Eds.), *Feature Extraction: Foundations and Applications* (pp. 137–165). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-35488-8_6

Lalitha, V., & Latha, B. (2022). A review on remote sensing imagery augmentation using deep learning. *Materials Today: Proceedings*. https://doi.org/https://doi.org/10.1016/j.matpr.2022.03.341

Langer, J., Jimenez de Aberasturi, D., Aizpurua, J., Alvarez-Puebla, R. A., Auguié, B., Baumberg, J. J., Bazan, G. C., Bell, S. E. J., Boisen, A., Brolo, A. G., Choo, J., Cialla-May, D., Deckert, V., Fabris, L., Faulds, K., García de Abajo, F. J., Goodacre, R., Graham, D., Haes, A. J., … Liz-Marzán, L. M. (2020). Present and Future of Surface-Enhanced Raman Scattering. *ACS Nano*, *14*(1), 28–117. https://doi.org/10.1021/acsnano.9b04224

Le, J. A., El-Askary, H. M., Allali, M., & Struppa, D. C. (2017). Application of recurrent neural networks for drought projections in California. *Atmospheric Research*, *188*, 100–106. https://doi.org/https://doi.org/10.1016/j.atmosres.2017.01.002

Li, K., Zhao, W., Peng, R., & Ye, T. (2022). Multi-branch self-learning Vision Transformer (MSViT) for crop type mapping with Optical-SAR time-series. *Computers and Electronics in Agriculture*, *203*, 107497. https://doi.org/https://doi.org/10.1016/j.compag.2022.107497

Li, L.-H., & Tanone, R. (2022). Vision Transformer Approach for Vegetables Recognition. *2022 International Seminar on Application for Technology of Information and Communication (ISemantic)*, 113–118. https://doi.org/10.1109/iSemantic55962.2022.9920435

Li, X., Li, X., Zhang, S., Zhang, G., Zhang, M., & Shang, H. (2022). SLViT: Shuffle-convolution-based lightweight Vision transformer for effective diagnosis of sugarcane leaf diseases. *Journal of King Saud University - Computer and Information Sciences*. https://doi.org/https://doi.org/10.1016/j.jksuci.2022.09.013

Li, Y., Qian, M., Liu, P., Cai, Q., Li, X., Guo, J., Yan, H., Yu, F., Yuan, K., Yu, J., Qin, L., Liu, H., Wu, W., Xiao, P., & Zhou, Z. (2019). The recognition of rice images by UAV based on capsule network. *Cluster Computing*, *22*(4), 9515–9524. https://doi.org/10.1007/s10586-018-2482-7

Li, Y., Wright, A., Liu, H., Wang, J., Wang, G., Wu, Y., & Dai, L. (2019). Land use pattern, irrigation, and fertilization effects of rice-wheat rotation on water quality of ponds by using self-organizing map in agricultural watersheds. *Agriculture, Ecosystems & Environment*, *272*, 155–164. https://doi.org/10.1016/J.AGEE.2018.11.021

Liang, J., Wang, B., Xu, X., & Xu, J. (2024). Integrating portable NIR spectrometry with deep learning for accurate Estimation of crude protein in corn feed. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *314*, 124203. https://doi.org/https://doi.org/10.1016/j.saa.2024.124203

Liu, Y., Pu, H., & Sun, D.-W. (2021). Efficient extraction of deep image features using convolutional neural network (CNN) for applications in detecting and analysing complex food matrices. *Trends in Food Science & Technology*, *113*, 193–204. https://doi.org/https://doi.org/10.1016/j.tifs.2021.04.042

Lodhi, V., Chakravarty, D., & Mitra, P. (2019). Hyperspectral Imaging System: Development Aspects and Recent Trends. *Sensing and Imaging*, *20*(1), 1–24. https://doi.org/10.1007/S11220-019-0257-8/METRICS

Lu, J., Chen, W., Lan, Y., Qiu, X., Huang, J., & Luo, H. (2024). Design of citrus peel defect and fruit morphology detection method based on machine vision. *Computers and Electronics in Agriculture*, *219*, 108721. https://doi.org/https://doi.org/10.1016/j.compag.2024.108721

Lu, W., Guo, Y.-M., Dai, D.-J., Zhang, C.-Y., & Wang, X.-Y. (2018). Rice Germination Rate Detection Based on Fluorescent Spectrometry and Deep Belief Network. *Guang Pu Xue Yu Guang Pu Fen Xi/Spectroscopy and Spectral Analysis*, *38*, 1303–1312. https://doi.org/10.3964/j.issn.1000-0593(2018)04-1303-10

Lu, Z., Zhao, M., Luo, J., Wang, G., & Wang, D. (2021). Design of a winter-jujube grading robot based on machine vision. *Computers and Electronics in Agriculture*, *186*, 106170. https://doi.org/https://doi.org/10.1016/j.compag.2021.106170

Matsui, T., Sugimori, H., Koseki, S., & Koyama, K. (2023). Automated detection of internal fruit rot in Hass avocado via deep learning-based semantic segmentation of X-ray images. *Postharvest Biology and Technology*, *203*, 112390. https://doi.org/https://doi.org/10.1016/j.postharvbio.2023.112390

Méndez-Albores, A., & Hernández, I. (2018). *DETECCIÓN DE RESIDUOS DE GLIFOSATO EN MAÍZ FRENTE A LOS RIESGOS A LA SALUD HUMANA Y A ABIÓTICOS, MEDIANTE ESPECTROSCOPIA RAMAN AUMENTADA POR LA SUPERFICIE.*

Milella, A., Marani, R., Petitti, A., & Reina, G. (2019). In-field high throughput grapevine phenotyping with a consumer-grade depth camera. *Computers and Electronics in Agriculture*, *156*, 293–306. https://doi.org/https://doi.org/10.1016/j.compag.2018.11.026

Miljkovic, D. (2017). Brief review of self-organizing maps. *2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2017 - Proceedings*, 1061–1066. https://doi.org/10.23919/MIPRO.2017.7973581

Momeny, M., Jahanbakhshi, A., Jafarnezhad, K., & Zhang, Y.-D. (2020). Accurate classification of cherry fruit using deep CNN based on hybrid pooling approach. *Postharvest Biology and Technology*, *166*, 111204. https://doi.org/https://doi.org/10.1016/j.postharvbio.2020.111204

Moore, D. S. (2014). Raman Spectroscopy Fundamentals. *Handbook of Spectroscopy: Second, Enlarged Edition*, *4–4*, 1813–1830. https://doi.org/10.1002/9783527654703.CH56

Mosca, S., Conti, C., Stone, N., & Matousek, P. (2021). Spatially offset Raman spectroscopy. *Nature Reviews Methods Primers 2021 1:1*, *1*(1), 1–16. https://doi.org/10.1038/s43586-021-00019-0

Müller, A. C., & Guido, S. (2016). *Introduction to machine learning with Python : a guide for data scientists*. 378. https://www.oreilly.com/library/view/introduction-to-machine/9781449369880/

Munera, S., Amigo, J. M., Aleixos, N., Talens, P., Cubero, S., & Blasco, J. (2018). Potential of VIS-NIR hyperspectral imaging and chemometric methods to identify similar cultivars of nectarine. *Food Control*, *86*, 1–10. https://doi.org/https://doi.org/10.1016/j.foodcont.2017.10.037

Naidu, G., Zuva, T., & Sibanda, E. M. (2023). A Review of Evaluation Metrics in Machine Learning Algorithms. *Lecture Notes in Networks and Systems*, *724 LNNS*, 15–25. https://doi.org/10.1007/978-3-031-35314-7_2

Navidan, H., Moshiri, P. F., Nabati, M., Shahbazian, R., Ghorashi, S. A., Shah-Mansouri, V., & Windridge, D. (2021). Generative Adversarial Networks (GANs) in networking: A comprehensive survey & evaluation. *Computer Networks*, *194*, 108149. https://doi.org/https://doi.org/10.1016/j.comnet.2021.108149

Nayak, J., Vakula, K., Dinesh, P., Naik, B., & Pelusi, D. (2020). Intelligent food processing: Journey from artificial neural network to deep learning. *Computer Science Review*, *38*, 100297. https://doi.org/https://doi.org/10.1016/j.cosrev.2020.100297

Noguera, M., Aquino, A., Ponce, J. M., Cordeiro, A., Silvestre, J., Calderón, R., Marcelo, M. da E., Pedro Jordão, Andújar, J. M., da Encarnação Marcelo, M., Pedro Jordão, & Andújar, J. M. (2021). Nutritional status assessment of olive crops by means of the analysis and modelling of multispectral images taken with UAVs. *Biosystems Engineering*, *211*, 1–18. https://doi.org/https://doi.org/10.1016/j.biosystemseng.2021.08.035

Okubo, K. (2021). NIR Hyperspectral Imaging. *Transparency in Biology: Making the Invisible Visible*, 203–222. https://doi.org/10.1007/978-981-15-9627-8_10

Osako, Y., Yamane, H., Lin, S.-Y., Chen, P.-A., & Tao, R. (2020). Cultivar discrimination of litchi fruit images using deep learning. *Scientia Horticulturae*, *269*, 109360. https://doi.org/https://doi.org/10.1016/j.scienta.2020.109360

O'Shea, K., & Nash, R. (2015). An Introduction to Convolutional Neural Networks. *International Journal for Research in Applied Science and Engineering Technology*, *10*(12), 943–947. https://doi.org/10.22214/ijraset.2022.47789

Panczer, G., De Ligny, D., Mendoza, C., Gaft, M., Seydoux-Guillaume, A. M., & Wang, X. (2012). Raman and fluorescence. *European Mineralogical Union Notes in Mineralogy*, *12*(1), 61–82. https://doi.org/10.1180/EMU-NOTES.12.2

Park, B., Shin, T., Cho, J.-S., Lim, J.-H., & Park, K.-J. (2023). Improving blueberry firmness classification with spectral and textural features of microstructures using hyperspectral microscope imaging and deep learning. *Postharvest Biology and Technology*, *195*, 112154. https://doi.org/https://doi.org/10.1016/j.postharvbio.2022.112154

Pathmanaban, P., Gnanavel, B. K., & Anandan, S. S. (2019). Recent application of imaging techniques for fruit quality assessment. *Trends in Food Science & Technology*, *94*, 32–42. https://doi.org/https://doi.org/10.1016/j.tifs.2019.10.004

Patil, P. U., Lande, S. B., Nagalkar, V. J., Nikam, S. B., & Wakchaure, G. C. (2021). Grading and sorting technique of dragon fruits using machine learning algorithms. *Journal of Agriculture and Food Research*, *4*, 100118. https://doi.org/https://doi.org/10.1016/j.jafr.2021.100118

Phaye, S. S. R., Sikka, A., Dhall, A., & Bathula, D. R. (2018). Dense and Diverse Capsule Networks: Making the Capsules Learn Better. *ArXiv*, *abs/1805.0*.

Pillay, R., Hardeberg, J. Y., & George, S. (2019). Hyperspectral imaging of art: Acquisition and calibration workflows. *Journal of the American Institute for Conservation*, *58*(1–2), 3–15. https://doi.org/10.1080/01971360.2018.1549919

Polder, G., Blok, P. M., de Villiers, H. A. C., van der Wolf, J. M., & Kamp, J. (2019). Potato Virus Y Detection in Seed Potatoes Using Deep Learning on Hyperspectral Images. *Frontiers in Plant Science*, *10*, 209. https://doi.org/10.3389/fpls.2019.00209

Pourdarbani, R., Sabzi, S., Kalantari, D., Paliwal, J., Benmouna, B., García-Mateos, G., & Molina-Martínez, J. M. (2020). Estimation of different ripening stages of Fuji apples using image processing and spectroscopy based on the majority voting method. *Computers and Electronics in Agriculture*, *176*, 105643. https://doi.org/https://doi.org/10.1016/j.compag.2020.105643

Qin, J., Chao, K., Kim, M. S., Lu, R., & Burks, T. F. (2013). Hyperspectral and multispectral imaging for evaluating food safety and quality. *Journal of Food Engineering*, *118*(2), 157–171. https://doi.org/https://doi.org/10.1016/j.jfoodeng.2013.04.001

R, M., & Rohini, R. (2016). *LASSO: A feature selection technique in predictive modeling for machine learning*. 18–20. https://doi.org/10.1109/ICACA.2016.7887916

Reyes, C., Medina-Gutiérrez, C., Medina-Valtierra, J., & Berrú, R. (2004). Aplicación de la espectroscopia Raman para la caracterización de pesticidas orgánicos. *Revista Internacional de Contaminación Ambiental*.

Rinnan, Å., van den Berg, F., & Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry*, *28*(10), 1201–1222. https://doi.org/https://doi.org/10.1016/j.trac.2009.07.007

Roweis, S. T., & Saul, L. K. (2000). Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science*, *290*(5500), 2323–2326. https://doi.org/10.1126/science.290.5500.2323

S, S., M, L. C., E, G. G., G, D. P., F, M.-P., & B, D. (2020). *AI WATCH. Defining Artificial Intelligence . KJ-NA-30117-EN-N (online)*. https://doi.org/10.2760/382730 (online)

Salzer, R., & Siesler, H. W. (2014). *Infrared and Raman spectroscopic imaging: Second, Completely Revised and Updated Edition*. 57–110. https://www.wiley.com/en-us/Infrared+and+Raman+Spectroscopic+Imaging%2C+2nd%2C+Completely+Revised+and+Updated+Edition-p-9783527336524

Sánchez-Maroño, N., Alonso-Betanzos, A., & Tombilla-Sanromán, M. (2007). Filter methods for feature selection - A comparative study. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *4881 LNCS*, 178–187. https://doi.org/10.1007/978-3-540-77226-2_19/COVER

Santosh, K. C., Das, N., & Ghosh, S. (2022). Chapter 3 - Deep learning models. In K. C. Santosh, N. Das, & S. Ghosh (Eds.), *Deep Learning Models for Medical Imaging* (pp. 65–97). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-823504-1.00013-1

Seraj, A., Mohammadi-Khanaposhtani, M., Daneshfar, R., Naseri, M., Esmaeili, M., Baghban, A., Habibzadeh, S., & Eslamian, S. (2023). Chapter 5 - Cross-validation. In S. Eslamian & F. Eslamian (Eds.), *Handbook of Hydroinformatics* (pp. 89–105). Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-12-821285-1.00021-X

Sharma, N., Sharma, R., & Jindal, N. (2021). Machine Learning and Deep Learning Applications-A Vision. *Global Transitions Proceedings*, *2*(1), 24–28. https://doi.org/https://doi.org/10.1016/j.gltp.2021.01.004

Shinoda, R., Kataoka, H., Hara, K., & Noguchi, R. (2023). Transformer-based Ripeness Segmentation for Tomatoes. *Smart Agricultural Technology*, 100196. https://doi.org/https://doi.org/10.1016/j.atech.2023.100196

Shuaishuai, L., Li, L., shiwang, C., Fanjia, M., Haihua, W., Zhanzhan, S., & Sigrimis, N. A. (2018). Prediction Model of Transpiration Rate of Strawberry in Closed Cultivation Based on DBN-LSSVM Algorithm. *IFAC-PapersOnLine*, *51*(17), 460–465. https://doi.org/https://doi.org/10.1016/j.ifacol.2018.08.171

Sohl-Dickstein, J., Weiss, E. A., Maheswaranathan, N., & Ganguli, S. (2015). Deep Unsupervised Learning using Nonequilibrium Thermodynamics. *32nd International Conference on Machine Learning, ICML 2015*, *3*, 2246–2255. https://doi.org/10.48550/arxiv.1503.03585

Song, P., Wang, J., Guo, X., Yang, W., & Zhao, C. (2021). High-throughput phenotyping: Breaking through the bottleneck in future crop breeding. *The Crop Journal*, *9*(3), 633–645. https://doi.org/https://doi.org/10.1016/j.cj.2021.03.015

Song, X., Zhang, G., Liu, F., Li, D., Zhao, Y., & Yang, J. (2016). Modeling spatio-temporal distribution of soil moisture by deep learning-based cellular automata model. *Journal of Arid Land*, *8*(5), 734–748. https://doi.org/10.1007/s40333-016-0049-0

Špiclin, Ž., Katrašnik, J., Bürmen, M., Pernuš, F., & Likar, B. (2010). Geometric calibration of a hyperspectral imaging system. *Applied Optics, Vol. 49, Issue 15, Pp. 2813-2818*, *49*(15), 2813–2818. https://doi.org/10.1364/AO.49.002813

Sun, D. W. (2010). Hyperspectral Imaging for Food Quality Analysis and Control. *Hyperspectral Imaging for Food Quality Analysis and Control*, 1–477. https://doi.org/10.1016/C2009-0-01853-4

Sun, Y., Liu, H., & Guo, Z. (2021). Capsule network-based approach for estimating grassland coverage using time series data from enhanced vegetation index. *Artificial Intelligence in Geosciences*, *2*, 26–34. https://doi.org/https://doi.org/10.1016/j.aiig.2021.08.001

Tempelaere, A., Minh Phan, H., Van De Looverbosch, T., Verboven, P., & Nicolai, B. (2023). Non-destructive internal disorder segmentation in pear fruit by X-ray radiography and AI. *Computers and Electronics in Agriculture*, *212*, 108142. https://doi.org/https://doi.org/10.1016/j.compag.2023.108142

Thai, H.-T., Tran-Van, N.-Y., & Le, K.-H. (2021). Artificial Cognition for Early Leaf Disease Detection using Vision Transformers. *2021 International Conference on Advanced Technologies for Communications (ATC)*, 33–38. https://doi.org/10.1109/ATC52653.2021.9598303

Theanjumpol, P., Wongzeewasakun, K., Muenmanee, N., Wongsaipun, S., Krongchai, C., Changrue, V., Boonyakiat, D., & Kittiwachana, S. (2019). Non-destructive identification and estimation of granulation in 'Sai Num Pung' tangerine fruit using near infrared spectroscopy and chemometrics. *Postharvest Biology and Technology*, *153*, 13–20. https://doi.org/10.1016/J.POSTHARVBIO.2019.03.009

Torkashvand, A. M., Ahmadi, A., & Nikravesh, N. L. (2017). Prediction of kiwifruit firmness using fruit mineral nutrient concentration by artificial neural network (ANN) and multiple linear regressions (MLR). *Journal of Integrative Agriculture*, *16*(7), 1634–1644. https://doi.org/https://doi.org/10.1016/S2095-3119(16)61546-0

Valor, E., & Bruto, A. (n.d.). *AgrInfo no40. Contribución del Sistema Agroalimentario a la Economía Española. 2022. 2 2. Importancia económica del Sistema Agroalimentario*.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention Is All You Need. *Advances in Neural Information Processing Systems*, *2017-December*, 5999–6009. https://doi.org/10.48550/arxiv.1706.03762

Velliangiri, S., Alagumuthukrishnan, S., & Thankumar joseph, S. I. (2019). A Review of Dimensionality Reduction Techniques for Efficient Computation. *Procedia Computer Science*, *165*, 104–111. https://doi.org/https://doi.org/10.1016/j.procs.2020.01.079

Wang, B., Sun, J., Xia, L., Liu, J., Wang, Z., Li, P., Guo, Y., & Sun, X. (2023). The Applications of Hyperspectral Imaging Technology for Agricultural Products Quality Analysis: A Review. *Food Reviews International*, *39*(2), 1043–1062. https://doi.org/10.1080/87559129.2021.1929297

Wang, C., Liu, B., Liu, L., Zhu, Y., Hou, J., Liu, P., Li, · Xiang, Li, X., & Wang, C. (2021a). A review of deep learning used in the hyperspectral image analysis for agriculture. *Artificial Intelligence Review*, *54*, 5205–5253. https://doi.org/10.1007/s10462-021-10018-y

Wang, C., Liu, B., Liu, L., Zhu, Y., Hou, J., Liu, P., & Li, X. (2021b). A review of deep learning used in the hyperspectral image analysis for agriculture. *Artificial Intelligence Review*, *54*(7), 5205–5253. https://doi.org/10.1007/s10462-021-10018-y

Wang, D., Vinson, R., Holmes, M., Seibel, G., Bechar, A., Nof, S., & Tao, Y. (2019). Early Detection of Tomato Spotted Wilt Virus by Hyperspectral Imaging and Outlier Removal Auxiliary Classifier Generative Adversarial Nets (OR-AC-GAN). *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-40066-y

Wang, K., Gou, C., Duan, Y., Lin, Y., Zheng, X., & Wang, F.-Y. (2017). Generative adversarial networks: introduction and outlook. *IEEE/CAA Journal of Automatica Sinica*, *4*(4), 588–598. https://doi.org/10.1109/JAS.2017.7510583

Wang, P., Hafshejani, B. A., & Wang, D. (2021). An improved multilayer perceptron approach for detecting sugarcane yield production in IoT based smart agriculture. *Microprocessors and Microsystems*, *82*, 103822. https://doi.org/https://doi.org/10.1016/j.micpro.2021.103822

Wendel, A., Underwood, J., & Walsh, K. (2018). Maturity estimation of mangoes using hyperspectral imaging from a ground based mobile platform. *Computers and Electronics in Agriculture*, *155*, 298–313. https://doi.org/https://doi.org/10.1016/j.compag.2018.10.021

Weng, S., Tang, L., Qiu, M., Wang, J., Wu, Y., Zhu, R., Wang, C., Li, P., Sha, W., & Liang, D. (2023). Surface-enhanced Raman spectroscopy charged probes under inverted superhydrophobic platform for detection of agricultural chemicals residues in rice combined with lightweight deep learning network. *Analytica Chimica Acta*, *1262*, 341264. https://doi.org/https://doi.org/10.1016/j.aca.2023.341264

Wu, M., Sun, J., Lu, B., Ge, X., Zhou, X., & Zou, M. (2019). Application of deep brief network in transmission spectroscopy detection of pesticide residues in lettuce leaves. *Journal of Food Process Engineering*, *42*(3), e13005. https://doi.org/10.1111/JFPE.13005

Xiao, Q., Li, G., Xie, L., & Chen, Q. (2018). Real-world plant species identification based on deep convolutional neural networks and visual attention. *Ecological Informatics*, *48*, 117–124. https://doi.org/https://doi.org/10.1016/j.ecoinf.2018.09.001

Xin, Z., Jun, S., Yan, T., Quansheng, C., Xiaohong, W., & Yingying, H. (2020). A deep learning based regression method on hyperspectral data for rapid prediction of cadmium residue in lettuce leaves. *Chemometrics and Intelligent Laboratory Systems*, *200*, 103996. https://doi.org/https://doi.org/10.1016/j.chemolab.2020.103996

Xiong, Y., Ohashi, S., Nakano, K., Jiang, W., Takizawa, K., Iijima, K., & Maniwara, P. (2021). Application of the radial basis function neural networks to improve the nondestructive Vis/NIR spectrophotometric analysis of potassium in fresh lettuces. *Journal of Food Engineering*, *298*, 110417. https://doi.org/https://doi.org/10.1016/j.jfoodeng.2020.110417

Xu, M., Sun, J., Yao, K., Cai, Q., Shen, J., Tian, Y., & Zhou, X. (2022a). Developing deep learning based regression approaches for prediction of firmness and pH in Kyoho grape using Vis/NIR hyperspectral imaging. *Infrared Physics & Technology*, *120*, 104003. https://doi.org/https://doi.org/10.1016/j.infrared.2021.104003

Xu, M., Sun, J., Yao, K., Cai, Q., Shen, J., Tian, Y., & Zhou, X. (2022b). Developing deep learning based regression approaches for prediction of firmness and pH in Kyoho grape using Vis/NIR hyperspectral imaging. *Infrared Physics & Technology*, *120*, 104003. https://doi.org/https://doi.org/10.1016/j.infrared.2021.104003

Yang, C. H., Xiong, L. Y., Wang, Z., Wang, Y., Shi, G., Kuremot, T., Zhao, W. H., & Yang, Y. (2020). Integrated detection of citrus fruits and branches using a convolutional neural network. *Computers and Electronics in Agriculture*, *174*, 105469. https://doi.org/https://doi.org/10.1016/j.compag.2020.105469

Yang, Q., Shi, L., Han, J., Yu, J., & Huang, K. (2020). A near real-time deep learning approach for detecting rice phenology based on UAV images. *Agricultural and Forest Meteorology*, *287*, 107938. https://doi.org/https://doi.org/10.1016/j.agrformet.2020.107938

Yu, S., Xie, L., & Huang, Q. (2023). Inception convolutional vision transformers for plant disease identification. *Internet of Things*, *21*, 100650. https://doi.org/https://doi.org/10.1016/j.iot.2022.100650

Yu, X., Lu, H., & Liu, Q. (2018). Deep-learning-based regression model and hyperspectral imaging for rapid detection of nitrogen concentration in oilseed rape (Brassica napus L.) leaf. *Chemometrics and Intelligent Laboratory Systems*, *172*, 188–193. https://doi.org/https://doi.org/10.1016/j.chemolab.2017.12.010

Yu, Y., Jiang, T., Gao, J., Guan, H., Li, D., Gao, S., Tang, E., Wang, W., Tang, P., & Li, J. (2022). CapViT: Cross-context capsule vision transformers for land cover classification with airborne multispectral LiDAR data. *International Journal of Applied Earth Observation and Geoinformation*, *111*, 102837. https://doi.org/https://doi.org/10.1016/j.jag.2022.102837

Zeng, F., Shao, W., Kang, J., Yang, J., Zhang, X., Liu, Y., & Wang, H. (2022). Detection of moisture content in salted sea cucumbers by hyperspectral and low field nuclear magnetic resonance based on deep learning network framework. *Food Research International*, *156*, 111174. https://doi.org/https://doi.org/10.1016/j.foodres.2022.111174

Zhang, J., He, L., Karkee, M., Zhang, Q., Zhang, X., & Gao, Z. (2018). Branch detection for apple trees trained in fruiting wall architecture using depth features and Regions-Convolutional Neural Network (R-CNN). *Computers and Electronics in Agriculture*, *155*, 386–393. https://doi.org/https://doi.org/10.1016/j.compag.2018.10.029

Zhang, L., An, D., Wei, Y., Liu, J., & Wu, J. (2022). Prediction of oil content in single maize kernel based on hyperspectral imaging and attention convolution neural network. *Food Chemistry*, *395*, 133563. https://doi.org/https://doi.org/10.1016/j.foodchem.2022.133563

Zhang, L., Wang, Y., Wei, Y., & An, D. (2022). Near-infrared hyperspectral imaging technology combined with deep convolutional generative adversarial network to predict oil content of single maize kernel. *Food Chemistry*, *370*, 131047. https://doi.org/https://doi.org/10.1016/j.foodchem.2021.131047

Zhu, C. (Ed.). (2021). Appendix B - Deep learning basics. In *Machine Reading Comprehension* (pp. 213–237). Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-323-90118-5.00015-1

Zhu, H., Yang, L., Ding, W., & Han, Z. (2022). Pixel-level rapid detection of aflatoxin B1 based on 1D-modified temporal convolutional network and hyperspectral imaging. *Microchemical Journal*, *183*, 108020. https://doi.org/https://doi.org/10.1016/j.microc.2022.108020

Zhu, H., Yang, L., Sun, Y., & Han, Z. (2021). Identifying carrot appearance quality by an improved dense CapNet. *Journal of Food Process Engineering*, *44*(1), e13586. https://doi.org/https://doi.org/10.1111/jfpe.13586

OVERVIEW OF THE RESEARCH

While fruits themselves serve as practical case studies in this context, the primary significance of the research presented here lies in the innovative methods and advanced technologies applied to solve these problems.

This doctoral thesis compiles four different studies that evaluate novel non-destructive analytical techniques such as hyperspectral imaging and Raman spectroscopy and machine learning approaches to solve common problems encountered in fruit production and commercialisation. Specifically, deep learning algorithms have been employed effectively for tasks such as bruise detection in plums using 3D Convolutional Neural Networks (3D-CNNs) and Gradient-Weighted Class Activation Mapping (Grad-CAM) to understand the model functioning, feature extraction and dimensionality reduction from hyperspectral data for persimmon firmness prediction using Variational Autoencoders (VAE), pest identification on complex pest mixtures commonly applied in apples and apple orchards using 1D Convolutional Neural Networks (1D-CNNs) and Grad-CAM to understand the model functioning, as well as predicting fruit quality attributes such as Total Soluble Solids (TSS), Titratatable Acidity (TA) and texture, to avoid food fraud by cultivar differentiation of 17 different cultivars, and Dutch consumer acceptance through artificial neural networks (ANNs). Complementing these advanced computational methods are non-destructive state-of-the-art technologies like hyperspectral imaging (HSI), able to determine internal quality in fruits (Munera et al., 2018) and which has proven valuable for bruise detection in plums, texture prediction in persimmons, and internal quality assessment and consumer acceptance prediction in strawberries. Besides, Raman spectroscopy has demonstrated its potential for rapid pesticide detection (Méndez-Albores & Hernández, 2018; Reyes et al., 2004).

The selection of specific fruits for these studies has been strategically motivated by their regional agricultural importance. Apples and plums hold significant economic relevance in Belgium (Butac, 2020); thus, together with Belgian research institutions, the studies have focused their efforts on addressing practical problems such as early-stage bruise detection to avoid rotting and food loss in dark-skinned plum cultivars where bruises are not visible by the human eye using HSI and CNNs. Additionally, in apples, the rapid identification of pesticides in complex pesticide mixtures commonly applied to apples was studied using Raman spectroscopy and CNNs to ensure food safety and the complement of EU legislation. Similarly, strawberry cultivation represents a vital agricultural sector in the Netherlands (Eurostat, 2023), where both Dutch-grown strawberries and imported varieties from Spain are extensively marketed. Consequently, this research includes strawberries from both countries to evaluate consumer acceptance among Dutch consumers using VIS-NIR spectral imaging combined with ANN modelling. Additionally, the persimmon cultivar 'Rojo Brillante' is traditionally cultivated in the Valencian region of Spain and has a protected designation of origin (PDO), and Spain is the second largest persimmon producer in the world (FAOSTAT, 2023). Due to its commercial importance locally and internationally and the importance of texture in consumer acceptance, this cultivar was selected for studying firmness assessment through hyperspectral imaging combined with convolutional variational autoencoders (VAE) for efficient data reduction.

Ultimately, the research presented here focuses not just on individual fruits but on demonstrating how innovative combinations of deep learning techniques, such as CNNs, VAEs, and multilabel classification methods, along with advanced imaging technologies like hyperspectral imaging and Raman spectroscopy, can effectively tackle the critical challenges faced by modern agriculture. The outcomes presented highlight opportunities for improving fruit quality control processes. They include enhanced food safety standards through quick pesticide detection methods, reduced waste by identifying bruises early, optimised storage conditions based on texture prediction models, and a better alignment between product offerings and consumer preferences.

AIMS

General aims

This research project aimed to explore the potential of novel non-destructive techniques, such as hyperspectral imaging and Raman spectroscopy, in combination with ANNs and other machine learning algorithms. The goal was to assess fruit quality and safety, predict the internal qualities and properties of various fruits, enhance our understanding of consumer perception, and ensure the safety of fruits, including persimmons, plums, strawberries, and apples.

Specific aims:

1. To study and predict different quality parameters of 17 varieties of strawberries together with overall consumer acceptance using HSI, machine, and deep learning models (Chapter I).
   a) To find relations and differences between the quality measurements of 17 different varieties of strawberries and consumer acceptance.
   b) To predict quality parameters such as TSS, TA and Bite using HSI and machine learning models.
   c) To predict overall consumer acceptance of the varieties.
   d) To avoid food fraud by identifying 17 different cultivars using HSI.

2. To classify fruits of the persimmon variety 'Rojo Brillante' in different texture groups for packing house sorting using Autoencoders (Chapter II).
   a) To study flesh-breaking force texture curves and create different texture groups.
   b) To create a deep learning model which classifies persimmon in different texture groups.
   c) To prove the ability of VAEs for feature extraction and data reduction as a solution for managing high-dimensional data effectively.

3. To detect invisible mechanical bruises in dark skin plums variety 'Presenta' using HSI and CNNs (Chapter III).
   a. to assess and compare three CNN models with different architectures: two models using transfer learning techniques and another customised from the ground up.
   b. To select the most important wavelengths for the detection of bruises.

c. to provide deeper insights into the operational mechanisms of each CNN model and to pinpoint the specific regions of interest within the images that these models focus on using a GradCam analysis.

4. To detect multiple pesticide active compounds present in pesticide mixtures usually applied to apple fruit using Raman spectroscopy (Chapter IV).
   a) Analyse and compare the different active compounds and mixture spectra.
   b) Create multilabel classification models to identify different active compounds present in pest mixtures.
   c) Prove that Raman spectroscopy is capable of identifying pesticide residues.
   d) Evaluate and understand the proposed CNN model's behaviour.

CHAPTER I. Advanced evaluation of strawberry quality, consumer preference, and cultivar discrimination through spectral imaging and neural networks.

Advanced evaluation of strawberry quality, consumer preference, and cultivar discrimination through spectral imaging and neural networks

Salvador Castillo-Girones[1], Jos Ruizendaal[2], Xiomara Salas-Valderrama[2], Sandra Munera[3], Jose Blasco[1], Gerrit Polder[2]

[1] Agroengineering center, Instituto Valenciano de Investigaciones Agrarias. Carretera CV-315, Km 10.7, 46113, Moncada, Valencia, Spain

[2]Greenhouse Horticulture Business Unit, Wageningen University & Research. Droevendaalsesteeg 4, 6708 PB, Wageningen, Netherlands

[3] Department of graphic engineering, Universitat Politècnica de València. Camino de Vera, s/n. 46022, Valencia, Spain

Food Control, 175, 111339

https://doi.org/https://doi.org/10.1016/j.foodcont.2025.111339

Abstract

Strawberries are among the most popular fruits, and meeting the rising demand for high-quality, flavorful varieties requires understanding consumer preferences. Accurately predicting these preferences, assessing quality, and preventing food fraud are crucial for breeders and sellers. This helps breeders develop superior cultivars and enables sellers to sort and market strawberries by taste and quality. This study explores the prediction of the quality and the acceptance of Dutch consumers of seventeen strawberry cultivars and their discrimination using VIS-NIR spectral imaging with a spectral range between 400 and 1,000 nm and Artificial Neural Networks (ANNs), which was not done before. A total of 3,564 samples were utilised. Three algorithms: Support Vector Machine, XGBoost, and a Multilayer Perceptron (MLP), were evaluated to predict quality parameters, consumer acceptance, and cultivar discrimination. MLP models showed the highest accuracy, with R^2 values of 0.85 for total soluble solids, 0.81 for titratable acidity, 0.76 for bite, and 0.78 for overall consumer acceptance. For cultivar discrimination, the MLP model achieved an F1 score of 0.84. These findings highlight the potential of ANNs in enhancing product quality assessment, preventing food fraud, and aligning products with consumer preferences in the food industry.

Keywords*:* strawberry; quality; consumer acceptance; spectral imaging; machine learning

1. Introduction

Strawberry (*Fragaria × Ananassa*) is an important fruit in the agriculture and food industry, known for its sweet taste and vibrant red colour. A total of 12.5 million tons were produced worldwide in 2021, with a total cultivated area of 389,665 ha. China was the leading producer with 3.4 million tons, followed by the USA with 1.2 million tons, and Türkiye with 669,195 tons (*FAOSTAT*, 2023). Regarding production in the EU, Spain was the leading producer in 2021, with 360,570 tons, followed by Poland with 162,900 tons and Germany with 130,630 tons. (*Eurostat*, 2023). As a versatile ingredient in various food products, strawberries offer many opportunities for the food and beverage industries. Additionally, this berry is a rich source of essential vitamins, minerals, and antioxidants, making it a part of a healthy diet. (Giampieri et al., 2014). Their unique combination of flavours and nutritional attributes has led to a substantial demand for strawberries.

Consumers expect strawberries to have high quality and intense flavour. Thus, understanding what quality attributes influence consumers the most and developing non-destructive tools to assess these attributes are gaining interest from packing houses and sellers. Physical and chemical properties, such as texture, total soluble solids (TSS), and titratable acidity (TA), determine the sensory qualities perceived by fruit consumers and often require the destruction of the samples for analysis. TSS is measured using a refractometer to analyse the fruit juice, while a titrator is utilised to assess TA from the juice. Additionally, texture measurements also involve the destruction of the sample, which is done using a texturometer. Since those measurements are destructive, only a few representative samples can be analysed, which does not guarantee the quality of the whole production. Therefore, non-destructive tools and analyses need to be developed to assess the quality of the products and the satisfaction of the consumers. Besides, the quality and the consumer acceptance of a product depend on the specific cultivar used (Ulrich & Olbricht, 2016), making consumer acceptance prediction difficult. This was the case in the study by Ross et al. (2010), who correlated TSS and TA of sweet cherry fruit with flavour intensity with unsatisfactory results. Moreover, Harker et al. (2009) found that acceptance preferences varied between countries, highlighting the importance of considering regional and cultural factors when predicting acceptance.

Spectral imaging and spectroscopy are extremely useful for monitoring crops and fruits, being able to measure on-site with portable devices, as in pears by Yu and Yao. (2022). While spectroscopy provides information from specific parts of a sample, spectral imaging captures both spatial and spectral information from the entire sample, thus providing more comprehensive data than spectroscopy alone (Polder et al., 2024). Furthermore, these technologies are becoming robust for assessing fruit quality (Walsh et al., 2020), and predicting consumer acceptance due to the ability of light to penetrate the fruit tissue and obtain internal information (Castillo-Girones et al. (2024); Rodríguez-Ortega et al. (2023)).

Typically, spectral data are analysed using machine learning algorithms. These types of models are trained on large labelled datasets to solve practical problems through classification and regression techniques. Algorithms, such as Artificial Neural Networks (ANNs), are gaining popularity due to their ability to identify hidden patterns and non-linear relationships within the data and solve complex problems like those involving spectral data. Multilayer Perceptrons (MLPs), and Convolutional Neural Networks (CNNs) are two of the most commonly used ANNs. MLPs, composed of a few layers of neurons, are typically used with spectral data (Elgendy, 2020). Previous studies have achieved good results using spectral imaging and different machine-learning techniques to evaluate similar internal properties in fruit similar to strawberries. For instance, in grapes, Gao & Xu. (2022) predicted TSS and achieved R^2 values of 0.90 using Partial Least Squares (PLS) in gooseberries, Blas Saavedra et al. (2024) achieved R^2 values of 0.90, 0.93, 0.53 and 0.43 in the estimation of vitamin C content, firmness, TSS and TA, respectively, using SVM models. In sweet cherries, Pullanagari & Li. (2021) obtained R^2 values of 0.88 for TSS and 0.60 for firmness prediction using a PLS model. For strawberries, most of the studies extracted only the mean spectra of each fruit from the images, such as ElMasry et al. (2007), who obtained R^2 values of 0.90 and 0.80 for TSS and TA, Nagata et al. (2005), who obtained an R^2 of 0.87 for TSS or Weng et al. (2020), who obtained R^2 values greater than 0.93 for TSS content, and 0.84 for TA. Only Su et al. (2021) used image data, achieving low R^2 values, 0.56. Nevertheless, no scientific publications were found related to predicting consumer acceptance of strawberries, which might be due to the complexity of the factors and their interactions that influence sensory perception (Barbin et al., 2012).

Therefore, the aim of this study was to assess strawberry quality using visible and near-infrared (Vis-NIR) spectral imaging in order to predict not only traditional quality parameters such as TSS, TA, and texture but also consumer acceptance using a Dutch panel, an aspect that has not been explored before using spectral data. Additionally, the study also aimed to differentiate between seventeen strawberry cultivars grown in The Netherlands and Spain based on their spectral profiles. Finally, advanced machine learning models, such as Support Vector Machine (SVM), XGBoost, and MLP, was evaluated to predict both quality attributes, cultivar and consumer acceptance. Through this approach, the goal was to align product development with consumer preferences, enhance quality control, and prevent fraud in the strawberry supply chain.

2. Material and methods

2.1. Fruit samples and experimental design

The experiments were performed in a laboratory located in the Greenhouse Horticulture & Flower Bulbs (Wageningen University & Research, Bleiswijk, The Netherlands). A total of 3,564 commercially mature strawberries from seventeen different cultivars were analysed, with eight cultivars grown in the Netherlands and nine in Spain (Table 1).

Table 2. Description of the strawberries used in this study.

Number	Cultivar	Origin	Week	Num. of samples
1	S2019-0081	Spain	1	243
2	FS2004	Spain	1	243
3	FS2108	Spain	1	243
4	Balada	Spain	1	243
5	Calinda	Spain	1	243
6	FF22S35	Spain	1	243
7	FF22S36	Spain	1	243
8	Fandango	Spain	1	243
9	V13	Spain	1	243
10	Sonata	Netherlands	1	243
11	Cadenza	Netherlands	2	162
12	FE2131	Netherlands	2	162
13	FE2132	Netherlands	2	162
14	FE2133	Netherlands	2	162
15	Jenkka	Netherlands	2	162
16	V15	Netherlands	2	162
17	Verdi	Netherlands	2	162

The experiment was conducted in the first week of May and the third week of June 2023. Ten strawberry cultivars were used in the first week, a set of 2,430 sample fruit. For each cultivar, nine boxes of strawberries were provided. In the second week, seven cultivars were used, including 1,134 samples and six boxes with strawberries were used for every cultivar. In all cases, the boxes contained 27 strawberries. The fruit cultivated in the Netherlands was collected three days before starting the experiment and transported to the laboratory in a refrigerated van at 5° C. The fruit cultivated in Spain was collected three days before the experiment and sent in a refrigerated transport at 5 °C to the laboratory. In both cases, the fruit was stored in the laboratory at 5º C and 90 % relative humidity (RH) to avoid degradation before the experiment.

2.2. Spectral image acquisition

Before image acquisition, the fruit was kept at room temperature (20 °C) for one hour to attemperate. All fruit (3,564 strawberry samples) was imaged in the laboratory at a controlled temperature of 20 °C using a line-scanning camera (Specim FX10 from Specim, Oulu, Finland) sensitive in the 400 to 1,000 nm spectral range. The images were captured in steps of 2.7 nm, with a spatial resolution of 0.28 mm/scan and an exposure time of 10 ms. The scene was illuminated by one halogen lamp (125 V, 3.12 A, 390 W from Phillips, Amsterdam, The Netherlands) for the NIR range and LED lights (Effi-Flex-HSI-300-910-970-TR-$, from Effilux, Les Ules, France) for the Vis-NIR range. Immediately after the image acquisition, samples were stored again at 5 °C and 90 % RH, and on the following day, quality measurements and the affective sensory evaluation by consumers were performed.

2.3. Destructive quality measurements

For the quality measurements, a total of 174 out of the 3,564 samples were used for destructive quality measurements in the two weeks of the experiment. They were randomly picked from all the boxes in all the cultivars equally. In the first week, 90 samples were analysed, and in the second week, 84 samples. Each of these 174 samples was cut in two halves: one half was used for destructive quality assessments, while the remaining half was used for sensory evaluation. First, single fruit cylinders (diameter 10 mm) were obtained per sample to measure the breaking force (N) of the flesh fruit -here referred to as bite- (50 N compression, 20 mm distance) measured by a texturometer (4301, Instron Engineering Corp., MA, USA). The remaining fruit flesh was blended individually, and the TSS (°Brix) was measured using a refractometer (Mettler Toledo, 4004 JK Tiel, Netherlands). TA (mmol H_3O^+ $100g^{-1}$) was measured with an automatic titrator (Mettler Toledo, 4004 JK Tiel, Netherlands) using 4 grams of the blended sample.

2.4. Affective sensory evaluation

The sensory acceptance evaluation involved 153 regular Dutch strawberry consumers, consisting of 109 women and 44 men aged 23 to 84. The evaluation was conducted at the Sensory Flavour Laboratory from the Business Unit Greenhouse Horticulture & Flower bulbs (Wageningen University & Research, Bleiswijk, The Netherlands), designed under the ISO 8589:2007 standard and at a constant temperature of 20 °C, following ISO 6658:2017 standard for sensory analysis.

During the sensory evaluation, 15 fruit samples were provided to each consumer for tasting, three pieces per cultivar, resulting in 2295 fruit samples. The consumers rated the samples on a hedonic scale ranging from 1 to 10, where 1 indicates an immense dislike/unacceptance, 5 indicates neither like nor dislike, and 10 indicates an extreme liking/acceptance. The participants were explained how to use the hedonic scale for rating before the evaluation. Samples were 3-digit coded and presented in a randomised order in a neutral environment. A red light was used to minimise the influence of colour on taste perception and water and salt-free crackers were offered to consumers for palate cleansing between samples.

2.5. Image processing and data extraction

The reflectance of the images was corrected using a white polytetrafluorethylene (PTFE) reference and dark references (image with the camera shutter closed). Then, as shown in Figure 1, the single strawberries were extracted from the images of the entire boxes by segmentation from the images using a pre-trained Detectron2 Fast R-CNN network. The Detectron2 framework is built on the Faster R-CNN architecture, which combines region proposal networks with CNNs for efficient object detection and segmentation. The model was previously developed for automatic object segmentation; it was trained on data from previous experiments to identify and segment various types of fruits in boxes. For this application, the model was fine-tuned to work specifically with strawberries, ensuring precise segmentation. Using this model, individual strawberries were extracted from images of entire boxes by applying segmentation masks generated by the network and saved.

Besides, the mean spectra of each strawberry was extracted, and Standard Normal Variate (SNV) pretreatment was applied to the spectra dataset for scatter correction. Using the extracted spectra, a spectral dataset was created to train the models. The spectra were also analyzed to identify spectral differences.

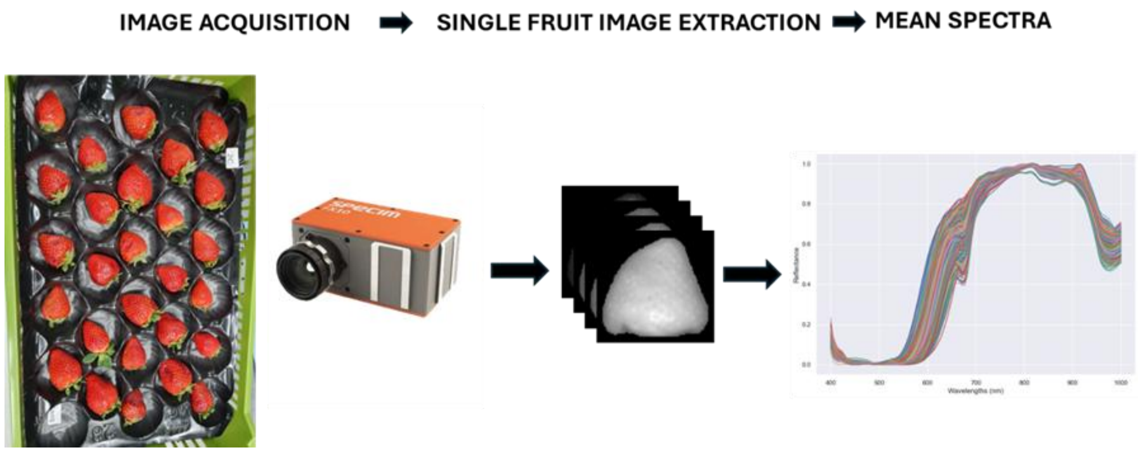


Figure 3. Data obtention and extraction.

2.6. Data analysis

2.6.1. Quality parameters and consumer acceptance

First, the normality of the quality parameters was assessed using the Kolmogorov-Smirnov Test. This test measures the maximum difference (D) between the empirical cumulative distribution function of the sample data and the theoretical distribution functions (Berger & Zhou, 2014)..

Subsequently, a multifactor analysis of variance (ANOVA) and Tukey's multiple range test were conducted to analyse the sensory and quality data, highlighting groups that exhibited significant differences (Prybutok, 2012).. Additionally, a Pearson correlation matrix was employed to explore potential relationships between quality parameters and to identify those most relevant to consumer acceptance. In this matrix, a value of 1 indicates a perfect positive correlation, while a value of -1 indicates a perfect negative correlation.

All analyses were performed using Statgraphics Centurion 18 (Statistical Graphics Corp., Rockville, MA, USA).

2.6.2. Model creation and optimisation

Three different models based on an SVM, XGBoost, and an MLP were trained using the mean spectra for each of the predicted variables (TSS, TA, bite, consumer acceptance and cultivar).

SVM is a supervised learning algorithm that works by finding an optimal hyperplane to separate data into different classes because of its effectiveness in high-dimensional spaces. XGBoost is a decision tree-based ensemble algorithm that iteratively builds models to minimise errors from previous iterations using gradient boosting. It can capture complex, non-linear relationships between variables while being computationally efficient. The MLP is a simple type of neural network with multiple layers of neurons, capable of modelling complex patterns in data (Müller & Guido, 2016).

Destructive measurements and consumer acceptance data were averaged per cultivar for model training rather than using individual sample values. Therefore, the average value of the parameters obtained from the destructively tested samples was applied to all samples of each cultivar. This was done to address the challenge of uneven data distribution while reducing noise and preventing overfitting during model training. The models maintained a more consistent input-output relationship by assigning the averaged values across cultivars, which is critical for effective training. Besides, predicting consumer liking, a subjective and complex attribute, poses a significant challenge for machine learning models. Averaging the data per cultivar reduces variability and helps capture the general trends associated with each cultivar, ultimately improving the capacity of the models to generalise (Aliferis & Simon, 2024).

Regarding train, validation and test datasets, 70 % of the samples were used for training the SVM and XGBoost models, and the remaining 30 %, unseen for the model during training, were used for testing. In the case of the MLP, 70 % of the data was used for training, 15 % for model validation while training, and the remaining 15 % unseen by the model for testing the model performance. The validation set in the neural networks helps fine-tune the model's hyperparameters and prevents overfitting by providing an independent dataset for evaluating the model's performance during training (Chollet, 2021)

The most suitable hyperparameters of the SVM and XGBoost were optimised to predict with the highest accuracy and lowest mean squared error (MSE) using Scikitlearn's RandomizedsearchCV and a 10-fold cross-validation. RandomizedSearchCV is a hyperparameter tuning method that randomly selects a fixed number of parameter combinations instead of testing all possible ones. It is useful when some parameters have little impact, as in the case of SVM or XGBoost, as it can find good configurations without exhaustive testing (Müller & Guido, 2016).

In the case of the MLP, Scikitlearn's GridSearchCV was used to optimise the number of layers, the best number of neurons on each layer, and the best activation function by trying different combinations of neurons and layers. A different MLP model was created to obtain the best results for every case. All MLP models were trained with the Adam optimiser. The Adam optimiser adjusts the learning rate for each parameter individually using momentum (past gradients) and adaptive scaling (past squared gradients). This helps neural networks converge faster and handle noisy data more effectively than standard gradient descent (Kingma & Ba, 2014). Early stopping in 100 epochs was used for training. Table 2 shows the hyperparameters selected for the regression and classification models.

Table 2. Hyperparameters selected for prediction and classification.

	Regression model hyperparameters					
	SVM		XGBoost		MLP	
TSS	Kernel:	Poly*	Min_child_weight:	100	Layers:	4
	Gamma:	Scale	Max_depth:	6	Neurons per layer:	75, 40, 5, 1
	Class weight:	Balanced	Learning_rate:	0.05	Activation function:	Softplus
	C:	48241.1	Gamma:	0.1	Last layer activation function:	Linear
			Colsample_bytree:	0.4	Loss function:	MSE
TA	Kernel:	Poly*	Min_child_weight:	100	Layers:	5
	Gamma:	Scale	Max_depth:	9	Neurons per layer:	200, 135, 70, 5, 1
	Class weight:	Balanced	Learning_rate:	0.1	Activation function:	Softplus
	C:	7840.3	Gamma:	0.3	Last layer activation function:	Linear
			Colsample_bytree:	0.5	Loss function:	MSE
Bite	Kernel:	Poly*	Min_child_weight:	100	Layers:	4
	Gamma:	Scale	Max_depth:	6	Neurons per layer:	200, 103, 15, 1
	Class weight:	Balanced	Learning_rate:	0.2	Activation function:	Softplus
	C:	7800.2	Gamma:	0.4	Last layer activation function:	Linear
			Colsample_bytree:	0.4	Loss function:	MSE
Consumer acceptance	Kernel:	Poly*	Min_child_weight:	1	Layers:	4
	Gamma:	Scale	Max_depth:	6	Neurons per layer:	75, 40, 5, 1
	Class weight:	Balanced	Learning_rate:	0.2	Activation function:	Softplus
	C:	78428.2	Gamma:	0.3	Last layer activation function:	Softplus
			Colsample_bytree:	0.5	Loss function:	MSE
Cultivar	Kernel:	RBF **	Min_child_weight:	3	Layers:	4
	Gamma:	10	Max_depth:	15	Neurons per layer:	200, 115, 30, 17
	Class weight:	Balanced	Learning_rate:	0.2	Activation function:	Relu
	C:	4.94	n_estimators:	100	Last layer activation function:	Softmax
			Colsample_bytree:	0.7	Loss function:	Categorical crossentropy

*Poly = polynomial of degree 3; ** RBF = Radial Basis Function.

The performance of the regression models was evaluated using the coefficient of determination (R^2), the mean absolute error (MAE), and the root mean squared error (RMSE). R^2 measures the proportion of the variance in the dependent variable explained by the independent variable in a regression model, and MAE measures the average absolute difference between the predicted and actual values. In contrast, RMSE measures the square root of the average of the squared differences between these predicted and observed values.

For classification models, the F1 score was used. F1 measures accuracy, combining precision and recall. Precision tells about how reliable the optimistic predictions of a model are, whereas recall or sensitivity measures how well the model identifies all positive values. Equation 1 shows how the F1 score was calculated, where TP means true positives, FP false positives, and FN false negatives.

$$F1\ Score = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} = \frac{TP}{TP + \frac{1}{2} * (FP + FN)}$$

Equation 1. F1 score calculation

Model training and data analysis were carried out using Python 3.9 with scikitlearn library, a 14-core processor (Intel i9 12th gen; Intel Inc., Santa Clara, CA, USA), 64 GB of RAM DDR4, and an 8 GB GPU (NVIDIA GeForce RTX 3070Ti GPU; Nvidia Inc., Santa Clara, CA, USA).

3. Results and discussion

3.1. Quality parameters and affective sensory evaluation

Table 3 summarises the average values for quality parameters and consumer scores across the seventeen strawberry cultivars, along with their standard deviations. Significant differences were observed among the cultivars for all measured attributes. For instance, cultivar V13 had the lowest TSS value (6.78 °Brix), while cultivar V15 exhibited the highest (10.29 °Brix). Regarding titratable acidity (TA), FS2004 displayed the lowest value (7.96 mmol $H_3O^+ \cdot 100g^{-1}$), whereas Calinda had the highest TA (14.36 mmol $H_3O^+ \cdot 100g^{-1}$). Bite force also varied substantially, with FS2004 showing the softest texture (0.29 N) and FE2132 being the firmest (1.51 N). In terms of sweetness-to-acidity balance, cultivar Verdi achieved the highest TSS/TA ratio (1.16), while FS2004 had the lowest (0.59). Similar TSS values were observed in strawberries by other authors, such as Saad et al., (2022), who found average values of 8.15 °Brix after evaluating three different cultivars (Festival, Gafiota and Florida).

Consumer acceptance scores also differed significantly among cultivars, reflecting varying levels of appeal. Cultivars FE2131, Jenkka, V15, and Sonata received the highest consumer ratings, with scores ranging from 7.2 to 7.4, indicating moderate to strong preference. These cultivars likely stood out due to their favourable balance of sweetness and acidity. For example, Jenkka and Sonata combined relatively high TSS values (8.51 °Brix and 10.00 °Brix, respectively) with balanced TA levels (9.56 mmol $H_3O^+ \cdot 100g^{-1}$ and 10.68 mmol $H_3O^+ \cdot 100g^{-1}$), contributing to their appealing taste profiles.

In contrast, cultivars FS2004 and FS2108 obtained the lowest consumer scores (5.5), indicating only slight appreciation or neutrality towards these varieties. The lower TSS value of FS2004 (8.48 °Brix) may have contributed to its diminished sweetness and overall attractiveness to consumers.

These findings highlight not only the diversity in quality attributes among strawberry cultivars but also their impact on consumer preferences, underscoring the importance of achieving an optimal balance of sweetness, acidity, and texture to enhance market acceptance.

Table 3. Results of the quality parameters analysis and consumer acceptance scores for each cultivar.

Cultivar	TSS (°Brix)	TA (mmol H_3O^+ 100g^{-1})	TSS/TA	Bite (N)	Consumer scores
	***	***	***	***	***
S2019-0081	8.29 ± 1.51 $^{c-e}$	8.71 ± 0.62 ^{g,h}	0.95 ± 0.18 ^{b,c}	0.99 ± 0.26 ^{c,d}	5.9 ± 2.0 ^{e}
FS2004	8.48 ± 0.44 $^{b-f}$	14.36 ± 0.75 ^{a,b}	0.59 ± 0.01 ^{h}	0.29 ± 0.08 ^{g}	5.5 ± 1.9 ^{f}
FS2108	9.01 ± 2.00 $^{b-d}$	10.93 ± 1.42 ^{e}	0.77 ± 0.12 $^{d-g}$	1.37 ± 0.71 ^{a,b}	6.6 ± 1.6 ^{c,d}
Balada	9.43 ± 1.64 $^{a-c}$	12.07 ± 1.45 ^{c,d}	0.78 ± 0.11 $^{d-g}$	0.53 ± 0.15 ^{g}	6.4 ± 1.7 ^{d}
Cadenza	7.83 ± 1.14 ^{e,f}	10.31 ± 1.11 ^{e,f}	0.76 ± 0.08 $^{e-h}$	0.92 ± 0.32 $^{c-f}$	6.5 ± 1.5 ^{c,d}
Calinda	9.76 ± 1.47 ^{a,b}	14.24 ± 1.40 ^{a}	0.69 ± 0.19 ^{g,h}	0.96 ± 0.36 $^{c-e}$	6.4 ± 1.8 ^{d}
FF22S35	8.03 ± 1.02 $^{d-f}$	9.44 ± 1.74 ^{f,g}	0.85 ± 0.08 $^{c-e}$	0.77 ± 0.33 $^{c-g}$	5.8 ± 1.8 ^{e,f}
FF22S36	10.27 ± 1.11 ^{a}	12.88 ± 0.68 ^{b,c}	0.80 ± 0.09 $^{d-g}$	0.62 ± 0.22 $^{d-g}$	6.4 ± 1.8 ^{d}
Fandango	9.68 ± 1.53 ^{a,b}	11.02 ± 1.09 ^{d,e}	0.88 ± 0.14 $^{c-e}$	0.75 ± 0.27 $^{c-g}$	6.6 ± 1.7 ^{c,d}
FE2131	8.29 ± 1.00 $^{c-e}$	7.96 ± 0.82 ^{h}	1.04 ± 0.06 ^{b}	0.92 ± 0.45 $^{c-}$f	7.3 ± 1.5 ^{a}
FE2132	8.96 ± 1.36 $^{b-d}$	12.62 ± 1.60 ^{c}	0.71 ± 0.17 ^{g,h}	1.51 ± 0.67 ^{a}	6.9 ± 1.3 ^{b,c}
FE2133	9.72 ± 0.78 ^{a,b}	12.60 ± 0.95 ^{c}	0.82 ± 0.16 $^{d-f}$	0.75 ± 0.30 $^{c-g}$	5.5 ± 1.9 ^{f}
V13	6.78 ± 1.45 ^{f}	8.12 ± 0.74 ^{h}	0.83 ± 0.16 $^{c-f}$	1.47 ± 0.61 ^{a}	5.7 ± 1.9 ^{e,f}
Jenkka	8.51 ± 1.19 $^{c-e}$	9.56 ± 0.88 ^{f,g}	0.89 ± 0.08 ^{c,d}	0.64 ± 0.22 $^{e-g}$	7.3 ± 1.7 ^{a,b}
V15	10.29 ± 1.04 ^{a}	13.93 ± 0.72 ^{a,b}	0.74 ± 0.14 $^{f-h}$	1.07 ± 0.26 ^{b,c}	7.2 ± 1.7 ^{a,b}
Sonata	10.00 ± 0.74 ^{a,b}	10.68 ± 1.42 ^{e}	0.94 ± 0.18 ^{b,c}	0.61 ± 0.24 ^{f,g}	7.4 ± 1.5 ^{a}
Verdi	10.24 ± 1.50 ^{a}	8.81 ± 0.74 ^{g,h}	1.16 ± 0.18 ^{a}	0.94 ± 0.46 $^{c-e}$	6.9 ± 1.8 ^{b,c}

Mean value ± standard deviation. Values followed by the same letter within the same column were not significantly different according to Tukey's least significant difference test. *** Significant at $p < 0.001$. TSS = total soluble solids; TA = titratable acidity.

The relationship between quality parameters and consumer acceptance provides valuable insights into the factors influencing consumer preferences. The ANOVA and Tukey tests revealed that TSS, TA, and bite did not independently affect acceptance. To further explore these relationships, a Pearson correlation matrix (Figure 2) was used to evaluate their interconnections. The strongest positive correlations were observed between TSS and TA (0.41), TSS and the TSS/TA ratio (0.45), and consumer acceptance with the TSS/TA ratio (0.37). Importantly, the TSS/TA ratio demonstrated a stronger correlation with consumer acceptance than TSS or TA alone, highlighting its significance in influencing taste perception. Nevertheless, none of these correlations were particularly strong, indicating the complexity of consumer acceptance.

These findings suggested that no single parameter can fully explain consumer preferences; instead, a combination of multiple factors is necessary to better predict acceptance. This aligns with previous research by Ross et al. (2010), who found that TSS and TA alone did not strongly correlate with flavour intensity in sweet cherry fruit. The results highlight the intricate nature of fruit taste and acceptance, where sweetness, acidity, and texture interact to influence consumer perception. The variability in consumer scores across cultivars (Table 3) further supports this complexity, as cultivars with balanced TSS and TA values, such as Jenkka and Sonata, achieved higher acceptance ratings compared to those with extreme values or imbalances like FS2004. This reinforces the importance of considering multiple quality parameters when evaluating fruit for market preferences.

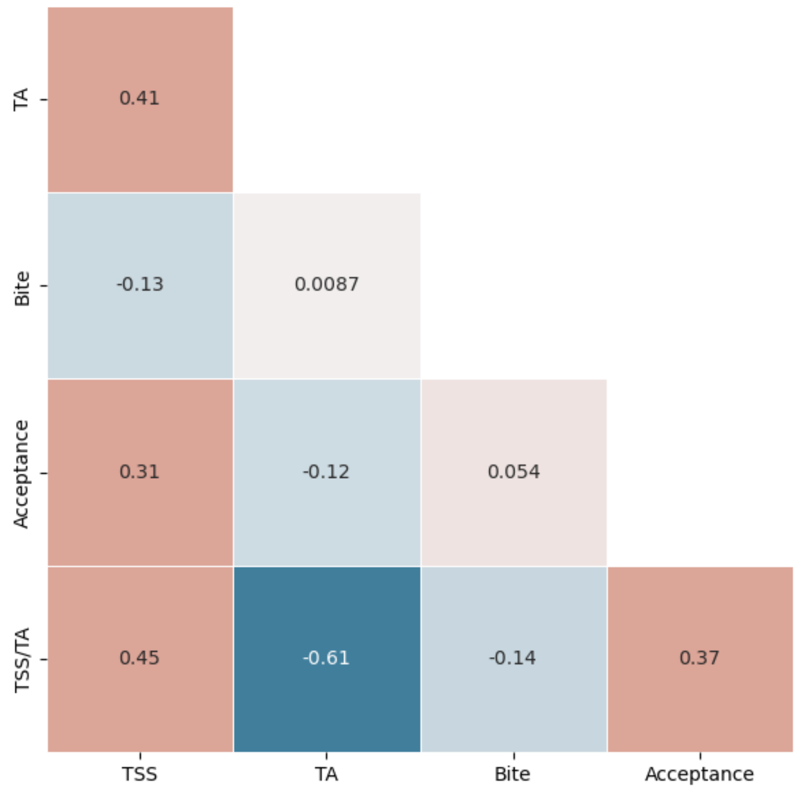


Figure 2. Pearson correlation matrix between the destructive measurements and consumer acceptance.

3.2. Spectral analysis

Figure 3 illustrates the mean reflectance spectra for each of the seventeen strawberry cultivars across the 400–1000 nm range, highlighting key differences in the Vis and NIR regions. These spectral variations are linked to important quality parameters such as colour, sugar content, water content, and structural integrity.

In the Vis range (400–700 nm), notable differences are observed between 400–450 nm, associated with carotenoid content (Wellburn, 1994). Variations between 530–750 nm reflect anthocyanin content around 535 nm (Lleó et al., 2011) and chlorophyll content near 680 nm (Wellburn, 1994). These pigments influence fruit colour, a critical factor for consumer acceptance. For example, cultivars with higher reflectance in these regions may appear lighter or redder, which can affect visual appeal.

In the NIR region (700–1000 nm), significant differences occur between 850–1000 nm. This range provides information on sugar content around 840 nm (ElMasry et al., 2007) and water content near 980 nm (Büning-Pfaue, 2003). Variations in this region are likely influenced by ripeness, texture, and cultivation practices. For instance, Jenkka exhibits one of the lowest reflectance values near 970 nm, suggesting higher free water content due to cell wall degradation (Tessmer et al., 2016), consistent with its low bite force value of 0.64 N. In contrast, FS2108 shows the highest reflectance at this wavelength, indicating lower free water content and a more stable structure, aligning with its higher bite force value of 1.37 N (Table 3).

The spectral data in Figure 3 provides additional insights into the relationships between quality parameters and consumer acceptance, complementing the findings from the Pearson correlation matrix (Figure 2) and the quality parameter table (Figure 1). The spectral reflectance curves reveal subtle differences among cultivars, particularly in the region around 970 nm, which is associated with water absorption. Higher water content, inferred from lower reflectance in this region, may contribute to juiciness, a trait often appreciated by consumers. For example, Sonata, which achieved one of the highest consumer scores (7.4), exhibits a balanced reflectance profile near 970 nm, aligning with its high TSS value (10.00 °Brix) and moderate TA (10.68 mmol $H_3O^+ \cdot 100g^{-1}$). This balance likely enhances its sweetness and overall flavour profile.

Similarly, cultivars like FE2131 and Jenkka, which also scored highly in consumer acceptance (7.3), demonstrate favourable combinations of TSS and TA values (e.g., Jenkka: TSS = 8.51 °Brix and TA = 9.56 mmol $H_3O^+ \cdot 100g^{-1}$), supported by spectral profiles indicative of balanced water content and internal composition. These findings align with the Pearson matrix results, where consumer acceptance showed a stronger correlation with the TSS/TA ratio (0.37) than with TSS or TA alone, highlighting the importance of sweetness-to-acidity balance in driving preference.

In contrast, cultivars such as FS2004 and FS2108, which received lower consumer scores of 5.5 and 6.6 respectively, highlight how imbalances in quality parameters can negatively affect acceptance. FS2004's low TSS value (8.48 °Brix) and low TSS/TA ratio (0.59) suggested insufficient sweetness relative to acidity, which may explain its reduced consumer scores. Additionally, FS2004's spectral profile showed higher reflectance near 970 nm, potentially indicating lower water content and reduced juiciness, further detracting from its sensory qualities.

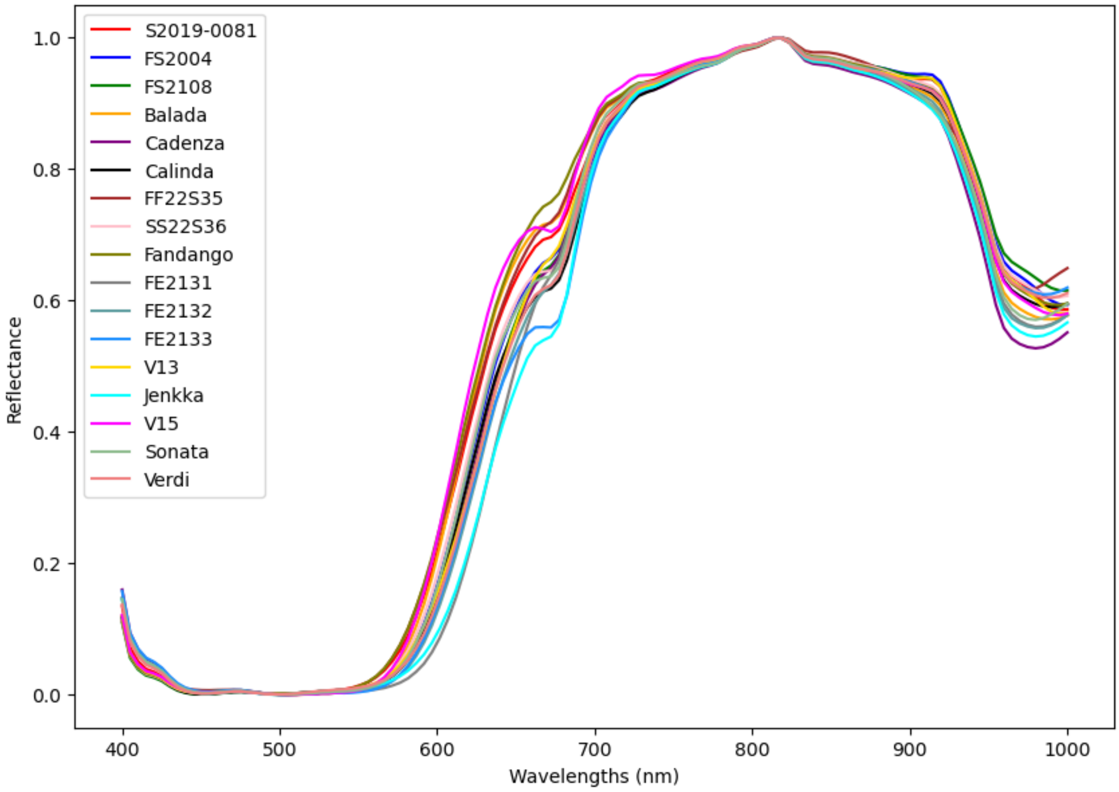


Figure 3. Mean spectra of each of the seventeen strawberry cultivars.

3.3. Prediction of quality parameters

Table 4 presents the performance metrics (R^2, MAE, and RMSE) of the models for predicting quality parameters in the test set. The MLP model consistently outperformed the others, achieving the highest precision across all parameters.

For TSS prediction, the MLP achieved an R^2 of 0.85, with an MAE of 0.63 and an RMSE of 0.80 °Brix, demonstrating its strong predictive capability. TA prediction followed closely, with the MLP obtaining an R^2 of 0.81, an MAE of 0.69, and an RMSE of 0.93 mmol $H_3O^+ \cdot 100g^{-1}$. Bite force prediction exhibited slightly lower precision compared to TSS and TA, but the MLP still performed best with an R^2 of 0.76, an MAE of 0.12, and an RMSE of 0.15 N.

The SVM model ranked second in performance, delivering results close to the MLP but slightly less precise across all parameters. In contrast, XGBoost exhibited the lowest predictive accuracy, with lower R^2 values and higher MAE and RMSE metrics for all quality parameters.

Table 4. TSS, TA and bite prediction test set prediction metrics

Model	TSS (°Brix)			TA (mmol H_3O^+ 100g^{-1})			Bite (N)		
	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE
SVM	0.81	0.68	0.87	0.73	0.86	1.07	0.67	0.14	0.18
XGBoost	0.71	0.76	1.04	0.69	0.86	1.16	0.62	0.15	0.20
MLP	0.85	0.63	0.80	0.81	0.69	0.93	0.76	0.12	0.15

R^2 = coefficient of determination; MAE = mean absolute error; RMSE = root mean squared error; TSS = total soluble solids; TA = titratable acidity.

These results are consistent with previous studies that have demonstrated the effectiveness of spectral imaging in predicting quality parameters for strawberries. For instance, ElMasry et al. (2007) reported a correlation coefficient (r) of 0.80 for predicting TSS using 77 samples from a single cultivar with a PLS model. Additionally, Nagata et al. (2005) found an r-value of 0.73 for TSS using five wavelengths based on 270 samples from one cultivar at three stages of ripeness. Weng et al. (2020) achieved R^2 values of 0.94 for TSS and 0.84 for TA by employing a PLS model with spectral and colour features from 120 samples of a single cultivar.

In the field of spectroscopy, Amodio et al. (2017) reported R^2 values of 0.85 for TSS and 0.58 for TA prediction using PLS models on 219 samples from three different farming systems and various maturity stages. Agulheiro-Santos et al. (2022), utilising 185 samples of one cultivar, and (Mancini et al., 2020), which examined five cultivars, documented R^2 values of 0.82 and 0.83 for TSS, respectively. (Mancini et al., 2020) also achieved an R^2 of 0.45 for predicting firmness. In contrast, Salata et al. (2022) reported slightly lower results for TSS with an R^2 of 0.75 but showed higher accuracy in TA prediction, achieving an R^2 of 0.75 using seven cultivars and Multi-Layer Perceptron (MLP) models. On the other hand, Sánchez et al. (2012), who examined nine cultivars at the commercial ripeness stage, obtained lower R^2 values of 0.35 for firmness, 0.61 for TSS, and 0.50 for TA using PLS models.

Compared to these studies, this work stands out by incorporating data from 3,564 samples of seventeen different cultivars that are commercially mature and sourced from two different origins. This broader dataset provides greater variability in quality attributes, enhancing the robustness and real-world applicability of the developed models.

3.4. Prediction of consumer acceptance

Table 5 presents the results of the models to predict estimated Dutch strawberry consumer acceptance of the test set. The highest performance was observed using the MLP model, with an R^2 of 0.78, MAE of 2.30, and RMSE of 2.92. Following the MLP, similar results were obtained with the SVM, achieving an R^2 of 0.73, MAE of 2.63, and RMSE of 3.26. XGBoost showed the least favourable performance, with an R^2 of 0.69, MAE of 2.75, and RMSE of 3.50.

This underscores the superior predictive capability of ANNs compared to other machine learning models.

Table 5. Consumer acceptance prediction on the test set.

Model	R^2	MAE	RMSE
SVM	0.73	2.63	3.26
XGBoost	0.69	2.75	3.50
MLP	0.78	2.30	2.92

R2 = coefficient of determination; MAE = mean absolute error; RMSE = root mean squared error.

To date, no studies have predicted the acceptability of strawberries using spectral imaging or spectroscopy. Research conducted on other fruits aimed at assessing consumer acceptance has often relied on indirect methods. Instead of predicting acceptance directly, these studies attempted to predict specific attributes that influence acceptance. For example, Harker et al. (2009) reviewed research on kiwifruit consumer acceptance. Rather than making direct predictions, they used spectral data to estimate dry matter content and then inferred consumer acceptance based on the correlation between dry matter and acceptance.

Similarly, Escribano et al. (2017) employed an approach similar to the present study, using spectroscopy to measure cherry properties. They divided each cherry into two halves: one for destructive measurements and the other for sensory evaluation. Involving 535 consumers, they had them rate the cherries on a hedonic scale from 0 to 10. They utilised PLS regression to predict the relationship between acceptance and quality measurements, ultimately finding that consumers preferred cherries with higher TSS content. However, like the previous study, they did not directly predict consumer acceptance from the spectral data; instead, they related it to the quality measurements predicted using PLS.

A few articles have directly assessed liking or consumer acceptance, such as Parpinello et al. (2013), who focused on table grapes. In their study, they used spectroscopy and PLS to create four clusters based on the predicted TSS content of the fruit. The grapes were then presented to consumers, who rated them on a 1-9 scale, which was later converted into three classes. This study achieved high classification accuracy, ranging from 75 % to 98 %, depending on the specific cluster class. However, these positive results may be partly attributed to the relatively small sample size (20 consumers) and the significant differences between the samples presented to the consumers.

3.5. Cultivar discrimination

Figure 4 shows the confusion matrices for the classification of samples into each cultivar class, along with the F1 scores for each model. High F1 scores were achieved across all models, with values exceeding 0.74. Among the models, the MLP outperformed the others, achieving the highest F1 score of 0.84. This demonstrates the superior capability of ANNs in handling complex classification tasks. The SVM model followed with an F1 score of 0.79, while XGBoost exhibited the lowest precision, achieving an F1 score of 0.74. These results align with the trends observed in the quality parameter predictions, where ANNs consistently delivered better performance compared to other machine learning models.

The confusion matrices illustrate the performance of each model in classifying the seventeen strawberry cultivars. They highlight key differences in model performance. While SVM achieved moderate accuracy overall, it exhibited higher misclassification rates for certain cultivars compared to MLP. XGBoost showed slightly lower accuracy than SVM and was particularly prone to errors in distinguishing between Dutch cultivars. The MLP model exhibited higher precision and recall across most cultivars, with fewer misclassifications compared to SVM and XGBoost. The superior performance of MLP can be attributed to its ability to learn non-linear patterns in the hyperspectral data, which are essential for discriminating between closely related classes.

For instance, cultivars such as Jenkka and V15 were accurately classified by the MLP, reflecting its ability to capture subtle spectral differences among cultivars.

Overall, despite all models achieving high correct classifications, cultivars 1 and 2 (S2019-0081 and FS2004), 3 and 8 (FS2108 and Fandango), and 4 and 9 (Balada and V13) showed higher levels of misclassification than the other varieties.

The confusion matrices revealed that all models struggled with certain cultivar pairs due to overlapping spectral signatures. For example, cultivars FF22S35 and FF22S36 were frequently misclassified across all models, highlighting the challenge of distinguishing between these closely related spectral profiles.

When comparing the confusion matrix results with the hyperspectral reflectance curves, Spanish cultivars tended to cluster together in their reflectance profiles, while Dutch cultivars also formed distinct but overlapping groups. In fact, Dutch cultivars such as FE2131, FE2132, and FE2133 showed some degree of confusion due to their similar genetic characteristics and growing conditions, since geographical origin and conditions play a key role in fruit quality, as proved by (Kitsios et al., 2025), who estimated the geographical origin of "Nova" mandarin juice based on multiple physicochemical and biochemical parameters.

These findings are consistent with previous studies that used spectroscopy for strawberry cultivar discrimination but often relied on fewer samples or cultivars. For example, Mancini et al., (2023) reported classification errors between 6.7 % and 13.3 % when classifying 216 fruits from four strawberry cultivars (Cristina, Romina, Sibilla, and Silvia) using PLS models. Similarly, Sánchez et al. (2012) achieved correct classification rates ranging from 57 % to 78 % for 300 samples from five cultivars (Antilla Fnm, Camarosa, Coral, Florida Fortuna, and Primoris Fnm) using a PLS model.

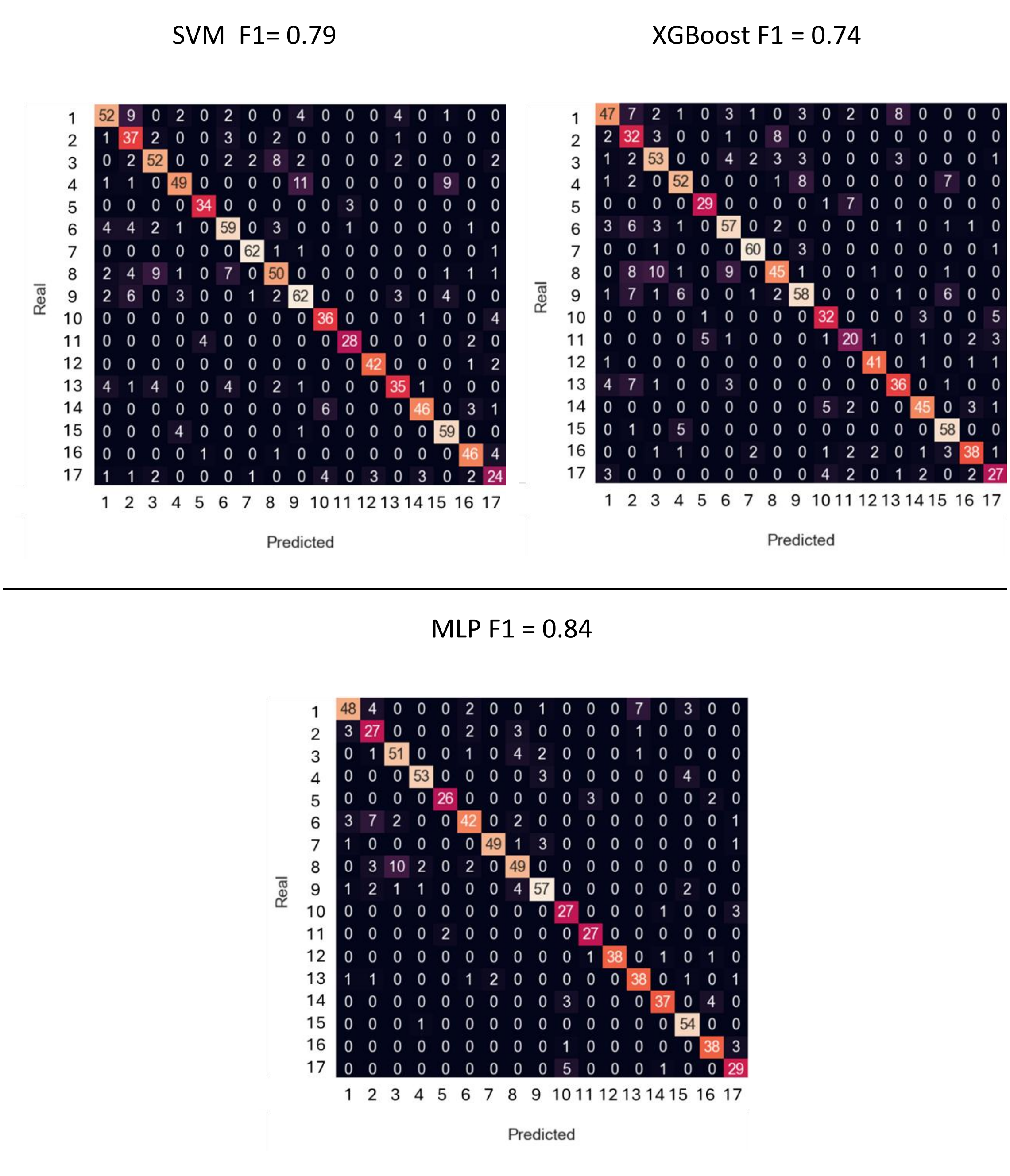


Figure 4. Results of the cultivar discrimination on the test set.

Overall, the aim of this study was to create general models that are applicable to the broader Dutch market, which primarily consists of Spanish and Dutch strawberries. Reducing the number of cultivars may enhance quality parameters prediction and improve variety classification or consumer acceptance by minimizing variability and simplifying predictions. However, it also poses a challenge, as the models may not be as general as intended. To achieve a better generalisation a large dataset is required to effectively train models such as ANNs, which are sensitive to the curse of dimensionality, especially in high-feature data like spectral imaging. This challenge occurs because having fewer samples in high-dimensional spaces results in sparse data, making pattern recognition difficult (Crespo Márquez, 2022). To improve the models, additional research should be conducted to expand the dataset.

4. Conclusions

This study evaluated the effectiveness of spectral imaging in predicting the quality of seventeen different strawberry cultivars and estimating Dutch consumer acceptance. First, a correlation analysis was conducted between quality parameters and consumer acceptability. Although quality metrics such as TSS, TA, and texture (bite) exhibited weak correlations with consumer acceptance, the TSS/TA ratio demonstrated the strongest correlation, underscoring the complexity of fruit taste preferences. Three machine learning techniques, SVM, XGBoost and MLP were evaluated for the prediction of quality, acceptability and cultivar. For predicting quality parameters, the MLP model showed the best performance, achieving R^2 values of 0.75 for TSS, 0.81 for TA, and 0.76 for bite prediction. The MLP model also produced the highest performance in predicting consumer acceptance with an R^2 of 0.78, and in cultivar classification, achieving an F1 score of 0.84.

These findings underscore the significance of spectral imaging in predicting the quality and consumer acceptance of seventeen different strawberry cultivars from two origins and seventeen different cultivars, a study that has not been conducted previously. This research serves as a foundational step for further investigations into this topic, which should include consumers from other countries, as fruit preferences can vary greatly across different cultures and regions. Future studies should explore a wider array of varieties and different seasons, along with the development of in-line or real-time prediction methods.

Acknowledgements

This work is part of the project PPS Smaakborging Groente en Fruit (Taste assurance of fruits and vegetables), partly funded by the Ministry of Economic Affairs, Secretariat Top Sector Horticulture and Starting Materials. The authors thank Fresh Forward, Brookberries, Onethird and Bakker Barendrecht for providing the fruit, support and technical supervision. Also, of MICIU AEI PID2023-150192OR-C31, C32 and C-33 with the support of FEDER funds and GVA-PROMETEO CIPROM/2021/014. Salvador Castillo Gironés thanks INIA for the FPI-INIA grant number PRE2020-094491, co-funded by EU EFS funds. Sandra Munera thanks the postdoctoral contract Juan de la Cierva-Formación (FJC2021-047786-I) co-funded by MCIN/AEI/10.13039/501100011033 and European Union NextGenerationEU/PRTR.

References

Agulheiro-Santos, A. C., Ricardo-Rodrigues, S., Laranjo, M., Melgão, C., & Velázquez, R. (2022). Non-destructive prediction of total soluble solids in strawberry using near infrared spectroscopy. *Journal of the Science of Food and Agriculture*, *102*(11), 4866–4872. https://doi.org/10.1002/JSFA.11849

Aliferis, C., & Simon, G. (2024). Overfitting, Underfitting and General Model Overconfidence and Under-Performance Pitfalls and Best Practices in Machine Learning and AI. In G. J. Simon & C. Aliferis (Eds.), *Artificial Intelligence and Machine Learning in Health Care and Medical Sciences: Best Practices and Pitfalls* (pp. 477–524). Springer International Publishing. https://doi.org/10.1007/978-3-031-39355-6_10

Amodio, M. L., Ceglie, F., Chaudhry, M. M. A., Piazzolla, F., & Colelli, G. (2017). Potential of NIR spectroscopy for predicting internal quality and discriminating among strawberry fruits from different production systems. *Postharvest Biology and Technology*, *125*, 112–121. https://doi.org/https://doi.org/10.1016/j.postharvbio.2016.11.013

Barbin, D. F., ElMasry, G., Sun, D.-W., & Allen, P. (2012). Predicting quality and sensory attributes of pork using near-infrared hyperspectral imaging. *Analytica Chimica Acta*, *719*, 30–42. https://doi.org/https://doi.org/10.1016/j.aca.2012.01.004

Berger, V. W., & Zhou, Y. (2014). Kolmogorov–Smirnov Test: Overview. *Wiley StatsRef: Statistics Reference Online*. https://doi.org/10.1002/9781118445112.STAT06558

Blas Saavedra, R., Cruz-Tirado, J. P., Figueroa-Avalos, H. M., Barbin, D. F., Amigo, J. M., & Siche, R. (2024). Prediction of physicochemical properties of cape gooseberry (Physalis peruviana L.) using near infrared hyperspectral imaging (NIR-HSI). *Journal of Food Engineering*, *371*, 111991. https://doi.org/https://doi.org/10.1016/j.jfoodeng.2024.111991

Büning-Pfaue, H. (2003). Analysis of water in food by near infrared spectroscopy. *Food Chemistry*, *82*(1), 107–115. https://doi.org/https://doi.org/10.1016/S0308-8146(02)00583-6

Castillo-Girones, S., Van Belleghem, R., Wouters, N., Munera, S., Blasco, J., & Saeys, W. (2024). Detection of subsurface bruises in plums using spectral imaging and deep learning with wavelength selection. *Postharvest Biology and Technology*, *207*, 112615. https://doi.org/https://doi.org/10.1016/j.postharvbio.2023.112615

Chollet, F. (2021). Deep Learning with Python, Second Edition. *Deep Learning with Python*. https://www.manning.com/books/deep-learning-with-python-second-edition

Crespo Márquez, A. (2022). The Curse of Dimensionality. *Springer Series in Reliability Engineering*, 67–86. https://doi.org/10.1007/978-3-030-97660-6_7

Elgendy, M., O'Reilly for Higher Education (Firm), & Safari, an O. M. Company. (2020). *Deep Learning for Vision Systems*. Manning Publications Co. https://books.google.com/books/about/Deep_Learning_for_Vision_Systems.html?hl=es&id=6gkLzAEACAAJ

ElMasry, G., Wang, N., ElSayed, A., & Ngadi, M. (2007). Hyperspectral imaging for nondestructive determination of some quality attributes for strawberry. *Journal of Food Engineering*, *81*(1), 98–107. https://doi.org/https://doi.org/10.1016/j.jfoodeng.2006.10.016

Escribano, S., Biasi, W. V, Lerud, R., Slaughter, D. C., & Mitcham, E. J. (2017). Non-destructive prediction of soluble solids and dry matter content using NIR spectroscopy and its relationship with sensory quality in sweet cherries. *Postharvest Biology and Technology*, *128*, 112–120. https://doi.org/https://doi.org/10.1016/j.postharvbio.2017.01.016

Eurostat. (2023). https://ec.europa.eu/eurostat/databrowser/view/apro_cpsh1__custom_8070269/default/table?lang=en

FAOSTAT. (2023). https://www.fao.org/faostat/en/#data/QCL

Gao, S., & Xu, J. (2022). Hyperspectral image information fusion-based detection of soluble solids content in red globe grapes. *Computers and Electronics in Agriculture*, *196*, 106822. https://doi.org/https://doi.org/10.1016/j.compag.2022.106822

Giampieri, F., Alvarez-Suarez, J. M., & Battino, M. (2014). Strawberry and Human Health: Effects beyond Antioxidant Activity. *Journal of Agricultural and Food Chemistry*, *62*(18), 3867–3876. https://doi.org/10.1021/jf405455n

Harker, F. R., Carr, B. T., Lenjo, M., MacRae, E. A., Wismer, W. V, Marsh, K. B., Williams, M., White, A., Lund, C. M., Walker, S. B., Gunson, F. A., & Pereira, R. B. (2009). Consumer liking for kiwifruit flavour: A meta-analysis of five studies on fruit quality. *Food Quality and Preference*, *20*(1), 30–41. https://doi.org/https://doi.org/10.1016/j.foodqual.2008.07.001

Kingma, D. P., & Ba, J. L. (2014). Adam: A Method for Stochastic Optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. https://arxiv.org/abs/1412.6980v9

Kitsios, A. P. M., Lazaridis, D. G., Koutoulis, A. S., Karabagias, V. K., Andritsos, N. D., & Karabagias, I. K. (2025). Geographical origin estimation of miniaturized samples of "Nova" mandarin juice based on multiple physicochemical and biochemical parameters conjointly with bootstrapping. *European Food Research and Technology*, 1–16. https://doi.org/10.1007/S00217-025-04684-3/METRICS

Lleó, L., Roger, J. M., Herrero-Langreo, A., Diezma-Iglesias, B., & Barreiro, P. (2011). Comparison of multispectral indexes extracted from hyperspectral images for the assessment of fruit ripening. *Journal of Food Engineering*, *104*(4), 612–620. https://doi.org/https://doi.org/10.1016/j.jfoodeng.2011.01.028

Mancini, M., Mazzoni, L., Gagliardi, F., Balducci, F., Duca, D., Toscano, G., Mezzetti, B., & Capocasa, F. (2020). Application of the Non-Destructive NIR Technique for the Evaluation of Strawberry Fruits Quality Parameters. *Foods 2020, Vol. 9, Page 441*, *9*(4), 441. https://doi.org/10.3390/FOODS9040441

Mancini, M., Mazzoni, L., Leoni, E., Tonanni, V., Gagliardi, F., Qaderi, R., Capocasa, F., Toscano, G., & Mezzetti, B. (2023). Application of Near Infrared Spectroscopy for the Rapid Assessment of Nutritional Quality of Different Strawberry Cultivars. *Foods*, *12*(17), 3253. https://doi.org/10.3390/FOODS12173253/S1

Müller, A. C., & Guido, S. (2016). *Introduction to machine learning with Python : a guide for data scientists*. 378. https://www.oreilly.com/library/view/introduction-to-machine/9781449369880/

Munera, S., Blasco, J., Amigo, J. M., Cubero, S., Talens, P., & Aleixos, N. (2019). Use of hyperspectral transmittance imaging to evaluate the internal quality of nectarines. *Biosystems Engineering*, *182*, 54–64. https://doi.org/https://doi.org/10.1016/j.biosystemseng.2019.04.001

Nagata, M., Tallada, J. G., Kobayashi, T., & Toyoda, H. (2005). *NIR Hyperspectral Imaging for Measurement of Internal Quality in Strawberries*. 1-. https://doi.org/10.13031/2013.19105

Parpinello, G. P., Nunziatini, G., Rombolà, A. D., Gottardi, F., & Versari, A. (2013). Relationship between sensory and NIR spectroscopy in consumer preference of table grape (cv Italia). *Postharvest Biology and Technology*, *83*, 47–53. https://doi.org/https://doi.org/10.1016/j.postharvbio.2013.03.013

Polder, G., Dieleman, J. A., Hageraats, S., & Meinen, E. (2024). Imaging spectroscopy for monitoring the crop status of tomato plants. *Computers and Electronics in Agriculture*, *216*, 108504. https://doi.org/https://doi.org/10.1016/j.compag.2023.108504

Prybutok, V. R. (2012). An Introduction to Statistical Methods and Data Analysis. *Technometrics*, *31*, 389–390. https://doi.org/10.1080/00401706.1989.10488572

Pullanagari, R. R., & Li, M. (2021). Uncertainty assessment for firmness and total soluble solids of sweet cherries using hyperspectral imaging and multivariate statistics. *Journal of Food Engineering*, *289*, 110177. https://doi.org/https://doi.org/10.1016/j.jfoodeng.2020.110177

Ross, C. F., Chauvin, M. A., & Whiting, M. (2010). Assignation of sweet cherry selections to 3 taste groupings based on perceived sweetness and sourness. *Journal of Food Science*, *75*(1). https://doi.org/10.1111/J.1750-3841.2009.01410.X

Saad, A. G., Azam, M. M., & Amer, B. M. A. (2022). Quality Analysis Prediction and Discriminating Strawberry Maturity with a Hand-held Vis–NIR Spectrometer. *Food Analytical Methods*, *15*(3), 689–699. https://doi.org/10.1007/S12161-021-02166-2/FIGURES/8

Salata, A., Ramón, H. M., Pandino, G., Najda, A., Amoriello, T., Ciccoritti, R., & Ferrante, P. (2022). Prediction of Strawberries' Quality Parameters Using Artificial Neural Networks. *Agronomy 2022, Vol. 12, Page 963*, *12*(4), 963. https://doi.org/10.3390/AGRONOMY12040963

Sánchez, M.-T., De la Haba, M. J., Benítez-López, M., Fernández-Novales, J., Garrido-Varo, A., & Pérez-Marín, D. (2012). Non-destructive characterization and quality control of intact strawberries based on NIR spectral data. *Journal of Food Engineering*, *110*(1), 102–108. https://doi.org/https://doi.org/10.1016/j.jfoodeng.2011.12.003

Su, Z., Zhang, C., Yan, T., Zhu, J., Zeng, Y., Lu, X., Gao, P., Feng, L., He, L., & Fan, L. (2021). Application of Hyperspectral Imaging for Maturity and Soluble Solids Content Determination of Strawberry With Deep Learning Approaches. *Frontiers in Plant Science*, *12*, 736334. https://doi.org/10.3389/FPLS.2021.736334/BIBTEX

Tessmer, M. A., Besada, C., Hernando, I., Appezzato-da-Glória, B., Quiles, A., & Salvador, A. (2016). Microstructural changes while persimmon fruits mature and ripen. Comparison between astringent and non-astringent cultivars. *Postharvest Biology and Technology*, *120*, 52–60. https://doi.org/https://doi.org/10.1016/j.postharvbio.2016.05.014

Ulrich, D., & Olbricht, K. (2016). A search for the ideal flavor of strawberry - Comparison of consumer acceptance and metabolite patterns in Fragaria × ananassa Duch. *Journal of Applied Botany and Food Quality*, *89*, 223–234. https://doi.org/10.5073/JABFQ.2016.089.029

Walsh, K. B., Blasco, J., Zude-Sasse, M., & Sun, X. (2020). Visible-NIR 'point' spectroscopy in postharvest fruit and vegetable assessment: The science behind three decades of commercial use. *Postharvest Biology and Technology*, *168*, 111246. https://doi.org/https://doi.org/10.1016/j.postharvbio.2020.111246

Wellburn, A. R. (1994). The Spectral Determination of Chlorophylls a and b, as well as Total Carotenoids, Using Various Solvents with Spectrophotometers of Different Resolution. *Journal of Plant Physiology*, *144*(3), 307–313. https://doi.org/https://doi.org/10.1016/S0176-1617(11)81192-2

Weng, S., Yu, S., Guo, B., Tang, P., & Liang, D. (2020). Non-Destructive Detection of Strawberry Quality Using Multi-Features of Hyperspectral Imaging and Multivariate Methods. *Sensors 2020, Vol. 20, Page 3074*, *20*(11), 3074. https://doi.org/10.3390/S20113074

Yu, Y., & Yao, M. (2022). A portable NIR system for nondestructive assessment of SSC and firmness of Nanguo pears. *LWT*, *167*, 113809. https://doi.org/https://doi.org/10.1016/j.lwt.2022.113809

CHAPTER II. Firmness classification in ‘Rojo Brillante’ persimmon fruit after cold storage using hyperspectral imaging and deep learning with variational autoencoders.

Firmness classification in 'Rojo Brillante' persimmon fruit after cold storage using hyperspectral imaging and deep learning with variational autoencoders

Salvador Castillo-Gironés[a], Ángel González[b], Sandra Munera[c], Marcelino Martínez-Sober[b], Sergio Cubero[a], Nariane Q. Vilhena[d], Ivan Blanco-Álvarez[b], Juan Gómez-Sanchís[b]

[a] Centro de Agroingeniería, Instituto Valenciano de Investigaciones Agrarias (IVIA), CV-315, km 10,7, 46113 Moncada (Valencia), Spain.

[b] Departamento de Ingeniería Electrónica, IDAL. Universitat de València, Av. Universitat, s/n, 46100 Burjassot (Valencia), Spain.

[c] Departamento de Ingeniería Gráfica. Universitat Politècnica de València. Camino de Vera, s/n, 46022 Valencia, Spain.

[d] Centro de Tecnología Postcosecha, Instituto Valenciano de Investigaciones Agrarias (IVIA), CV-315, km 10,7, 46113 Moncada (Valencia), Spain.

LWT. (Under review)

Abstract

Firmness is one of the most important quality parameters of 'Rojo Brillante' persimmon fruit, and it can be significantly affected during cold storage. Industry requirements demand 40 N at harvest for storage and 20 N for commercialisation. This study evaluated the use of hyperspectral imaging to determine the firmness of persimmons stored under industrial conditions. To achieve this, fruit samples were stored for up to three months under various industrial conditions to produce a wide range of firmness levels. Hyperspectral images of each sample fruit were captured within the 420–1010 nm range, and the firmness was measured by a texturometer. An innovative methodology was applied, combining a convolutional variational autoencoder (VAE) for feature extraction from firmness curves and images and clustering analysis to identify firmness levels. The clustering analysis identified three distinct firmness categories (hard, medium and soft) corresponding to the requirements of the industry. Using six wavelengths, a support vector machine (SVM) classification model accurately assigned 81.3% of fruit samples to the corresponding categories, thus paving the way for the utilisation of in-line detection using multispectral images. This approach demonstrates the potential of combining hyperspectral imaging and advanced machine learning techniques for non-destructive firmness prediction in 'Rojo Brillante' persimmon fruit quality assessment.

Keywords: fruit; persimmon; firmness; deep learning; spectral imaging

1. Introduction

Persimmon (*Diospyros kaki*) is a fruit with a rich cultural history in Asian countries such as China, Korea, and Japan and has become an important agricultural product worldwide (Yamada et al., 2012). China is the largest persimmon producer in the world, with a production of 3.4 million tonnes in 2024. However, Spain has now become the second largest producer in the world, with 404,131 tonnes (FAOSTAT, 2023), which highlights the economic and agricultural importance of persimmon production in this country.

'Rojo Brillante' persimmon is one of the most important cultivars produced in Spain and is predominantly grown in the Valencian region under the designation of origin 'Kaki Ribera del Xúquer' (Besada et al., 2008). Traditionally, this cultivar was consumed locally, as its astringency made it suitable for consumption only when overripe, when the tannins present in the pulp become insoluble, and astringency is no longer present (Tessmer et al., 2016). Nevertheless, the introduction of a postharvest treatment by exposing the fruit to high concentrations of CO_2 allows the removal of fruit astringency while maintaining a high level of flesh firmness. This development has resulted in a significant increase in the cultivation of this product in recent decades.

After the astringency removal treatment, inadequate storage conditions can adversely affect the quality of 'Rojo Brillante' persimmons, leading to alterations in flesh firmness due to temperature or storage duration. Industry standards recommend a firmness of 40 N at harvest to ensure the fruit's suitability for storage under industrial conditions, with a minimum of 20 N required for commercialisation after storage. Persimmon with lower than 20 N could result in rejection from consumers who prefer fruits with high firmness (Vilhena et al., 2022). Consequently, the development of accurate tools to assess the flesh firmness of these fruits before they enter the market is of great interest to packing houses. Such tools would help prevent consumer rejection and ultimately have a positive impact on future sales.

Destructive firmness assessment using texturometers is not feasible as it requires destruction of the sample. Besides, acoustic sensors are also helpful and not expensive in firmness assessment (Ramos et al., 2005), as in apples (Fathizadeh et al., 2020) or in kiwifruit (Tian et al., 2022). However, their industrial use for in-line detection is still challenging due to limitations such as sensitivity to environmental noise, challenges in interpreting complex signals for certain materials, and potential difficulty in maintaining consistent sensor-to-surface contact during high-speed production processes. (Lesage & Destain, 1996). CT scanning can provide valuable information on the internal structure of fruits and vegetables. However, its low speed and high costs make it unsuitable for industrial rapid firmness assessment (Nugraha et al., 2019). Most studies focus on microstructure and porosity, as proven in cranberries (Liu et al., 2025) or pears (Cui et al., 2021). Also, on pears, (Joseph et al., 2023) used CT scans and spectroscopy, finding that the porosity identified by CT scans correlated with the scattering coefficient at 760 nm, demonstrating that spectroscopy can be an alternative to CT scans. Hyperspectral imaging and spectroscopy have emerged as a proven technology for predicting internal quality attributes in a non-destructive manner (Castillo-Girones et al., 2024; Munera et al., 2018). Spectral technologies, including hyperspectral and multispectral imaging, are more suitable and cost-effective for the persimmon industry to assess fruit firmness non-destructively. In particular, multispectral cameras are an effective option as they can evaluate fruit quality using affordable cameras that capture several specific wavelengths at a fast speed allowing in-line detection (Gaikwad & Tidke, 2022). Regarding hyperspectral imaging assessment and spectroscopy, several studies conducted across different fruit including apples (Mendoza et al., 2011; Peng & Lu, 2008), pears (Li et al., 2016; Yu et al., 2018), peaches (Lu & Peng, 2006; Xuan et al., 2022), bananas (Xie et al., 2018), sweet cherries (Pullanagari & Li, 2021), blueberries (Leiva-Valenzuela et al., 2013; Park et al., 2023) and grapes (Xu et al., 2022), have successfully demonstrated the effectiveness of this technology for firmness prediction. However, no studies have been found to predict ‘Rojo Brillante’ persimmon firmness using hyperspectral imaging.

The use of hyperspectral images, however, generates a substantial amount of data, posing challenges in terms of processing. To address this issue, wavelength selection is often employed to reduce data dimensionality, thus facilitating more efficient processing. In recent years, Autoencoders (AE) have garnered attention due to their ability to uncover non-linear relationships and extract essential features from data. The encoder component of an AE compresses the input data into a lower-dimensional representation, thereby reducing data complexity and size. This dimensionality reduction of all kinds of data allows for easier machine learning (ML) model training (Bank et al., 2021).

AEs offer several advantages over traditional dimensionality reduction techniques like PCA and t-SNE. Unlike PCA, which is limited to linear transformations, VAEs can capture complex non-linear relationships in the data, making them suitable for high-dimensional datasets with intricate structures, such as hyperspectral images. VAEs outperform t-SNE and UMAP in preserving global geometry while still capturing local structures, which is crucial for maintaining meaningful data representations. For quality assessment of fruit, VAEs are advantageous because they can model subtle variations in fruit features (e.g., colour and texture) (Battey et al., 2021).

However, the application of AE has been limited, with only a few studies using them in research involving pears (Yu et al., 2018), grapes (Xu et al., 2022), kiwifruit (Torkashvand et al., 2017), blueberries (Park et al., 2023) and apples (Fathizadeh et al., 2021).

No articles were found regarding the use of AE for firmness analysis and prediction and its use for hyperspectral imaging feature extraction combination. In firmness curve analysis, traditional methods often focus on specific features like slope or peak force to assess firmness, as done in Pears by Yu et al. (2018), who used AEs with the image mean spectra. However, the use of AE offers a more comprehensive approach. By processing entire firmness curves, AE captures comprehensive curve characteristics, providing better insights while significantly reducing data dimensionality. Similarly, autoencoders are underutilised in hyperspectral imaging to retain essential features while reducing data complexity.

This study introduces an innovative methodology for predicting the firmness of the 'Rojo Brillante' persimmon, a variety that dominates European production and accounts for approximately 90% of Spain's output. The primary aim of this research was to evaluate the potential of hyperspectral imaging to determine persimmon firmness under these conditions. Another key aim of the study was to achieve data reduction by utilising only a few selected wavelengths and employing variational autoencoders for feature extraction. This approach enables the use of multispectral cameras for real-time applications, as the selection of fewer wavelengths minimises data acquisition requirements, and the autoencoder effectively reduces the dimensionality of the data, facilitating rapid and efficient predictions. The novelty of the approach lies in the use of a convolutional variational autoencoder (VAE) for extracting features from firmness curves and hyperspectral images. This technique is rarely applied in fruit studies and never in persimmon fruit, which can provide a successful solution for effectively managing high-dimensional data. Within this innovative approach, clustering methods such as K-means and hierarchical clustering (HC) were used to identify and analyse the different firmness levels of the fruit after storage. The final aim was to develop a predictive model using the support vector machine (SVM) algorithm to accurately classify the fruit into the identified firmness levels. SVMs are a better choice for classification tasks than Partial Least Squares or Decision Trees because they are robust to overfitting, handle small datasets effectively, and can model non-linear relationships using kernels (Aurélien Géron, 2019), ensuring reliable and accurate predictions for quality control.

This research fills a significant gap in the current literature and provides valuable insights for improving quality control processes in 'Rojo Brillante' persimmon production, emphasising the potential of combining hyperspectral imaging with advanced machine learning techniques.

2. Material and methods

2.1 Fruit samples

A total of 2380 persimmon fruits cv 'Rojo Brillante' were used in this experiment. Persimmon was obtained from a fruit cooperative (CANSO) located in L'Alcúdia, Valencia at commercial maturity and from different producers and locations in the 2022 season. The maturity was measured using the external colour index (CI), commonly employed for 'Rojo Brillante' (Salvador et al., 2007), with an average CI of 4.5. Then, the fruit was transported to the IVIA facilities in Moncada, Valencia, Spain, where it was stored for up to three months under varying conditions to observe changes in flesh firmness. The storage conditions included temperatures of 0 °C, 1 °C, and 5 °C, with 90% relative humidity, with and without 1-MCP treatment and storage times of 1, 2 and 3 months. The storage conditions simulate the conditions commonly used in the industry. This resulted in different combinations of storage conditions to obtain a wide range of firmness, reproducing the different conditions that may occur in the industry (Vilhena, 2022). 490 fruits were evaluated directly after harvest, and the remaining were stored to observe changes in flesh firmness and to obtain a significant degree of variability. Each month, a random set of 720 fruits was taken from each storage condition: with 1-MCP treatment (105 samples for each of the three temperatures) and without 1-MCP (105 samples for each of the three temperatures), and stored at room temperature (20 ºC) for 24 hours prior doing the measurements. Then, hyperspectral images of the fruit were acquired, and the flesh firmness was analysed using a texturometer.

2.2 Hyperspectral imaging acquisition and processing

Images were acquired using a hyperspectral system, as shown in Figure 1. The system is composed of a camera (CoolSNAP ES, Photometrics, AZ, USA), and two tuneable Vis and NIR filters (Varispec VIS-07 and NIR-07, Cambridge Research and Instrumentation, Inc., MA, USA) with a spectral range of 400-1050 nm, taking 1392×1040-pixel images with a spatial resolution of 0.14 mm/pixel every 10 nm. The range of 420-1010 nm was selected, excluding noisy bands at the beginning and the end, which resulted in 60 total wavelengths. The integration time of each band was calibrated to avoid saturated images and optimise the hyperspectral system's dynamic range.

Fruit samples were placed manually in a fruit holder, which elevated the fruit depending on its size, to obtain focused images. The fruits were illuminated using diffuse light produced by twelve 37 W halogen lamps (Eurostar IR Halogen MR16, Ushio America, Inc., CA, USA) powered by a 12 V direct current. Lamps were placed equidistantly inside a hemispherical aluminium diffuser painted white with a rough firmness on its inner surface to maximise reflectivity and minimise directional reflections, which might produce bright spots.

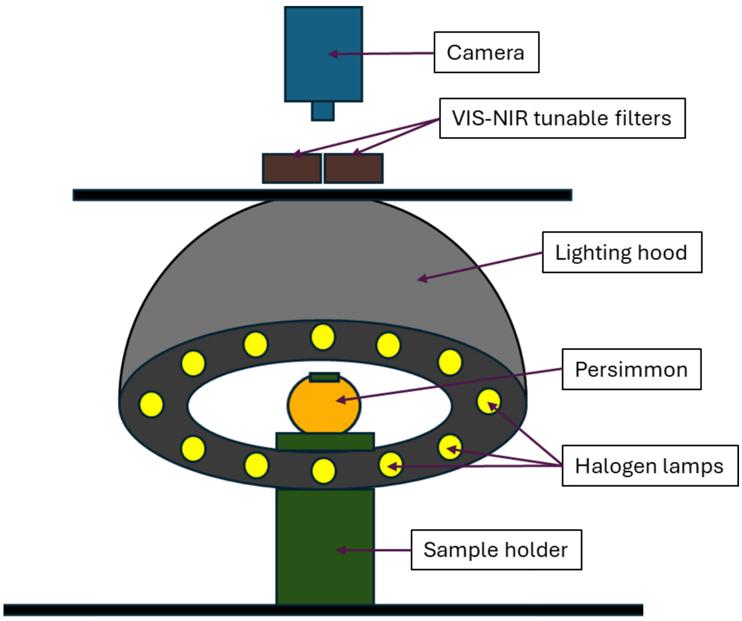


Figure 4. Hyperspectral imaging system

Equation 1 details the method for correcting images at every wavelength: Where $\rho^{Reference}(\lambda)$ is the reflectance value of the standard surface at a specific wavelength (99 % in the case of

Spectralon®, Labsphere, Inc, NH, USA), $Im_{Dark(\lambda)}$ and $Im_{White(\lambda)}$ are the dark and white refeences at that specific wavelength, and $Im_{Raw(\lambda)}$ is the raw image at that wavelength.

$$Im_{Refl} = \rho^{Reference}(\lambda)\frac{Im_{Raw(\lambda)} - Im_{Dark(\lambda)}}{Im_{White(\lambda)} - Im_{Dark(\lambda)}}$$

Equation 3. HSI radiometric calibration

The first step of the image processing was to adjust the size of the images to the size of the fruit, removing most of the background. To this end, the image at 850 nm was selected for visualisation, as it provided the highest contrast between the fruit and the background for image segmentation to remove the background and visualisation. After that, a morphological closure operation in all images was applied to fit the image to the size of the fruit. Each image was resized using zero padding from its centre to match the dimensions of the biggest image due to different fruit sizes. Lastly, the hyperspectral images were reduced to a final dimension of 711×866 pixels (length and width) to fit the model used. Figure 2 shows an example of the size reduction proposed. After that step, the size of the images was compressed by 60.8% without loss of information. Moreover, a min-max normalisation was performed to restrict the intensity of each pixel between 0 and 1.

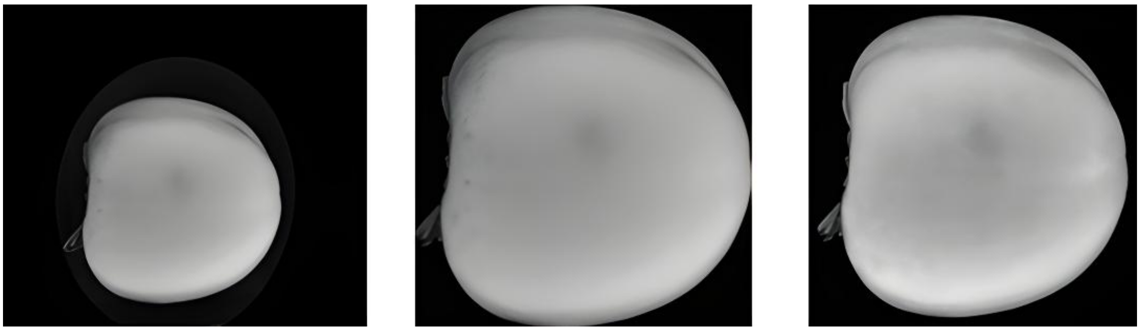


Figure 2. Example of the procedure for reducing the size of a hyperspectral image. The image was at 850 nm (left), and the same image was adjusted to the size of the fruit (middle) and the maximum size of any persimmon in the dataset after the zero-padding operation (right). The axes show the size of the image (length and width) in pixels, and the images are shown on grayscale.

The next step of the image processing was the reduction of the spectral dimensionality. To do so, a linear correlation dimensionality reduction method was used (Eid et al., 2013). This method analyses the dependencies and linear relationships between the different wavelengths in the dataset, removing the highly correlated ones. To apply this dimensionality reduction method, a single spectrum of each image was obtained, and the average intensity of all the pixels in each hyperspectral image was calculated. A dataset containing the average spectra of each image was obtained, the rows being the mean spectra of each fruit and the columns, the wavelengths. After that, the correlation between each pair of wavelengths was calculated. Finally, the method selected only the wavelengths that were less correlated to each other (correlation below 0.8). The threshold of 0.8 was chosen based on the visual inspection of the correlation plots (Figure 11 in the Results section 3.2), which shows the linear correlation and a scatter plot between each selected wavelength. Visually, the scatter plots indicate low correlation among the selected wavelengths, further justifying the chosen threshold after applying the size reduction preprocessing and wavelength selection.

After that, images containing only the selected wavelengths were created, and those resulting multispectral images were used for training the model.

2.3 Firmness analysis and curves processing

Immediately after image acquisition, the firmness of each sample fruit was measured using a universal testing machine (4301, Instron Engineering Corp., MA, USA) with an 8 mm puncture probe. The skin of the fruit was carefully removed (1 mm) with a knife, and the fruit was placed on the plate horizontally. The force required to break the flesh of the fruit (in N) was recorded in a load curve.

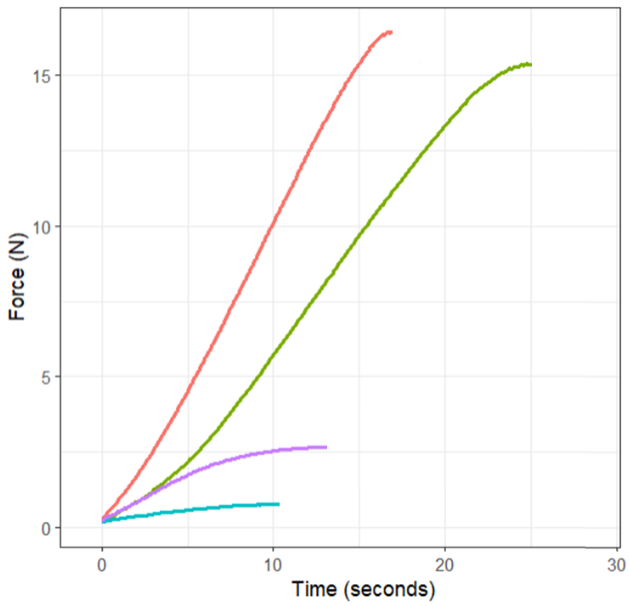


Figure 3. Examples of raw firmness curves of four persimmon samples.

As illustrated in Figure 3, the firmness curves exhibited an initial phase where the applied force increases until reaching a maximum force, the breaking point. To address these variations and prepare the data for analysis, the maximum peak of each curve was identified and extended to match the longest duration among all curves, which was 87 seconds force. This extension ensured that all curves could be uniformly analysed without altering their inherent information up to the breaking.

A Bayesian VAE proposed by Kingma and Welling (2013) was used to extract the most important features of the curves in a latent space. The reason behind analysing the complete curves is that when evaluating the firmness of a persimmon, it is imperative to analyse the entire force-displacement curve rather than relying solely on peak force or slope measurements. This comprehensive approach allows for a better understanding of the mechanical properties of the fruit, including firmness, elasticity, and viscosity throughout the deformation process. By considering the entire curve, the insights into the firmness characteristics of the persimmon fruit can be better studied. This approach analyses the complete curve with the VAE and ensures a more accurate representation of the firmness profile of the fruit.
The used VAE is a deep learning model capable of transforming an input vectorial space (in this case, firmness curves) into another one, called latent space, of a lower dimension than the original one, which follows a multivariate normal distribution. In this way, the features in the latent space are uncorrelated. This part is called the 'encoder'. A decompression process is performed from these latent components in which the input vectorial space is reconstructed. This part is called the 'decoder' (Huang et al., 2021).

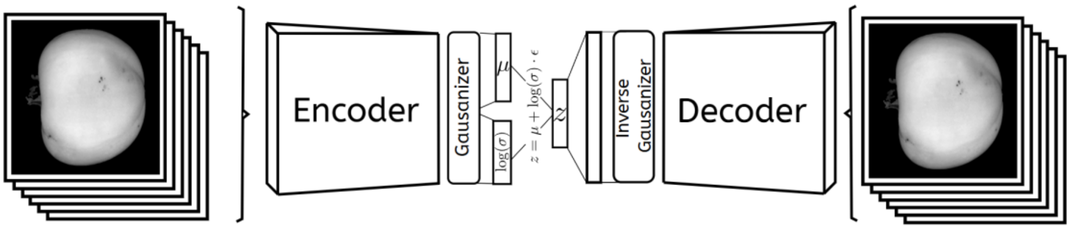


Figure 4. Generic VAE using, for example, hyperspectral images as input.

This type of model (Figure 4) learns by two sources of error: original and reconstructed data are compared to see how well it is capturing the information, and latent components and a theoretical multivariate normal distribution are compared to ensure these latent dimensions are uncorrelated. The parameters of the neural model are then updated accordingly. Repeating this process for a specified number of iterations will reduce global error to a point where the original and reconstructed images are closely related. At this final point, the autoencoder will compress the data in features of the latent space, obtaining the most valuable information (Rybkin et al., 2021).

To effectively analyse the firmness curves using the VAE, it was essential to standardise the input data by ensuring all inputs had a uniform shape and values scaled between 0 and 1. This process was crucial because each curve varied in duration until reaching the flesh breaking point, which depended on the specific structure of the measured fruit.

By adapting the curves in this manner, they were made compatible with the VAE, which is designed to extract key features from the data up to the point of highest force, thereby capturing the most critical characteristics for analysis. This process did not modify the essential information of the curves, as the data up to the breaking force remained intact.

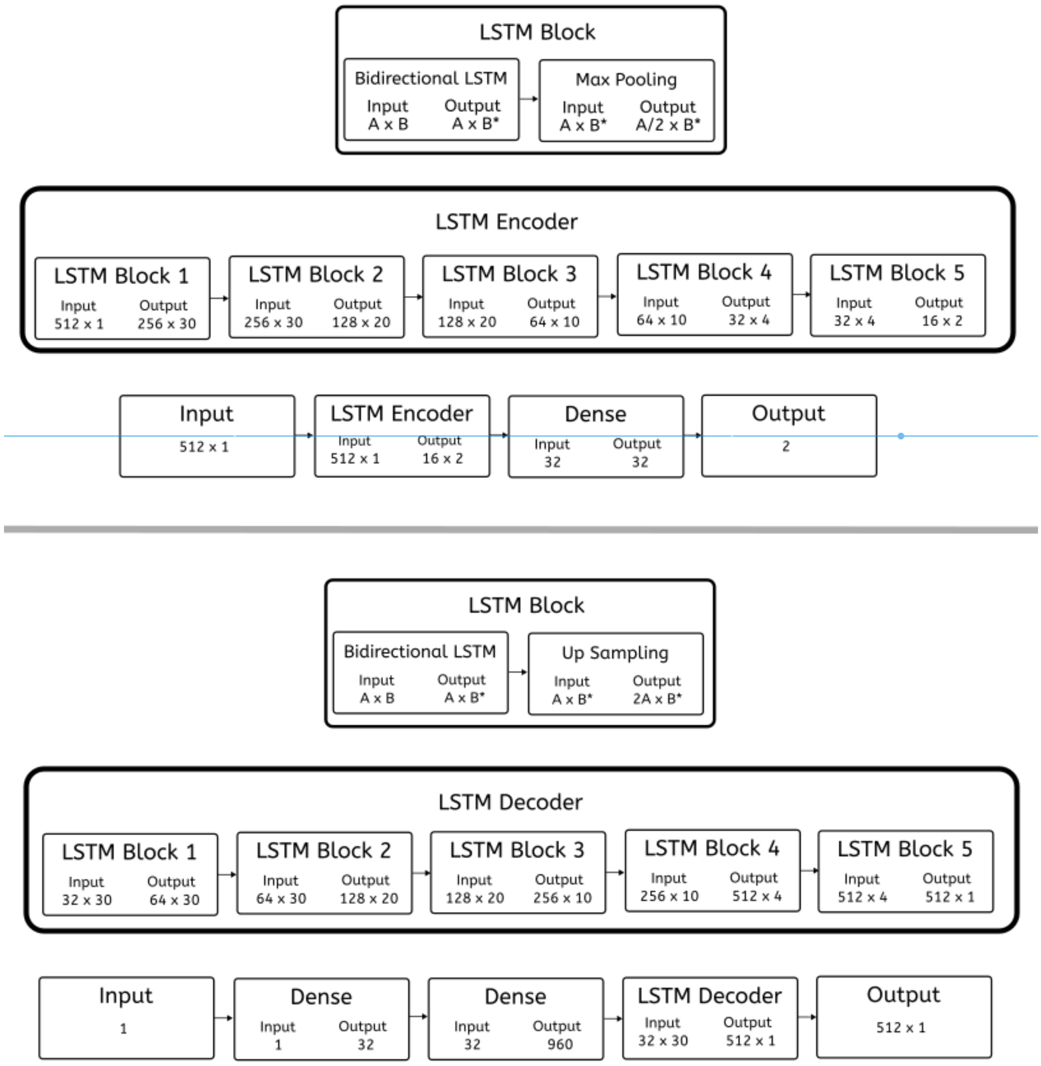


Figure 5. Structure of the VAE model used to extract features of the firmness curves. The top part of the image represents the encoder, and the bottom part represents the decoder.

Figure 5 shows the VAE structure used to extract features from the firmness curves. The encoder and decoder parts used to compress the curves were based on long short-term memory (LSTM) networks, which are layers used when one dimension is temporal. These layers can remember what has happened in specific moments in the past to make new predictions. In addition to this, bidirectional variants were used. This layer consists of one LSTM that learns data structures from left to right and another one which learns data structures from right to left. Then, the most important features were collected and used as input for the next layer. In this model, only one dimension was used to collect all the information. Therefore, the encoder had a final output of two neurons, and the decoder had only one. The encoder consisted of 7738 parameters, and the decoder consisted of 41,816.

Thus, with the help of the VAE, firmness curves were compressed to one single component, which means a reduction of 99.9 %.

2.4 Firmness clustering

Feature extraction from the firmness curves was performed using a variational autoencoder (VAE), with the extracted features subsequently used as inputs for two unsupervised clustering methods: K-means and Hierarchical Clustering (HC). These clustering techniques were applied to identify distinct groups of firmness among the persimmon samples. These clustering approaches were used for three main reasons: i) exploration and comparison: by applying multiple clustering techniques, the results obtained from each method were compared and assessed to determine which aligned best with the nature of the problem; ii) robustness and stability: random initialisation or sensitivity to specific parameters can sometimes influence a clustering method. By employing two methods, it becomes possible to verify whether the results are consistent and stable; iii) context and data complexity: some problems (datasets) may have characteristics that are better suited to a specific method.

On the one hand, K-means clustering (Shi et al., 2010) is based on separating the dataset into K clusters, minimising their intra-cluster distance (the distance between points in the same cluster), and maximising inter-cluster distance (the distance between different clusters). To obtain the optimal number of clusters, from 2 to 30 clusters were tested and their inertias (sum of the squared distance of samples to their cluster centroid) were plotted. The location of the elbow of this plot determined the number of clusters (Madhulatha, 2012).

On the other hand, HC (Patel et al., 2015) is based on merging clusters that are closest to each other. As it is an iterative process, distances of different clusters are tracked and represented. The best number of clusters is the one that, in the next iteration, increases the mean distance between clusters beyond 20%.

2.5 Firmness classification using multispectral images

Models were built from the measured data. For that, a holdout cross-validation method was used. Two random subsets were created for training and testing; 80% of the samples (1895 samples) were used for training and 20% for testing (485 samples). This separation was maintained in all the methods applied in this work.

As was carried out with the firmness curves, a VAE was used to extract the most essential features, in this case, of the multispectral images. Once the features from latent space had been computed from the multispectral images (the most relevant bands of the hyperspectral images were chosen in the preprocessing stage) using the VAE encoder and each sample had been assigned to a firmness group characterised by its firmness curve, it was possible to perform image classification using an SVM to classify them into the different firmness classes.

Figure 6 shows the structure of the VAE trained to compress the multispectral images. The encoder and decoder parts used were convolutional: several specified kernels pass through the entire image and retain all the spatial information of the image. For these convolutional layers, 3×3 kernels and parametric rectified linear unit (PReLU) activation functions were used in all cases. The first three convolutional layers of the encoder had a stride hyperparameter of two units (the output of these convolutional layers was half the input). The encoder consisted of 471,624 parameters, whereas the decoder consisted of 657,392.

With the help of the VAE, the multispectral images were compressed from 72,000,000 values of pixel reflectance (dimensions) to 7 features in the latent space, a total reduction of 99.9%. The number of features was chosen empirically.

The VAE model consisted of two main parts: the encoder and the decoder. The encoder output has 14 neurons, where 7 represent the average value (μ) of a hidden vector, and the other 7 represent the variation (σ) of that hidden vector. A technique called the 'reparameterisation trick' is used to connect the encoder and decoder. It involves a simple reparameterisation that consists of creating a new vector called z using the values of μ and ϵ. The formula is z = μ + ϵ. Here, ϵ is a random number chosen from a specific mathematical distribution. The resulting z vector has 7 latent features and acts as input for the decoder part of the model (Figure 6), which shows this reparameterisation. The decoder then uses z to reconstruct the original data.

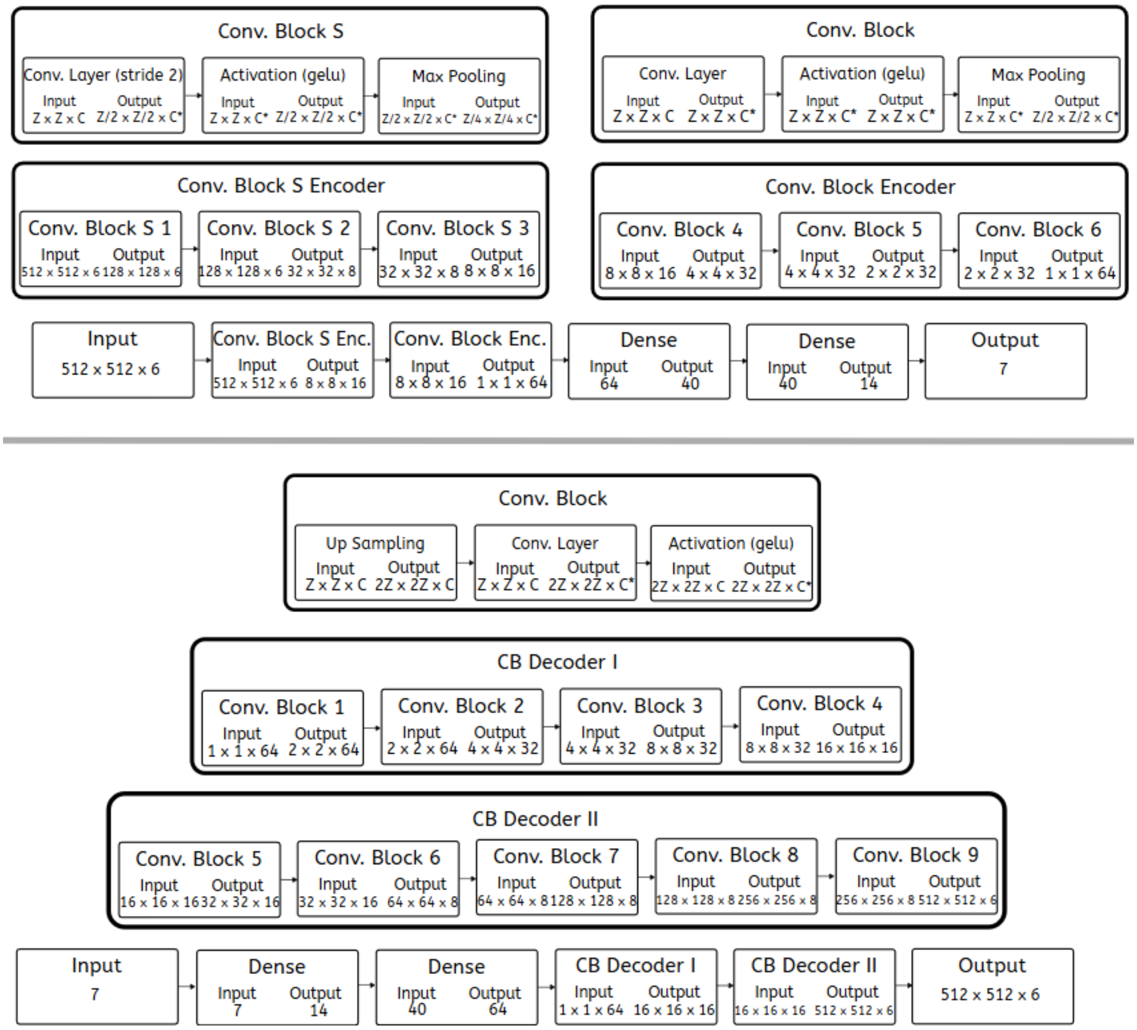


Figure 6. Structure of the sigma-VAE model used to extract the features of the hyperspectral images. The top part of the figure represents the encoder, and the bottom part represents the decoder.

A visual inspection tool was used to determine whether the 7 latent features (defined for each multispectral image) could separate the classes of firmness characterised by the cluster analysis of the firmness curves. There are several techniques in the literature showing how to visualise (how to map) a high-dimensional space (in our problem 7 dimensions) in 2 or 3 dimensions. Uniform manifold approximation and projection (UMAP) was used to visualise a high-dimensional space with 7 dimensions, which was the latent space of the AE. It is a non-linear manifold learning technique based on building a graph from data while keeping neighbourhood relationships (McInnes et al., 2018). The difference between this algorithm and those normally used, such as t-SNE (Van Der Maaten & Hinton, 2008), is that it is more scalable and does not suffer as much when dealing with large datasets.

Once the classes were separable in the latent feature space, the final step involved training an SVM classifier using the encoded data obtained from the VAE. This classifier was used to discriminate between the three firmness levels using the multispectral images with six selected wavelengths.

Figure 7 shows a schematic representation of the methodology used to classify the images into the three firmness groups defined by the cluster analysis.

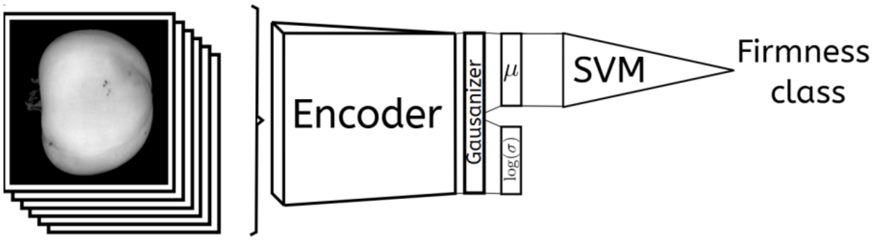


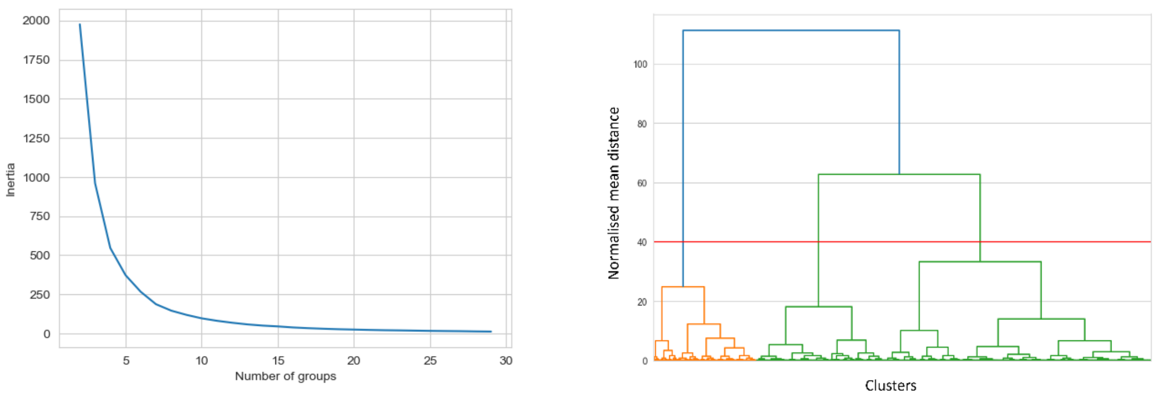
Figure 7. General schematic view of the firmness classification system.

The SVM algorithm was selected to classify different groups of firmness in the sample. To find their optimal hyperparameters, the 'Hyperopt' Python library was used with the 'Tree Parzen Estimators' option (C = 65.85, $coef_0$ = 0.11, degree = 4, gamma = 'auto' and tol = 0.0009, with a MinMaxScaler preprocessing whose parameters were clip = True and feature_range =[-1.0, 1.0]). Accuracy and confusion matrices were used to evaluate the performance of the classification model. The first was used to assess the overall performance of the model, whereas the second was used to evaluate the performance within each class.

3. Results and discussion

3.1 Firmness Clustering

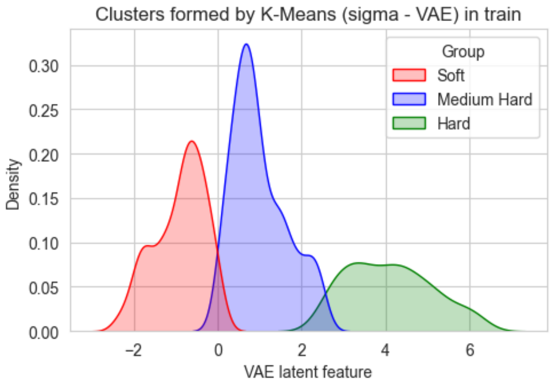
K-means and HC were utilised for cluster analysis in a non-supervised manner, using features extracted by the VAE from the breaking force curves. Figure 8a shows the elbow curve (inertia vs. number of clusters) in the clustering performed by the K-means clustering algorithm. The elbow of this curve is located between 3 and 5 clusters, meaning that the ideal number of clusters lies between 3-5 clusters. Figure 8b shows the dendrogram provided by the HC analysis, and the red line shows the threshold established to determine the number of clusters provided by the HC, namely, 3 clusters. Beyond the red line at 3 clusters, the mean distance between clusters increased by 20% when the number of clusters decreased to 2. Therefore, based on Figures 8a and 8b, the most consistent number of clusters using both techniques was 3. This was the point where the elbow curve exhibited the most significant change, and the distance between clusters remained more consistent.



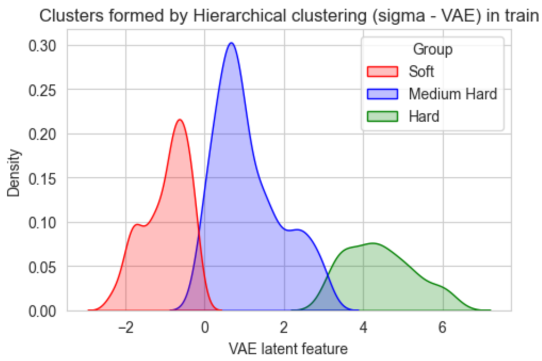
a) b)

Figure 8. The curve resulting from the K-means analysis (a) and the dendrogram resulting from the agglomerative HC (b). The red line represents the cut-off by which each sample belongs to which class is chosen.

Figure 9 shows the probability density function of the latent variable used to perform the 3 clusters obtained by K-means and HC. Both algorithms showed similar results regarding the probability density function of each group. Apparent differences were observed in the distributions for each of the groups in both algorithms. Groups 0 and 2 were differentiated from each other, while for group 1, there was a certain degree of overlap concerning the first two groups. This occurs because cluster 1 belongs to the persimmons with an intermediate firmness (medium hard firmness). This fact showed that the latent feature provided by the VAE encoder, built to compress data of firmness curves, provided information about the firmness of persimmons.



a)



b)

Figure 9. Probability density distributions of each cluster were obtained using K-means and hierarchical clustering on the training set samples.

Figure 10 shows the mean firmness curve of each of the three groups created by K-means clustering. The green line belongs to persimmons with the hardest flesh firmness since the texturometer needed the highest force to break the flesh and less time. The blue line belongs to persimmons with medium hard firmness since they required more time to break the flesh than the hard ones, showing that changes in firmness were produced. Finally, the red line belongs to soft or ripe persimmons. In this case, very little force was needed to break the flesh due to the structural changes caused by the fruit softening during storage (Tessmer et al., 2016) in the same variety of persimmon.

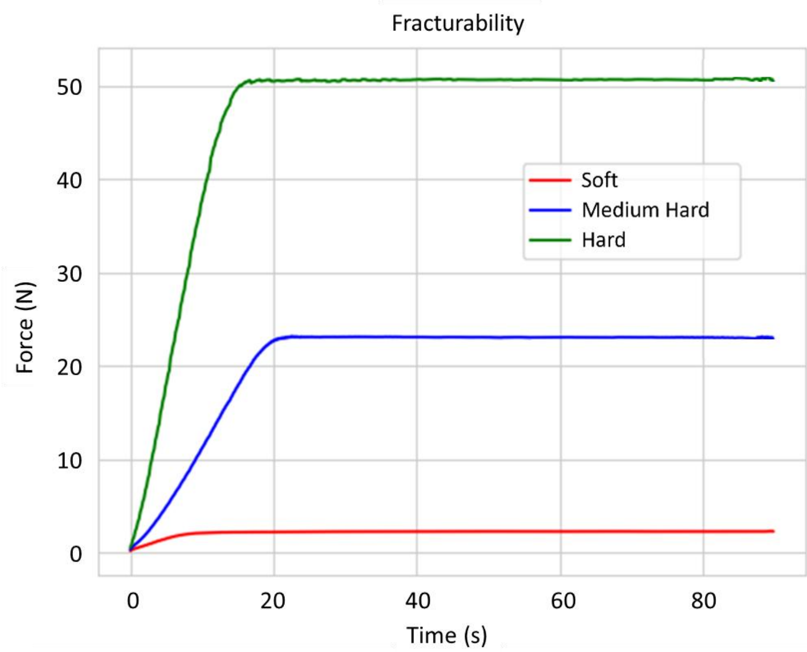


Figure 10. The mean firmness curve for each of the three firmness groups was obtained with the K-means algorithm.

Commercial standards require a firmness of at least 40 N at harvest for storage at 0 ºC –1 °C and above 20 N after 40 days of storage for commercialisation (Vilhena et al., 2022). This study successfully used clustering methods to sort persimmons into three groups, meeting these industry standards. The "hard" group, with an average breaking force of 51 N, is suitable for cold storage, as it exceeds the 40 N required for long-term storage. The "medium" group, averaging 23 N, meets the minimum 20 N needed for commercialisation. The "soft" group includes fruit with firmness below the levels needed for commercial use.

From 2380 samples, 692 were classified and labelled as hard, 114 as medium-hard, and the remaining 574 as soft. After the separation of the train and test datasets, a total of 548, 889, and 458 hard, medium-hard, and soft fruits were present in the train set. In the test set, 144, 225, and 116 hard, medium hard and soft fruits were present

3.2 Spectral dimensionality reduction

The linear correlation was used to reduce the spectral dimensionality of the hyperspectral images. Only the wavelengths that showed less correlation to each other (correlation below 0.8) were selected, resulting in six final wavelengths: 440 nm, 460 nm, 520 nm, 580 nm, 660 nm and 1000 nm. Figure 11 shows the linear correlation and a scatter plot between each wavelength selected after applying the size reduction preprocessing and wavelength selection.

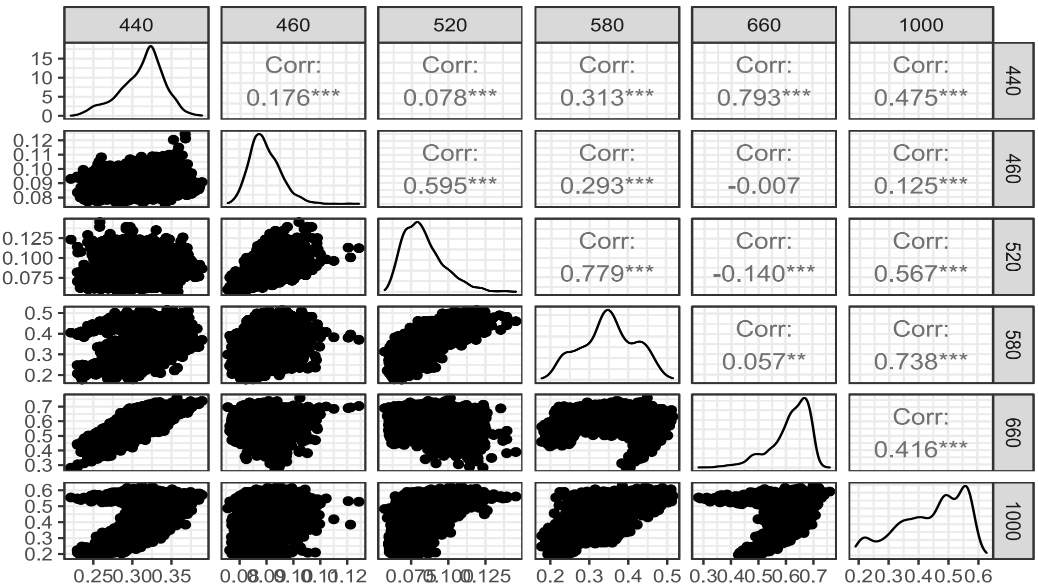


Figure 11. A pair plot of wavelengths was selected after applying the correlation-based method. The relational plots showing the relationships between the reflectance selected are shown in the bottom-left part, density plots are shown in the middle, and the correlation coefficients are displayed in the upper-right part. The correlation coefficients marked with "***" are significant at p-value < 0.001.

Figure 12 shows the mean spectra of each firmness group obtained by the two clustering methods. It can be observed that the selected wavelengths are those with the highest differences between mean spectra. These disparities predominantly manifest in crucial spectral regions, notably the water peak centred around 980 nm, indicative of free water content (Büning-Pfaue, 2003), and the peaks associated with carotenoids and chlorophyll around 500 nm and 680 nm, respectively (Wellburn, 1994). Accordingly, the chosen wavelengths at 460 nm, 520 nm, 580 nm, 660 nm and 1000 nm correspond to distinct alterations in flesh structure and ripening processes induced by varying storage temperatures. Throughout the ripening phase, there is an increase in carotenoid content and a decrease in chlorophyll content, observable through changes in reflectance at the 520 nm, 580 nm, and 660 nm bands, as outlined by Büning-Pfaue (2003). Additionally, over time, a progressive loss of parenchymal structure is attributed to membrane and cell wall degradation, leading to an augmented presence of free water, quantifiable at 980 nm, as Tessmer et al. (2016) demonstrated. Soft samples, characterised by higher free water content, exhibited lower reflectance levels than harder samples. Likewise, samples abundant in chlorophyll or carotenoids displayed lower reflectance levels. This phenomenon can be explained by the interaction of light with the sample constituents. In soft samples with elevated free water content, light penetrates deeper due to the presence of more water, leading to increased absorption and scattering within the material. As a result, less light is reflected back to the detector, resulting in lower reflectance levels. Similarly, chlorophyll and carotenoids absorb specific wavelengths within the visible spectrum. When present in higher concentrations, these pigments absorb more incident light, showing lower reflectance levels (Sun, 2010).

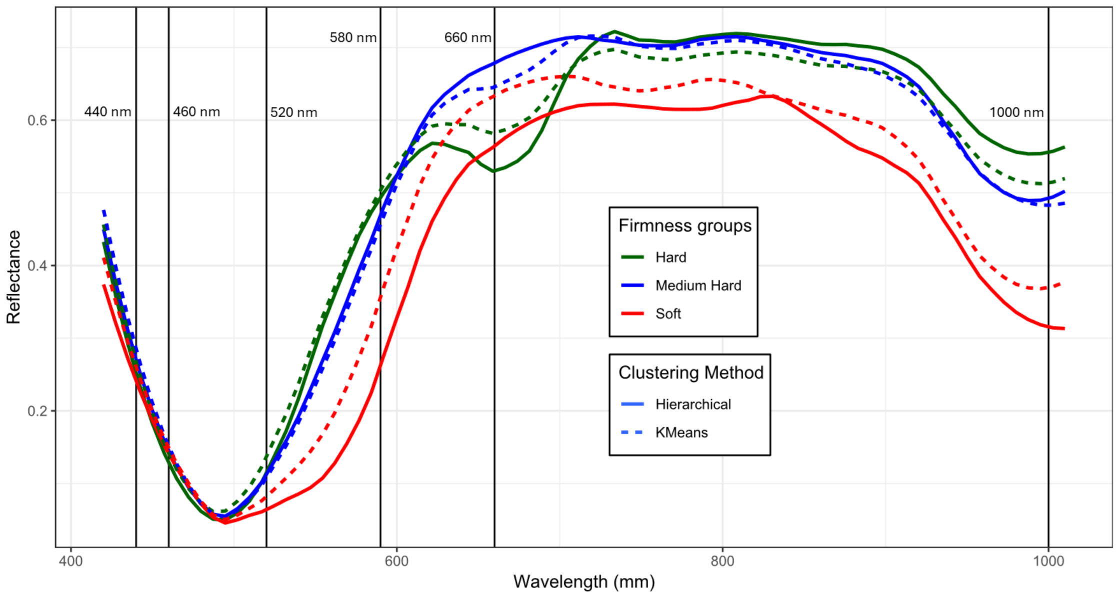


Figure 12. The mean spectrum of the persimmon fruit samples is based on the three firmness groups. The black vertical lines correspond to the six wavelengths selected.

3.3 Latent variable representation

Figure 13 shows the clustering results using UMAP for the latent space (feature selection) of the VAE. The firmness levels were separated (despite the fact that there was some overlapping between these classes) in the UMAP space. The red points belonged to medium hard firmness fruit, which overlaps with the hard (in green) and soft (in blue) firmness levels. This result was consistent with the results shown in Figure 9, where the overlapping of the probability density distribution of each class was shown for the K-means and HC clustering methods.

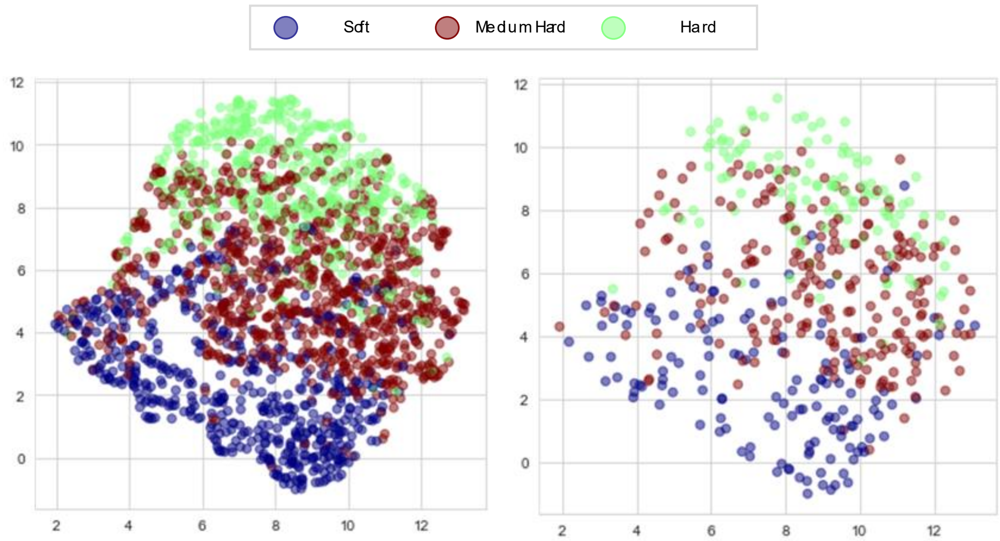


Figure 13. 2D UMAP representation of the latent components obtained in the multispectral image model. The training dataset is on the left, and the test dataset is on the right. Green represents the 'hard' firmness fruit; red corresponds to 'medium-hard', and blue corresponds to 'soft'.

3.4 Firmness classification

The SVM model was trained using the encoded data from the VAE, achieving a training accuracy of 85.0% and a test accuracy of 81.4%. The confusion matrices, shown in Table 1, provide a detailed overview of the classification performance.

The results demonstrate that the combination of multispectral images, AE and SVM algorithm classified the persimmon firmness accurately. Specifically, when considering the test set, hard firmness persimmons were classified with an 83.3% accuracy rate, medium-hard firmness persimmons with an accuracy rate of 85.3%, and soft persimmons with a 75.6% accuracy rate. Due to the application of spectral dimensionality reduction of the images, the model was trained with images containing six distinct wavelengths (440 nm, 460 nm, 520 nm, 580 nm, 660 nm, and 1000 nm). This enables the utilisation of multispectral cameras instead of hyperspectral ones for real-time detection tasks. This approach is advantageous because hyperspectral images, which capture a bigger range of wavelengths, typically require more computing resources and time for processing, making them slower and less efficient for real-time applications. Whereas multispectral cameras offer a simpler and faster data acquisition process, and the detection system becomes more practical and accessible for real-world scenarios.

Table 1. Confusion matrices of firmness discrimination in the training and test sets.

Class	Training set				Test set			
	Hard	Medium hard	Soft	A (%)	Hard	Medium hard	Soft	A (%)
Hard	477	71	0	87.0	120	24	0	83.3
Medium hard	42	756	91	85.0	21	170	34	75.6
Soft	0	81	377	82.3	0	17	99	85.3

A = accuracy

Firmness determination in fruit using hyperspectral imaging has been extensively studied, although mainly employing regression methods such as Partial Least Squares (PLS). Li et al. (2016) predicted firmness characteristics in different pear varieties and obtained an R^2 value of 0.88. Leiva-Valenzuela et al. (2013) obtained an R^2 value of 0.87 for predicting firmness in blueberries. Xie et al. (2018) predicted banana firmness, obtaining an R^2 value of 0.76. Mendoza et al. (2011) determined apple firmness with R^2 values ranging from 0.84 to 0.95, depending on the apple variety. On the other hand, Lu and Peng (2006) and Xuan et al. (2022) used MLR and obtained R^2 values of 0.58, 0.77 and 0.82, respectively, for three different varieties. Peng and Lu (2008) also used MLR and obtained an R^2 value of 0.89 for predicting apple firmness. Moreover, Pullanagari and Li (2021) applied Gaussian Process regression for firmness prediction in sweet cherries and achieved an R^2 value of 0.60. The only study found in the literature that correlated spectral data (spectroscopy) with firmness prediction in persimmon achieved an R^2 value of 0.89 using PLS (Hemrattrakun et al., 2021).

Only a few authors have explored the use of a VAE for feature extraction and dimensionality reduction coupled with an ML classifier. Xu et al. (2022) employed a similar approach; they predicted grape firmness using spectral images as input for the AE but used SVM for regression and achieved an R^2 value of 0.90. Yu et al. (2018) employed the mean spectra as input to the AE to predict firmness in pears and a multi-layer perceptron as a regressor and achieved an R^2 value of 0.89.

It should be noted that the accuracy of firmness prediction is not determined only by the classification model or data management techniques employed. Different fruits exhibit unique internal characteristics and structures and undergo distinct changes in firmness during ripening or cold storage.

4. Conclusions

The use of hyperspectral imaging was evaluated to determine the firmness of 'Rojo Brillante' persimmon fruit after cold storage under industrial conditions. This study introduces a novel approach combining VAEs for feature extraction from firmness curves and hyperspectral images. This method rarely applied to fruits and was unprecedented in persimmons, effectively reducing over 99% of the dimensionality of both datasets while retaining critical information, enabling efficient and accurate analysis. Furthermore, three firmness levels (high, medium, and soft), essential for determining storage and commercialisation limits, were successfully identified using K-means clustering and hierarchical clustering applied to firmness curves. Using only seven features derived from hyperspectral images at six wavelengths, the SVM algorithm correctly classified 81.3% of the fruit into the three firmness levels. The results of this study highlight hyperspectral imaging, combined with VAEs, as a promising tool for rapid, non-destructive firmness monitoring of 'Rojo Brillante' persimmons in industrial settings. Notably, the reduction of hyperspectral data to six wavelengths enables the potential development of multispectral cameras, which can facilitate in-line detection and sorting of fruit firmness. Integrating this innovation into industrial processes could significantly reduce costs by enabling efficient and automated quality assessment.

Nevertheless, this research represents a foundational step, and future studies must address several limitations. Additional harvest seasons and a broader range of persimmon cultivars are necessary to confirm the generalizability and robustness of the findings. Moreover, further investigation into the technical feasibility of implementing in-line detection systems using multispectral cameras that capture the selected wavelengths in this study is required to ensure their practical application and reliability in diverse industrial environments. The promising combination of hyperspectral imaging, VAE-based feature extraction, and machine learning algorithms offers significant potential to transform fruit quality monitoring processes. Such advancements will help reduce operational costs and enhance industrial efficiency while maintaining high product quality.

Acknowledgements

This work was partially funded through the project AEI PID2023-150192OR-C31 and C33 with the support of FEDER funds, project AEI TED2021-130117B-C31 and C33, by the project GVA-PROMETEO CIPROM/2021/014 and by the project PID2021-127975OR-C21. Salvador Castillo thanks INIA for the FPI-INIA grant number PRE2020-094491, partially supported by European Union FSE funds. Nariane Q. Vilhena thanks the Ministerio de Ciencia e Innovación of Spain for grant PRE2018-085833. Sandra Munera expresses her thanks for the postdoctoral contract Juan de la Cierva - Formación (FJC2021-047786-I) co-funded by MCIN/AEI/10.13039/501100011033 and European Union NextGeneration EU/PRTR.

References

Aurélien Géron. (2019). Hands-on machine learning with Scikit-Learn, Keras and TensorFlow: concepts, tools, and techniques to build intelligent systems. O'Reilly Media, 851. https://www.oreilly.com/library/view/hands-on-machine-learning/9781492032632/

Battey, C. J., Coffing, G. C., & Kern, A. D. (2021). Visualizing population structure with variational autoencoders. G3: Genes|Genomes|Genetics, 11(1), jkaa036. https://doi.org/10.1093/G3JOURNAL/JKAA036

Bank, D., Koenigstein, N., Giryes, R., 2021. Autoencoders. https://doi.org/10.48550/arXiv.2003.05991

Besada C., 2008. Mejora de la tecnología de desastringencia, conservación y comercialización de caqui "Rojo Brillante" para consumo en fresco. In Doctoral Thesis, Universidad Politécnica de Valencia..

Büning-Pfaue, H., 2003. Analysis of water in food by near infrared spectroscopy. Food Chem. 82(1), 107–115. https://doi.org/https://doi.org/10.1016/S0308-8146(02)00583-6

Castillo-Girones, S., Van Belleghem, R., Wouters, N., Munera, S., Blasco, J., & Saeys, W. (2024). Detection of subsurface bruises in plums using spectral imaging and deep learning with wavelength selection. Postharvest Biology and Technology, 207, 112615. https://doi.org/https://doi.org/10.1016/j.postharvbio.2023.112615

Cui, Z., Wang, N., Duan, Y., Xu, X., Wang, R., Zhang, S., & Ma, C. (2021). High-Resolution Microstructure Analysis of Cork Spot Disordered Pear Fruit "Akizuki" (Pyrus pyrifolia Nakai) Using X-Ray CT. Frontiers in Plant Science, 12, 715124. https://doi.org/10.3389/FPLS.2021.715124/BIBTEX

Eid, H., Hassanien, A. E., Kim, T.-H., Banerjee, S., 2013. Linear Correlation-Based Feature Selection For Network Intrusion Detection Model. In Commun. Comput. Inf. Sci. (Vol. 381). https://doi.org/10.1007/978-3-642-40597-6_21

FAOSTAT, 2023. https://www.fao.org/faostat/en/#data/QCL. Accessed 16.12.24

Fathizadeh, Z., Aboonajmi, M., Hassan-Beygi, S. R., 2021. Nondestructive methods for determining the firmness of apple fruit flesh. Inf. Process.Agric. 8(4), 515–527. https://doi.org/https://doi.org/10.1016/j.inpa.2020.12.002

Fathizadeh, Z., Aboonajmi, M., & Beygi, S. R. H. (2020). Nondestructive firmness prediction of apple fruit using acoustic vibration response. Scientia Horticulturae, 262, 109073. https://doi.org/https://doi.org/10.1016/j.scienta.2019.109073

Gaikwad, S., & Tidke, S. (2022). Multi-Spectral Imaging for Fruits and Vegetables. International Journal of Advanced Computer Science and Applications, 13(2), 743–760. https://doi.org/10.14569/IJACSA.2022.0130287

Hemrattrakun, P., Nakano, K., Boonyakiat, D., Ohashi, S., Maniwara, P., Theanjumpol, P., Seehanam, P. 2021. Comparison of Reflectance and Interactance Modes of Visible and Near-

Infrared Spectroscopy for Predicting Persimmon Fruit Quality. Food Anal Methods. 14(1), 117–126. https://doi.org/10.1007/s12161-020-01853-w

Huang, Y., Wang, J., Li, N., Yang, J., Ren, Z. 2021. Predicting soluble solids content in "Fuji" apples of different ripening stages based on multiple information fusion. Pattern Recognit Lett. 151, 76–84. https://doi.org/https://doi.org/10.1016/j.patrec.2021.08.003

Joseph, M., Van Cauteren, H., Postelmans, A., Nugraha, B., Verreydt, C., Verboven, P., Nicolai, B., & Saeys, W. (2023). Porosity quantification in pear fruit with X-ray CT and spatially resolved spectroscopy. Postharvest Biology and Technology, 204, 112455. https://doi.org/10.1016/J.POSTHARVBIO.2023.112455

Kingma, D. P., and Welling, M., 2013. Auto-Encoding Variational Bayes. 2nd, ICLR 2014 - Conference Track Proceedings. https://arxiv.org/abs/1312.6114v10

Leiva-Valenzuela, G. A., Lu, R., Aguilera, J. M., 2013. Prediction of firmness and soluble solids content of blueberries using hyperspectral reflectance imaging. J Food Eng. 115(1), 91–98. https://doi.org/https://doi.org/10.1016/j.jfoodeng.2012.10.001

Lesage, P., & Destain, M.-F. (1996). Measurement of tomato firmness by using a non-destructive mechanical sensor. Postharvest Biology and Technology, 8(1), 45–55. https://doi.org/https://doi.org/10.1016/0925-5214(95)00057-7

Li, B., Hou, B., Zhang, D., Zhou, Y., Zhao, M., Hong, R., Huang, Y., 2016. . Pears characteristics (soluble solids content and firmness prediction, varieties) testing methods based on visible-near infrared hyperspectral imaging. Optik, 127(5), 2624–2630. https://doi.org/https://doi.org/10.1016/j.ijleo.2015.11.193

Liu, C., Komatsu, T., Lopez-Moreno, H., Mura, J., Zalapa, J., & Ikeda, S. (2025). Internal structure, density, and mechanical properties of cranberry fruits. Journal of Food Science, 90(3), e70153. https://doi.org/10.1111/1750-3841.70153

Lu, R., and Peng, Y., 2006. Hyperspectral Scattering for assessing Peach Fruit Firmness. Biosyst Eng, 93(2), 161–171. https://doi.org/https://doi.org/10.1016/j.biosystemseng.2005.11.004

Madhulatha, T. S., 2012. An Overview on Clustering Methods. IOSR j. eng., 02(04), 719–725. https://doi.org/10.9790/3021-0204719725

Matsuo, T., and Itoo, S., 1982. A model experiment for de-astringency of persimmon fruit with high carbon dioxide treatment: In vitro gelation of kaki-tannin by reacting with acetaldehyde. Agric. Biol. Chem. 46, 683–689.

McInnes, L., Healy, J., Melville, J., 2018. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. https://arxiv.org/abs/1802.03426v3

Mendoza, F., Lu, R., Ariana, D., Cen, H., Bailey, B., 2011. Integrated spectral and image analysis of hyperspectral scattering data for prediction of apple fruit firmness and soluble solids content. Postharvest Biol. Technol. 62(2), 149–160. https://doi.org/https://doi.org/10.1016/j.postharvbio.2011.05.009

Munera, S., Rodríguez-Ortega, A., Aleixos, N., Cubero, S., Gómez-Sanchis, J., Blasco, J., 2021. Detection of Invisible Damages in 'Rojo Brillante' Persimmon Fruit at Different Stages Using Hyperspectral Imaging and Chemometrics. Foods, 10(9). https://doi.org/10.3390/foods10092170

Munera, S., Amigo, J. M., Aleixos, N., Talens, P., Cubero, S., & Blasco, J. (2018). Potential of VIS-NIR hyperspectral imaging and chemometric methods to identify similar cultivars of nectarine. Food Control, 86, 1–10. https://doi.org/https://doi.org/10.1016/j.foodcont.2017.10.037

Nugraha, B., Verboven, P., Janssen, S., Wang, Z., & Nicolaï, B. M. (2019). Non-destructive porosity mapping of fruit and vegetables using X-ray CT. Postharvest Biology and Technology, 150, 80–88. https://doi.org/10.1016/J.POSTHARVBIO.2018.12.016

Park, B., Shin, T., Cho, J.-S., Lim, J.-H., Park, K.J., 2023. Improving blueberry firmness classification with spectral and textural features of microstructures using hyperspectral microscope imaging and deep learning. Postharvest Biol. Technol. 195, 112154. https://doi.org/https://doi.org/10.1016/j.postharvbio.2022.112154

Patel, S., Sihmar, S., Jatain, A. 2015. A study of hierarchical clustering algorithms. 2015 2nd (INDIACom), 537–541.

Peng, Y., and Lu, R., 2008. Analysis of spatially resolved hyperspectral scattering images for assessing apple fruit firmness and soluble solids content. Postharvest Biol. Technol., 48(1), 52–62. https://doi.org/https://doi.org/10.1016/j.postharvbio.2007.09.019

Pullanagari, R. R., and Li, M. 2021. Uncertainty assessment for firmness and total soluble solids of sweet cherries using hyperspectral imaging and multivariate statistics. J Food Eng., 289, 110177. https://doi.org/https://doi.org/10.1016/j.jfoodeng.2020.110177

Ramos, F., Valero, C., Homer, I., Puig-Mauri, J., Ruiz-Altisent, M. (2005). Non destructive fruit firmness sensors A review. Span. J. Agric. Res., 3. https://doi.org/10.5424/sjar/2005031-125

Rybkin, O., Daniilidis, K., Levine, S., 2021. Simple and Effective VAE Training with Calibrated Decoders. https://doi.org/10.48550/arXiv.2006.13202

Salvador, A., Besada, C., Crisosto, C. H., 2020. Persimmon. In: Manual on postharvest handling of Mediterranean tree fruits and nuts (pp. 88-133). Wallingford UK: CABI.

Salvador, A., Arnal, L., Besada, C., Larrea, V., Quiles, A., I. Pérez-Munuera (2007). Physiological and structural changes during ripening and deastringency treatment of persimmon cv 'Rojo Brillante'. Postharvest Biology and Technology, 46, 181-188. doi:10.1016/j.postharvbio.2007.05.003

Shi, N., Liu, X., Guan, Y. 2010. Research on k-means clustering algorithm: An improved k-means clustering algorithm. 3rd IITSI 2010, 63–67. https://doi.org/10.1109/IITSI.2010.74

Sun, D. W. (2010). Hyperspectral Imaging for Food Quality Analysis and Control. Hyperspectral Imaging for Food Quality Analysis and Control, 1–477. https://doi.org/10.1016/C2009-0-01853-4

Tessmer, M. A., Besada, C., Hernando, I., Appezzato-da-Glória, B., Quiles, A., Salvador, A., 2016. Microstructural changes while persimmon fruits mature and ripen. Comparison between astringent and non-astringent cultivars. Postharvest Biol. Technol. 120, 52–60. https://doi.org/https://doi.org/10.1016/j.postharvbio.2016.05.014

Tian, S., Wang, J., & Xu, H. (2022). Firmness measurement of kiwifruit using a self-designed device based on acoustic vibration technology. Postharvest Biology and Technology, 187, 111851. https://doi.org/10.1016/J.POSTHARVBIO.2022.111851

Torkashvand, A. M., Ahmadi, A., Nikravesh, N. L., 2017. Prediction of kiwifruit firmness using fruit mineral nutrient concentration by artificial neural network (ANN) and multiple linear regressions (MLR). J. Integr. Agric. 16(7), 1634–1644. https://doi.org/https://doi.org/10.1016/S2095-3119(16)61546-0

Van Der Maaten, L., Hinton, G., 2008. Visualizing Data using t-SNE. J. Mach. Learn. Res. 9, 2579–2605.

Vilhena, N. Q., Quiles, A., Gil, R., Llorca, E., Fernández-Serrano, P., Vendrell, M., & Salvador, A. (2022). Slight Changes in Fruit Firmness at Harvest Determine the Storage Potential of the 'Rojo Brillante' Persimmon Treated with Gibberellic Acid. Horticulturae, 8(2). https://doi.org/10.3390/horticulturae8020140

Wellburn, A. R., 1994. The Spectral Determination of Chlorophylls a and b, as well as Total Carotenoids, Using Various Solvents with Spectrophotometers of Different Resolution. J. Plant Physiol. 144(3), 307–313. https://doi.org/https://doi.org/10.1016/S0176-1617(11)81192-2

Xie, C., Chu, B., He, Y., 2018. Prediction of banana color and firmness using a novel wavelengths selection method of hyperspectral imaging. Food Chem. 245, 132–140. https://doi.org/https://doi.org/10.1016/j.foodchem.2017.10.079

Xu, M., Sun, J., Yao, K., Cai, Q., Shen, J., Tian, Y., Zhou, X., 2022. Developing deep learning-based regression approaches for prediction of firmness and pH in Kyoho grape using Vis/NIR hyperspectral imaging. Infrared Phys. Technol. 120, 104003. https://doi.org/https://doi.org/10.1016/j.infrared.2021.104003

Xuan, G., Gao, C., Shao, Y., 2022. Spectral and image analysis of hyperspectral data for internal and external quality assessment of peach fruit. Spectrochim. Acta A Mol. Biomol. Spectrosc. 272, 121016. https://doi.org/https://doi.org/10.1016/j.saa.2022.121016

Yu, X., Lu, H., Wu, D., 2018. Development of deep learning method for predicting firmness and soluble solid content of postharvest Korla fragrant pear using Vis/NIR hyperspectral reflectance imaging. Postharvest Biol. Technol. 141, 39–49. https://doi.org/https://doi.org/10.1016/j.postharvbio.2018.02.013

CHAPTER III. Detection of subsurface bruises in plums using spectral imaging and deep learning with wavelength selection.

Detection of subsurface bruises in plums using spectral imaging and deep learning with wavelength selection

Castillo-Girones, S[1]., Van Belleghem, R[2]., Wouters, N[2]., Munera, S[3]., Blasco, J[1]., Saeys, W[2].

[1] Centro de Agroingeniería. Instituto Valenciano de Investigaciones Agrarias (IVIA). Ctra. Moncada-Náquera km 4.5. 46113. Moncada. Valencia (Spain).

[2] KU Leuven, Department of Biosystems, MeBioS, Kasteelpark Arenberg 30, 3001, Leuven, Belgium.

[3] Departamento de Ingeniería Gráfica. Universitat Politècnica de València. Camino de Vera. s/n. 46022 Valencia (Spain).

Postharvest Biology and Technology, 207, 112615.

https://doi.org/https://doi.org/10.1016/j.postharvbio.2023.112615

Abstract

Plums are widely consumed, both fresh and processed. During harvest, handling, or transportation, they are exposed to static and dynamic compression forces exceeding the critical stress for tissue damage. Compression-related damage typically develops further as the fruit ripens and softens, facilitating the occurrence of rot, which might cause significant losses in the supply chain. Early detection of these damages is crucial to sorting the damaged fruit out and deviating it to processing, thus preventing food waste. However, early-stage bruises or damages on plums are not visible, especially not in dark-skin cultivars. Therefore, this study aimed to explore the potential of hyperspectral imaging in the 430 to 1,000 nm range and deep learning algorithms to detect these invisible bruises at an early stage. To this end, 'Presenta' plums were impacted at three different levels to simulate varying degrees of damage. Images of both bruised and non-bruised plums were taken immediately after bruising and 24 and 48 hours after bruise induction. Three distinct CNNs were trained to analyse the images. Two of these networks were implemented using transfer learning (ResNet and HSCNN), while the third was custom-designed for this specific purpose. The most informative wavelengths were identified as inputs for the CNNs employing PCA. F1 scores over 81 % were obtained in all cases, and almost 100 % accuracy was obtained in classifying the bruised plums with the highest impact energy of 0.50 J. Thus, detecting and classifying bruised plums using only three wavelengths is possible, paving the way for in-line sorting with multispectral cameras in packing houses.

Keywords: Spectral imaging, postharvest, invisible bruises, plums, deep learning

1. Introduction

Plums (*Prunus domestica L.*) are highly favored for their delicious taste, appealing texture, and abundant antioxidants and vitamins (Jayasankar et al., 2016). Worldwide, the area cultivated with plums is 2 619 471 ha, with a production of 11 758 thousand tons in 2017, 25.9 % in Europe (Butac, 2020). This ranks plums in fourth place in terms of production among the temperate fruit trees after apple, pear, and peach.

The main issue of this fruit in postharvest is that mechanical and compression damages happen during harvesting and management. This damage causes invisible internal lesions when the fruit is inspected and packed, leading to early spoilage and the latter rejection by consumers, resulting in significant losses in the fruit supply chain (Martínez-Romero et al., 2003). When bruises are produced, the critical contact stress is exceeded, and plums experience cell wall rupture and tissue integrity loss, releasing compounds like polyphenols into the fruit matrix. When exposed to oxygen, these compounds cause tissue browning (Soliva-Fortuny et al., 2001). Additionally, bruised fruit exhibit increased climacteric ethylene and respiration ripening peaks (Martínez-Romero et al., 2003; Serrano et al., 2004). The changes in biochemical composition, such as elevated ethylene and respiration rates and increased abscisic acid levels, provide clues for detecting bruises by analysing ethylene and respiration production through chromatography (Serrano et al., 2004). Still, chromatography methods are time-consuming and unsuitable for industrial in-line detection of invisible bruises. Therefore, the development of new technologies for early internal lesion detection in plums is a growing need to avoid wastage in retail.

However, identifying early bruises is challenging, especially in dark-skinned cultivars. Advances in optical technologies, such as hyperspectral imaging, allow for measuring internal fruit quality (Munera et al., 2019) and can be a promising technology for invisible internal bruise detection and can be a good technology for invisible internal bruise detection as it can obtain information from light that has partially penetrated the fruit tissue and be scattered towards the camera (Rodríguez-Ortega et al., 2023). In recent years, several scientific investigations have addressed the issue of fruit bruise detection and classification through diverse hyperspectral imaging techniques with relatively good accuracy. Yin et al. (2022) employed the Otsu binarisation method, achieving an accuracy rate of 96 % in detecting bruises in loquat. In persimmon, Huang et al. (2022) applied SVM to classify damaged and non-damaged persimmon, obtaining 100 % accuracy, and Munera et al. (2021) detected internal bruises using PCA, determining the time of damage induction with an accuracy of 99 %. Li et al. (2021) undertook a study involving the classification of bruised peaches, obtaining a 100 % accuracy 48 h after being hit with SVM. Yuan et al. (2021) bruised jujube fruit and monitored the progression of bruises over time, getting an accuracy of almost 100 % using PLS-DA. (Velásquez C et al., 2023) used several classification methods to early detect anthracnose damage in mango by hyperspectral imaging, detecting the disease before the development of visible symptoms.

However, (Arango et al., (2021) were the first to use deep learning algorithms for fruit bruise detection. They applied convolutional neural networks (CNNs) to both RGB and NIR images of bruised apples, obtaining an accuracy of 97 % in both imaging scenarios.

CNNs are a specific type of Artificial Neural Network (ANN) designed for image analysis. Compared to other machine learning methods, ANNs can extract more complex relationships from the data (Arul, 2021). CNNs apply small spatial filters to different regions of an input image that slide across the image and perform convolution operations. This process extracts features like edges, textures, or patterns (Elgendy et al., 2020). This feature extraction and non-linear modeling make CNNs particularly effective for complex problems like fruit damage detection, where bruises can vary widely in appearance or be internal and invisible, as demonstrated in olive anthracnose detection (Fazari et al., 2021).

Nevertheless, hyperspectral imaging technology can be computationally intensive. To cope with the high sorting speeds required in the industry, the number of wavelengths is typically reduced to retain only the most relevant wavelengths in a multispectral system, reducing computational demands (Amigo., 2019). In this sense, PCA is a method commonly used for dimensionality reduction in hyperspectral images of fruit such as loquat (Yin et al., 2022), kiwi fruit (Lü & Tang, 2012), apples (Hu et al., 2020) and peaches (Li et al., 2018). This method replaces the original, highly correlated spectral variables with a smaller number of orthogonal linear combinations, known as PCs, that capture most of the variance in the data (Saeys et al., 2019).

Despite the extensive research on hyperspectral imaging applications in many fruit species, no reports were found on its use for detecting subsurface bruises in plums using deep learning. Visual bruise inspection is particularly challenging in dark-skinned cultivars like ‘Presenta’ used in this study. Thus, the primary aim of this study was to investigate the feasibility of employing hyperspectral images processed using CNNs as a tool for the early detection of bruises in dark plums, considering the severity of the damage (feeble, medium, hard). The secondary aims of this study were: i) to assess and compare three CNN models with different architectures: two models using transfer learning techniques and another customised from the ground up; ii) to select the most important wavelengths for the detection of the bruises using PCA; and iii) to provide deeper insights into the operational mechanisms of each CNN model and to pinpoint the specific regions of interest within the images that these models focus on using a GradCam analysis.

2. Materials and Methods

2.1 Plant material and bruising

Plums (*Prunus domestica* L.) cv. 'Presenta' is a cultivar characterised by its yellow flesh and dark purple skin, which makes it challenging to detect visually minor bruises. The first lot of 255 fruit was obtained in the second week of September, and a second lot of 120 plums was purchased two weeks later. All fruit was obtained from a local producer (Fruithandel Romain Wouters & Co nv, Geetbets, Belgium) during the 2022 season when they had reached commercial maturity. Upon arrival at KU Leuven, the fruit was placed into boxes of 15 plums each. Fruit from the first lot was packed in 17 boxes, while eight boxes were used for the second lot of 120 plums. The fruit was stored in a cold room at 6 °C and 80 % relative humidity until 24 hours before the start of each experiment when they were moved to room temperature (20 °C).

After the fruit was tempered for 24 hours, bruising was produced by automatically dropping a pendulum attached to an electromagnet from a defined angle. The impact occurred around the equator of the plum (Figure 1 left) and was measured as described in Keresztes et al. (2016, 2017) and van Zeebroeck et al. (2003). The device used in this study involves a load cell (Dytran 1051V3, max. load 444 N, sensitivity 11.2 Mv/N) and a high-resolution encoder (Heidenhain RON275, max. speed 3,000 rev/min, resolution 90,000 pulses/rev). It consisted of a 0.505 m long wooden arm with a spherical aluminium impactor (25 mm curvature radius) that contained a force sensor to record the force during impact. Three different angles were used to produce different impact energies, referred to as feeble (0.12 J and 20°), medium (0.32 J and 30°), and hard (0.50 J and 40°). Each box of 15 fruit was arranged in rows of two and three fruit (Figure 1 right). Nine fruit of each box were bruised, corresponding to the rows with three fruit. Figure 1 right shows the fruit damaged with the lowest energy (top row), medium energy (central row), and the highest energy (bottom row).

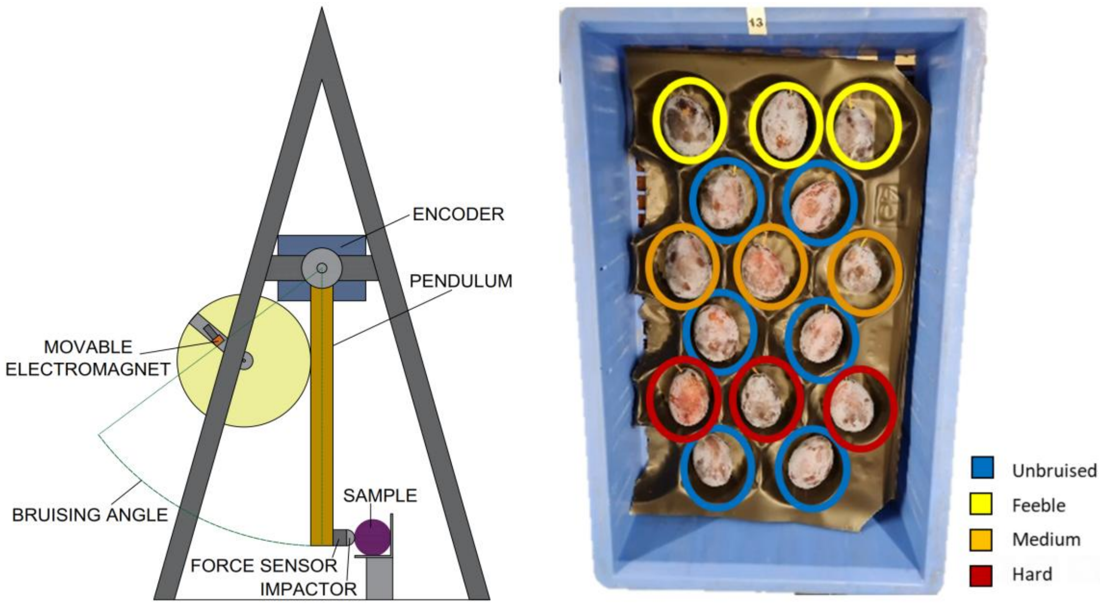


Figure 5. Pendulum device used for creating the bruises while measuring the impact force (left) and an example of a measured box with plum samples (right).

2.2 Hyperspectral image acquisition and processing

Hyperspectral images of the fruit were taken at 0, 24, and 48 hours from inducing the bruises. The images were acquired using a line-scan camera (Specim FX10 from Specim, Oulu, Finland) in the wavelength range from 400 to 1,000 nm with a spectral resolution of 2.7 nm (211 wavelengths per hyperspectral image) and a F/1.7 aperture. Samples were scanned with a resolution of 0.28 mm/scan and an exposure time of 10 ms. In this experiment, the spectral range from 430 to 1 000 nm was measured as the camera had low sensitivity for wavelengths shorter than 430 nm, resulting in a low signal-to-noise ratio. Therefore, the acquired hypercubes had 211 spectral bands.

Diffuse-indirect light to avoid specular reflections illuminated the fruit from 8, 12V 55W halogen spotlights Phillips H7 (Phillips, Amsterdam, Netherlands). The lamps were arranged equidistant from each other inside a white hemispherical diffuser, as shown in Figure 2 left. The hemispherical diffuser was located 20 cm above the box with the plums.

White and dark references were acquired after every four hyperspectral images to correct the influence of the artificial light and the sensitivity of the equipment. White referencing was performed with a diffuse reference target (Spectralon CSRT-MS-120 MP-10915-T0B, Labsphere Inc., New Hampshire, USA), and dark images were taken with the camera's shutter closed. The raw intensity values of the images were later converted into relative reflectance values using the white and dark reference values.

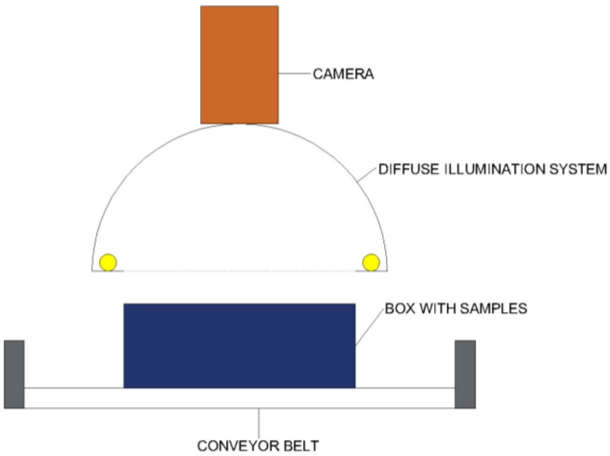


Figure 6. Schematic illustration of the hyperspectral imaging system with an indication of its main components

Each image contained an entire box. To extract the hyperspectral images of the single plums, the first step was to calculate the ratio from the reflectance images acquired at 500 nm and 800 nm, calculated as in Eq. (1). Then, adaptative segmentation was performed using the Otsu algorithm to obtain a binary mask of the plums. Finally, the hyperspectral images of each plum were extracted and saved in hdf5 file format, resulting in two datasets, one for the first lot of fruit composed of 765 hypercubes and the other for the second lot composed of 360 hypercubes corresponding to the 375 individual fruit, captured in three moments: immediately after bruising, and after 24 and 48 hours. All images were finally resized to 224 by 224 pixels.

$$Im_{ratio} = \frac{Im_{500\ nm} - Im_{800\ nm}}{Im_{500\ nm} + Im_{800\ nm}}$$

Equation 3. Reflectance ratio for performing Otsu segmentation.

2.3. Data analysis

The two lots of fruit were used to build and externally validate the model. The first lot of fruit was used to build the model and divided into two datasets: a training set composed of 12 boxes and an internal validation set composed of the remaining 5 boxes. The training set was used to select the important wavelengths and tune up the hyperparameters, while the validation set was used to obtain the performance of the models. The second lot of fruit, composed of other plums purchased on a different day, was used to validate the model with an external and independent test set of samples.

The values of the spectral bands for the individual pixels of all hyperspectral images were standardised by subtracting the mean value and dividing by the standard deviation.

2.3.1 Wavelength selection

Dimensionality reduction was performed with PCA. To select the most important wavelengths, 161,400 randomly sampled spectra from all images in the training data (all days) were used. Important wavelengths, responsible for most of the spectral variance, were identified as those corresponding to the peaks (both positive and negative) in the loading vectors of selected PCs. Only peaks with a loading value higher than 0.05 were considered to limit the number of selected wavelength variables, as proposed by Zhu et al. (2019).

2.3.2 Deep Learning Models

Two already developed DL models were developed. The first one, designed specifically for hyperspectral image processing, was based on an HSCNN proposed by Varga et al. (2021). The second was adapted from ResNet (He et al., 2016). ResNet was initially developed to receive as input RGB images, so it had to be adapted by extending the kernel depth of the first convolution layer from three to the required wavelengths (211).

In addition, a third model based on a customised 3-dimensional (3D) CNN was created from scratch. CNN-based models with 3D convolutional layers can extract more information from spectra than 2D CNN models (Elgendy et al., 2020). The architecture consisted of three convolutional layers with varying filter sizes (16, 32, and 64 filters), each using a 3x3 kernel and ReLu activation functions. Subsequently, three max-pooling layers with a pool size of 2 and three batch normalisation layers were applied. Following the last batch normalisation layer, a global average pooling 3D layer was introduced to consolidate the information. Then, two final dense layers were used for classification. The first dense layer featured 128 neurons and was followed by a dropout layer with a rate of 0.2 to avoid overfitting. The last dense layer consisted of a single artificial neuron with a sigmoid activation function for binary classification. All networks were adapted for a binary classification, with the class labels being 0 (unbruised) or 1 (bruised).

Four input dimensions were tested: the full hypercubes with 211 channels, hypercubes composed of the selected wavelengths, 3-channel hypercubes that combine the selected bands that individually gave the best performance, and one-channel images with the best-performing wavelength. Each input model was optimised with the same hyperparameters listed in Table 1. The training was done on a personal computer (PC) with a graphics processing unit (GPU) GeForce RTX 3060 Ti GPU (NVIDIA, California, USA) with 8 GB of VRAM. The number of epochs was determined empirically based on the training and validation F1 scores. As a model optimiser, an adapted version of the Adam optimiser (AdamW) was used, proposed by Loshchilov and Hutter. (2019). The optimiser selection is critical since it allows the model weights to be adjusted to maximise the loss function used to measure the model performance. A decaying learning rate increased model stability over the 300 epochs to help find the optimal weights. Finally, to determine which of the selected wavelengths were most discriminative for bruise detection, a model was built for every single wavelength of the selected wavelengths.

Table 4. Hyperparameters used for training the deep learning models.

Number of epochs	300 (800 in the 3D-CNN)	
Batch size	10	
Optimiser	AdamW (Loshchilov & Hutter, 2019)	
Loss function	Binary Crossentropy	
Learning rate	Epoch 0-150	0.001
	Epoch 150-200	0.0001
	Epoch 200-300	0.00001

The model was built using the images of the first lot of fruit. Out of the 17 boxes, 12 were used for training, comprising 180 plums and 540 images, while the remaining 5, containing 75 plums and 225 images, were used for internal validation.

During training, an online data augmentation was applied randomly to the input images to prevent overfitting. Two types of aumgentation were performend. On the one hand, an intensity change combined a random baseline (0 to ± 0.1) and a random stretch (from 0.6 to 1.4). On the other hand, spatial changes were performed by horizontal flipping and random rotation (between 0° to 45°).

F1 scores were used to measure the model performance. This measure combines precision and recall to capture better problems with an uneven class distribution. F1 scores better measure the incorrectly classified classes than accuracy by considering both measures of precision and recall (Equation 2).

$$F1\ score = \frac{2 * Precision * Recall}{Precision + Recall}$$

Equation 2. F1 score calculation.

Where precision is the number of positives correctly predicted relative to the number of positive predictions. Recall is the number of positives correctly predicted relative to the number of known positives.

2.4 GradCAM heatmap visualization

GradCAM heatmap visualizations were performed for a more comprehensive understanding of which image regions significantly influence the classification decision in each model, particularly to determine which models truly focus on the bruise for their predictions. These visualizations aid in identifying the image areas that have contributed most to the classification decision. It was performed for all the models for the two best-performing inputs. This analysis aims to identify the image parts that have contributed most to the classification decision. It generates heatmaps representing the activation patterns of the input image (Selvaraju et al., 2020). For each model, the last feature map before the fully connected layers was taken as the target layer.

3. Results

3.1. Important wavelengths

PCA was applied to 161 400 randomly pixel spectra from all images in the training data, and 6 PCs were selected since they represented more than 99 % of the variance.

The absolute loading spectra for the first 6 PCs are illustrated in Figure 3, where the selected wavelength variables are marked, namely 621, 661, 680, 686, 694, 705, 735, 822, 884, and 985 nm.

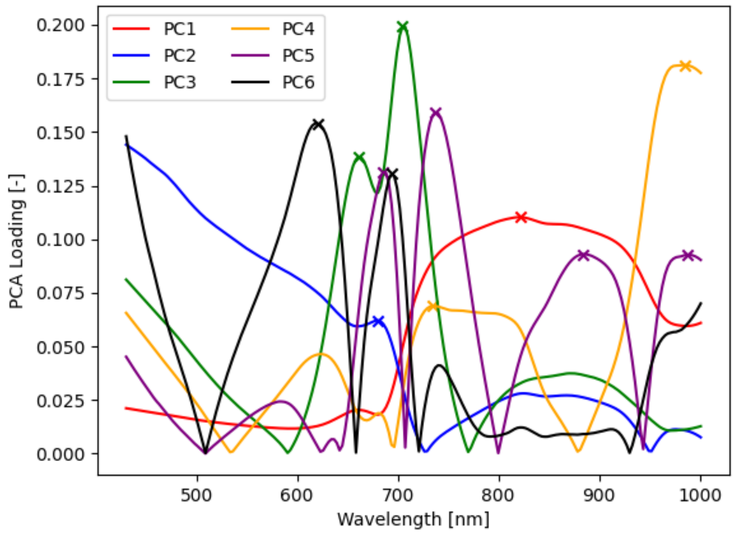


Figure 3. Absolute PCA loading spectra for the first 6 PCs were calculated from the plum spectra in the training set. Crosses indicate selected wavelength variables corresponding to a peak higher than 0.05.

In order to understand how PCA works on the spectral images, Figure 4 shows an RGB image of one of the plums from the test set, the corresponding image at 990 nm, and the corresponding PC score images. While the bruise is not apparent in the RGB image, it can be observed in the image at 980 nm and the score images for PCs 2, 4, and 5. The PC1 score image clearly distinguishes between the fruit and the background, which forms the largest source of variation. The score images for PC2 and PC3 highlight the variation in skin colour, especially for PC2, where the skin colour changes that can be seen in the RGB images. The bruise becomes visually detectable in the PC4 and PC5 images, while the PC6 image shows the peduncle.

The selected wavelengths around 680 nm (661, 680, 686, 694 nm) may be attributed to chlorophyll in the plum tissue (Wellburn, 1994). Khatiwada et al. (2016) also found similar wavelengths to be key in apple bruising and tissue browning using the PLS regression coefficients in his case. They related the range between 550 and 760 nm to the browning caused by the oxidation of polyphenolic compounds resulting from the release of polyphenol oxidase from the busted cells. Similar results were found by Du et al. (2019), who measured polyphenol oxidase in mechanically injured peaches and found a correlation between the close wavelengths and the polyphenol oxidase. Besides, the selected wavelengths in the 688–900 nm range can be linked to the third overtone of the C-H and N-H vibration, where phenolic compounds or soluble solids can be measured (Sun et al., 2021). The wavelength variable at 985 nm can be related to the absorption peak of water around 970 nm (3^{rd} overtone of the stretching vibrations of O-H) (Büning-Pfaue, 2003; Sun et al., 2021) This peak indicates the release of water from the cells due to cell wall rupture and loss of tissue integrity caused by bruising (Martínez-Romero et al., 2003; Serrano et al., 2004). Following the 985 nm wavelength, the 822 nm and 705 nm wavelengths had the highest loadings, being the most important ones. They were then used to create a multispectral image with three false-colour channels. This multispectral image was used as an input to the deep learning networks, and their performance was compared against networks using the full hypercubes (211 wavelengths) and single wavelength images at 985 nm.

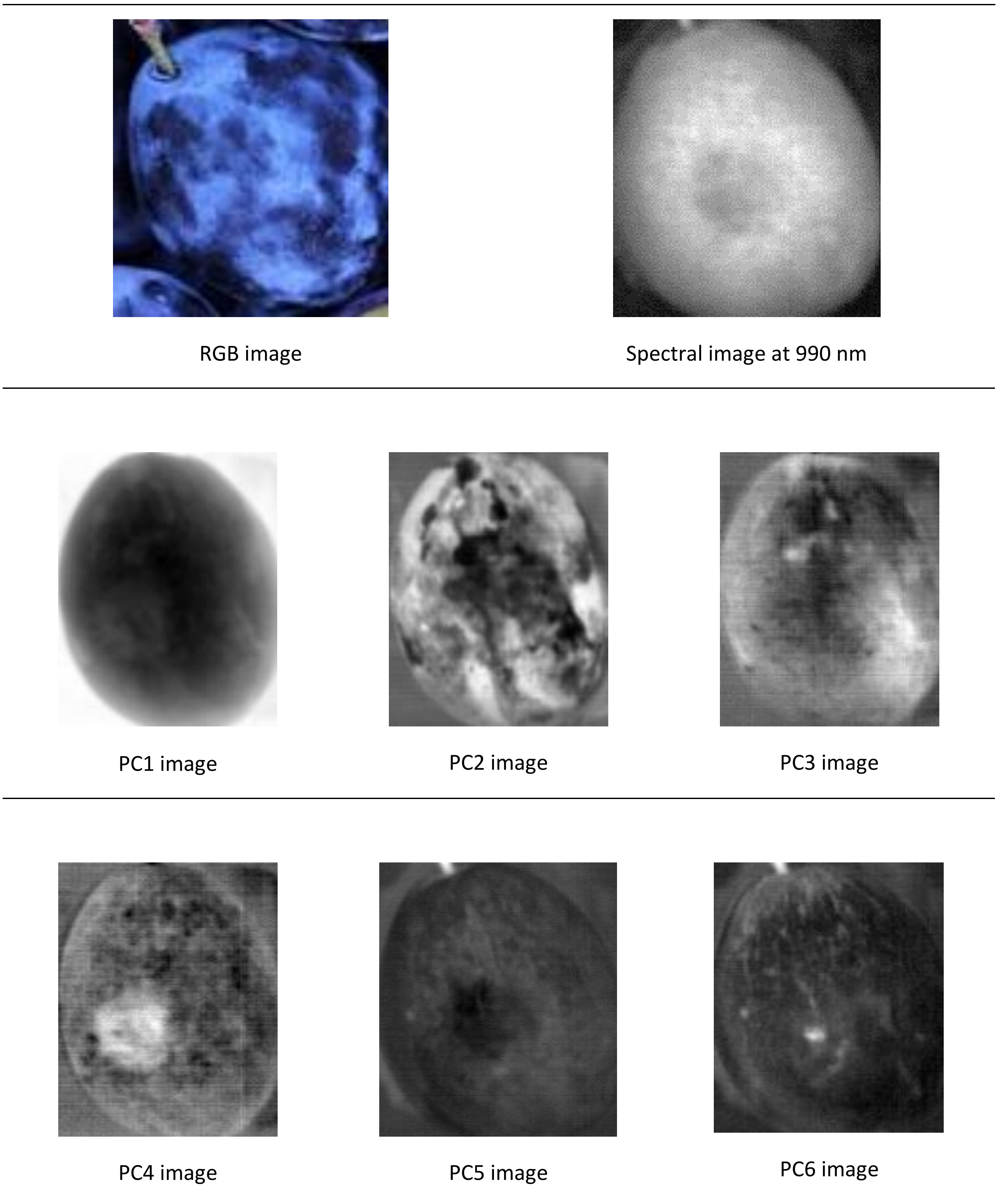


Figure 4. RGB image of a hard-bruised plum sample from the independent test set and its corresponding spectral image at 990 nm (top right), and images for the first six PC images calculated from the hypercube.

3.2. Bruise detection models

The three different architectures, ResNet, the HSCNN model proposed by Varga et al. (2021), and the specifically created 3D-CNN model, were trained using images with all wavelengths, with the 10 most important wavelengths, the 3 most important and only one wavelength. The F1 scores are shown in Table 2, and the confusion matrices in Table 3.

The highest performance was obtained by supplying the full hypercubes with 211 wavelengths to the HSCNN model, resulting in an F1 score of 90 %. From the results listed in Table 2, which shows F1 scores, and Table 3, which shows the accuracies and confusion matrices of the test set, it can be seen that the HSCNN network consistently outperformed the ResNet and the 3D-CNN model. However, it should be noted that the HSCNN model with only three wavelengths as input is very close in performance to the results obtained using the HSCNN using all wavelengths. For the ResNet model, using only three wavelengths, the F1 score was slightly higher than the ResNet model with all wavelengths (88 % and 87 %, respectively). It should be noted that due to the strong influence of random elements when training and selecting a neural network, it can be argued that these slight differences may be negligible given the small sample numbers in the test set (360 samples). The developed 3D-CNN showed results similar to those of ResNet. However, it slightly outperformed ResNet when using all, and 10 wavelength images were used as input (87 % vs. 87 % with all wavelengths and 86 % vs. 81 % when using 10 wavelengths). Regarding training times, there is a clear difference between using the full spectrum or only a limited number of bands**¡Error! No se encuentra el origen de la referencia.**. The time required to train the networks based on the full hypercube was seven times larger than those based on three wavelengths. Similarly, inference time was much lower when using three wavelengths, which is crucial for future industrial in-line applications.

Table 4 explains the average results already commented as presented in Tables 2 and 3, showing the confusion matrices for each of the bruising levels (unbruised, feeble, medium, and hard) for the best performing model (HSCNN) using all wavelengths, 10 wavelengths and 1 wavelength images as input. This was done intentionally to analyse in detail how well each model could distinguish between bruised and unbruised fruit at each specific severity level. It can be seen that unbruised and hard bruised plums can be classified with almost 100 % accuracy, especially in the case of the hard bruised plums, where only when using 1 wavelength one plum was mislabeled. However, in the case of the feeble plums, the accuracy dropped significantly to 63 % when using only one wavelength as an input. In all cases, when using fewer wavelengths as an input, accuracy decreased, and so more mislabeled samples appeared.

Table 5. F1 scores were obtained by the three different model architectures and inputs.

Model	All wavelengths	10 wavelengths	3 wavelengths	1 wavelengths
HSCNN	90.0 %	89.0 %	89.0 %	88.0 %
ResNet	87.0 %	82 %	88.0 %	82.0 %
3D-CNN	87.0 %	86 %	86.0 %	81.0 %

Table 6. Confusion matrices for the three models use all 10, 3, and 1 wavelengths.

Model	Class	All wavelengths		10 wavelengths		3 wavelengths		1 wavelength	
		U	B	U	B	U	B	U	B
HSCNN	U	160	20	160	20	160	20	157	23
	B	17	163	20	160	19	161	20	160
ResNet	U	155	25	146	34	156	24	145	35
	B	21	159	31	149	20	160	31	149
3D-CNN	U	156	24	152	28	153	27	145	35
	B	21	159	23	157	22	158	32	148

U = unbruised plums; B= bruised plums

Table 7. Confusion matrix for each bruising level for the best-performing model using all wavelengths, 10 wavelengths, and one wavelength.

Damage level	Class	All wavelengths		10 wavelengths		3 wavelengths		1 wavelength	
		U	B	U	B	U	B	U	B
Unbruised	U	175	5	175	5	175	6	175	8
	B	5	175	6	174	6	174	7	173
Feeble	U	118	62	116	64	115	65	114	66
	B	65	115	65	115	66	114	68	112
Medium	U	158	22	157	23	156	24	154	26
	B	20	160	21	159	21	159	23	157
Hard	U	180	0	180	0	180	0	169	1
	B	0	180	0	180	0	180	0	180

U = unbruised; B= bruised

Figure 5 shows a more detailed analysis of the test performance for the best-performing model, namely the HSCNN model with the full hypercube as an input and an F1 score of 90 %. It can be seen that the performance increases for higher impact levels. Plums that underwent the highest impact level (hard) were consistently detected as bruised, while detection rates for feeble plums varied between 50 % and 71 %, and medium-bruised plums ranged from 83 % to 92 %. Only 10 % of the unbruised samples measured on day zero were wrongly classified as defective, with no false positives observed on other days. This indicates that the unbruised samples were hard to distinguish from those subjected to the impacts directly after bruising. However, performance decreased over time for samples subjected to feeble and medium impact levels. This decline is likely attributed to the gradual degradation of the cell wall and solubilisation ripening, as demonstrated in persimmons by Tessmer et al. (2016). This degradation leads to a similarity between the fruit's flesh and artificially damaged flesh, making the plum classification more challenging.

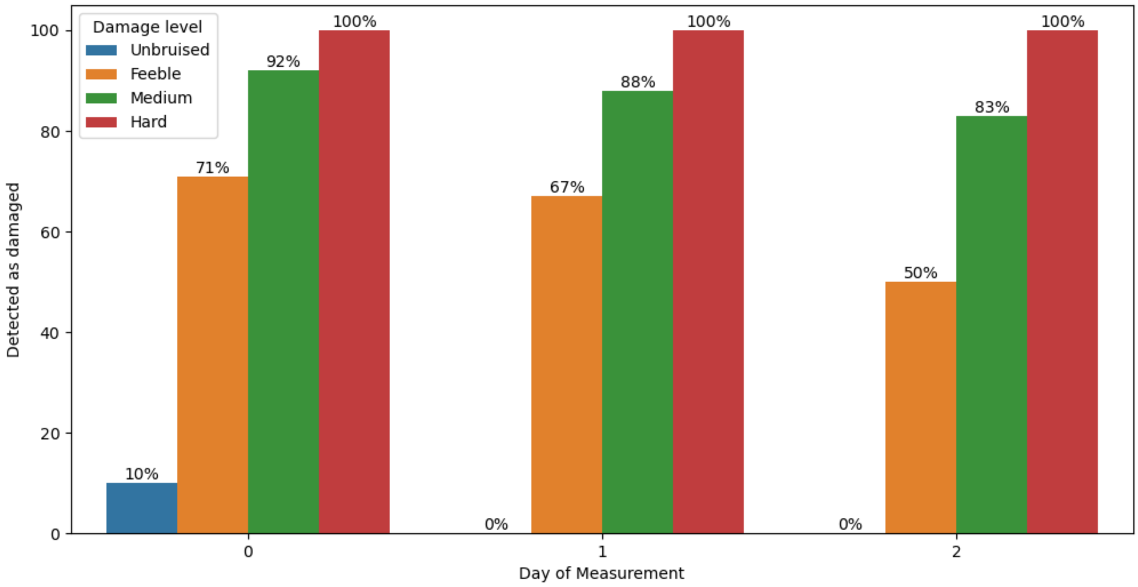


Figure 5. The fraction of plums detected as defective by the HSCNN model with the full hypercube as input as a function of bruising level and the time after impact.

3.3. GradCam analysis

A GradCam was also performed to obtain insight into the model behavior and the information in the images used to make the classification decision. The GradCAM heatmap visualizations for the different networks are illustrated in Figure 6. In this example, a random plum sample from the independent test set exposed to the highest impact level was used for the heatmap visualization. The coloured regions indicate where the model focuses on prediction, with the red-coloured zones showing the highest focus. The heatmaps are overlayed on false colour images reconstructed from either the entire hypercube ('full') at the wavelengths where bruises were most visible (960 nm) or the 3 selected wavelengths based on the PCA loadings ('3WL' selected').

The ResNet network focuses on the entire region of the bruise when given the full hypercube as input while focusing on the edges of the bruise when given an input with only three selected wavelengths. On the contrary, the HSCNN network focuses on the bruise region when trained with the three selected wavelengths, whereas it focuses on the edges of the fruit when using the full image. However, the custom-made 3D-CNN, which is a much smaller and simpler network, may yield comparable results to the other models. Despite this, the 3D-CNN cannot extract deeper image features and concentrate exclusively on identifying bruises in either scenario. Consequently, its reliability diminishes compared to the other models despite similar F1 scores. In general, it can be seen that the GradCAM activations for the HSCNN network are more distributed over the entire plum. In contrast, the ResNet activations are more focused on the bruise itself, which gives more confidence to this architecture concerning its performance on a different dataset, for example, a different variety. However, as the bruised area is the main point of interest on all the heatmaps, it is clear that the HSCNN and ResNet models were not facing a 'Clever Hans' situation, which refers to a scenario where a model seems to perform exceptionally well on a given task, but in reality, relies on unintended cues or artifacts in the data rather than truly learning the desired task (Lapuschkin et al., 2019), which might be the case for the 3D-CNN. That can also be because the 3D-CNN only has 3 convolutional layers, which does not allow the model to go deeper and focus more, and, thus, the specially created model does not provide clear added value or better performance than using transfer learning from the other standard networks.

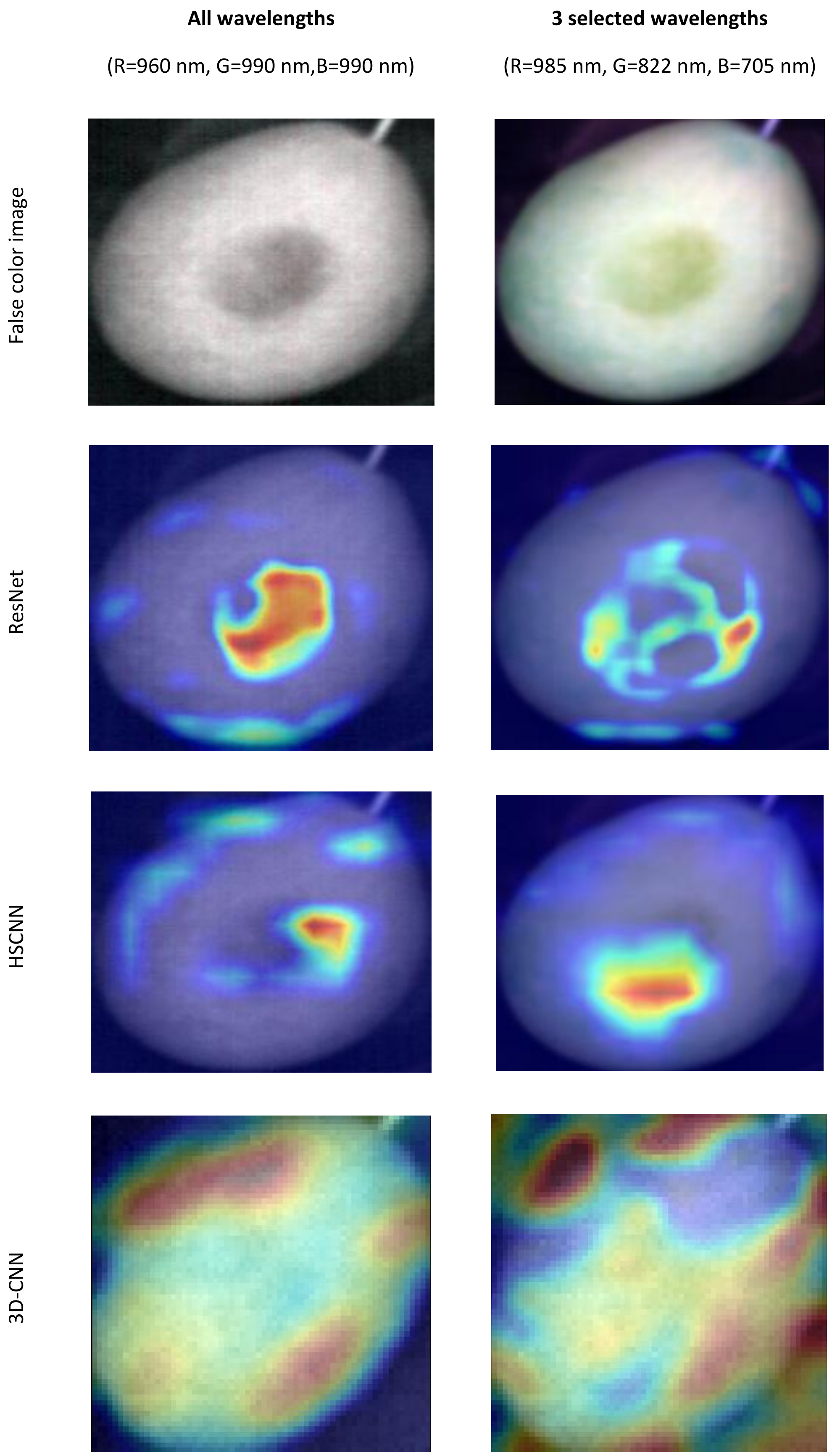


Figure 6. GradCAM heatmap visualization of regions used to classify a plum from h test set bruised with the high impact level by the different models. Red regions are the areas where the model focuses on prediction.

4. Discussion

The bruise overall accuracy detection of 89 %, which reached 100 % in the case of the hard bruised plums with the best model obtained in this study, is in line with those reported for other fruit using traditional machine learning models. For example, 91 % in tomatoes using PLS discriminant analysis (Sun et al., 2021), 86 % in kiwifruit using PCA (Lü & Tang, 2012), 93 % in peaches using XGBoost (Li et al., 2022), 91 % in jujube using PLS discriminant analysis (R. Yuan et al., 2021), and a 96 % and 97 % accuracy in loquat with PCA and XGBoost, respectively (Yin et al., 2022; Munera et al., 2021a). However, this study focused on using the whole image and deep learning with wavelength selection since it might allow fast commercial in-line application more suitable for food industries and sorting lines than pixel-wise classification.

Similar to the different bruising levels considered in this study, (Ye et al., 2018) tested 3 bruising levels in potatoes and reported a minimum accuracy of 88 %. In other studies reported in the literature, 6 bruising levels were used in yellow peaches (Li et al., 2023) and 2 bruising levels in tomatoes (Sun et al., 2021), resulting in accuracy above 88 %. Since different bruising levels were used, those results are the most comparable to this study, and similar results were obtained.

In other studies using neural networks for classification in bruise detection, an accuracy of 92 % was obtained in strawberries using an Autoencoder (Liu et al., 2022), while 97 % accuracy was obtained in apples using hyperspectral imaging together with 3D information and a CNN (Hu et al., 2020). Similar results were obtained in apples with deep learning models such as YOLO and Faster R-CNN (Yuan et al., 2022) and strawberries using VGG architecture (Zhou et al., 2022). Arango et al. (2021) and Yang et al. (2022) used ResNet architecture for bruise detection in apples and nectarines. They obtained classification accuracies of over 97 %, which is higher than the values obtained in this study. However, they only focused on one bruising level, making it challenging to directly compare results from other studies involving different fruits and varied experimental conditions, such as diverse bruising methods and forces applied.

Although this study produced results similar to those reported by other researchers for different fruit types, it is worth highlighting that the overall accuracy was slightly lower. This outcome divergence could be attributed to the specific experimental conditions used, particularly the lower intensity of bruising induced in the present study. The best model obtained almost 100 % accuracy when predicting unbruised plums. However, in the case of the feeble bruising plums, accuracies dropped to 71 % on day 0 and continued decreasing to 50 % on day 2 due to fast ripening, which made the plum internal structure more similar to the bruising area due to the flesh structure changes. Regarding medium-level bruising plums, accuracies were higher: 92 % on day 0 and 84 % on day 2. Moreover, it is essential to consider that comparing bruising observations across different fruit types may have limitations due to their unique structural and compositional characteristics.

It should also be noted that the custom-built model did not provide better results than the transfer learning of widely used networks, while it takes much more time in fine-tuning and training than doing transfer learning. This can most likely be attributed to the limited number of images available for training the network from scratch.

5. Conclusion

This study evaluated the potential of hyperspectral imaging in the 430 to 1000 nm range and deep learning models for early-stage bruise detection in dark-skinned plums. Three CNNs were employed, HSCNN and ResNet, both utilizing transfer learning and a 3D-CNN designed from scratch to create bruise detection models. PCA was applied to accelerate the bruise discrimination process, and ten crucial wavelengths were identified. Notably, the HSCNN model outperformed ResNet and 3D-CNN models, achieving an F1 score of 90 % when fed full images as input. Its performance in bruise detection improved with increasing impact levels, reaching 100 % accuracy in high-impacted plums across all days. Models using only the most important three wavelengths (985, 822, and 705 nm) from the ten selected achieved comparable performance with shorter training times. These models achieved F1 scores exceeding 81 % and up to 88 %, suggesting that classifying low-impact bruised plums based on multispectral images involving a few wavelengths is feasible. The alternative developed holds potential for future in-line inspection using multispectral cameras. In addition, GradCAM heatmaps illuminated key image regions utilized by the models. ResNet activations concentrated on the bruise itself, enhancing confidence in its transferability to different datasets. This study also demonstrated that transferring pre-trained HSCNN and ResNet networks was more effective than training a custom 3D CNN model from scratch.

Acknowledgements

This work was partially funded by projects GVA-PROMETEO CIPROM/2021/014. Salvador Castillo Gironés thanks INIA for the FPI-INIA grant number. PRE2020-094491, partially supported by European Union FSE funds and the COST action CA19145 for granting the research stay at KU Leuven. Remi Van Belleghem thanks the Research Foundation-Flanders (FWO, Brussels, Belgium) for his Ph.D. Fellowship strategic basic research (1S28522N). Sandra Munera thanks the postdoctoral contract Juan de la Cierva-Formación (FJC2021-047786-I) co-funded by MCIN/AEI/10.13039/501100011033 and European Union NextGenerationEU/PRTR. The authors thank Mieke Thoelen from Belorta for providing the packing sheets and Wim Wouters from Fruithandel Romain Wouters & Co. nv for his flexibility in providing the plums.

References

Amigo, J. M. (2019). Chapter 1.1 - Hyperspectral and multispectral imaging: setting the scene. In J. M. Amigo (Ed.), *Hyperspectral Imaging* (Vol. 32, pp. 3–16). Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-444-63977-6.00001-8

Arango, J. D., Staar, B., Baig, A. M., & Freitag, M. (2021). Quality control of apples by means of convolutional neural networks - Comparison of bruise detection by color images and near-infrared images. *Procedia CIRP*, *99*, 290–294. https://doi.org/https://doi.org/10.1016/j.procir.2021.03.043

Arul, V. H. (2021). 5 - Deep learning methods for data classification. In D. Binu & B. R. Rajakumar (Eds.), *Artificial Intelligence in Data Mining* (pp. 87–108). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-820601-0.00001-X

Baranowski, P., Mazurek, W., & Pastuszka-Woźniak, J. (2013). Supervised classification of bruised apples with respect to the time after bruising on the basis of hyperspectral imaging data. *Postharvest Biology and Technology*, *86*, 249–258. https://doi.org/https://doi.org/10.1016/j.postharvbio.2013.07.005

Baranowski, P., Mazurek, W., Wozniak, J., & Majewska, U. (2012). Detection of early bruises in apples using hyperspectral data and thermal imaging. *Journal of Food Engineering*, *110*(3), 345–355. https://doi.org/https://doi.org/10.1016/j.jfoodeng.2011.12.038

Büning-Pfaue, H. (2003). Analysis of water in food by near infrared spectroscopy. *Food Chemistry*, *82*(1), 107–115. https://doi.org/https://doi.org/10.1016/S0308-8146(02)00583-6

Butac, M. (2020). *Plum Breeding*. https://doi.org/10.5772/intechopen.92432

Du, X., Li, X., Liu, Y., Zhou, W., & Li, J. (2019). Genetic algorithm optimized non-destructive prediction on property of mechanically injured peaches during postharvest storage by portable visible/shortwave near-infrared spectroscopy. *Scientia Horticulturae*, *249*, 240–249. https://doi.org/https://doi.org/10.1016/j.scienta.2019.01.057

Elgendy, M., O'Reilly for Higher Education (Firm), & Safari, an O. M. Company. (2020). *Deep Learning for Vision Systems*. Manning Publications Co. https://books.google.com/books/about/Deep_Learning_for_Vision_Systems.html?hl=es&id=6gkLzAEACAAJ

Fazari, A., Pellicer-Valero, O. J., Gómez-Sanchıs, J., Bernardi, B., Cubero, S., Benalia, S., Zimbalatti, G., & Blasco, J. (2021). Application of deep convolutional neural networks for the detection of anthracnose in olives using VIS/NIR hyperspectral images. *Computers and Electronics in Agriculture*, *187*, 106252. https://doi.org/https://doi.org/10.1016/j.compag.2021.106252

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2016-Decem*, 770–778. https://doi.org/10.1109/CVPR.2016.90

Hu, Z., Tang, J., Zhang, P., & Jiang, J. (2020). Deep learning for the identification of bruised apples by fusing 3D deep features for apple grading systems. *Mechanical Systems and Signal Processing*, *145*, 106922. https://doi.org/https://doi.org/10.1016/j.ymssp.2020.106922

Jayasankar, S., Dowling, C., & Selvaraj, D. K. (2016). Plums and Related Fruits. In B. Caballero, P. M. Finglas, & F. Toldrá (Eds.), *Encyclopedia of Food and Health* (pp. 401–405). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-384947-2.00549-3

Keresztes, J. C., Diels, E., Goodarzi, M., Nguyen-Do-Trong, N., Goos, P., Nicolai, B., & Saeys, W. (2017). Glare based apple sorting and iterative algorithm for bruise region detection using shortwave infrared hyperspectral imaging. *Postharvest Biology and Technology*, *130*, 103–115. https://doi.org/https://doi.org/10.1016/j.postharvbio.2017.04.005

Keresztes, J. C., Goodarzi, M., & Saeys, W. (2016). Real-time pixel based early apple bruise detection using short wave infrared hyperspectral imaging in combination with calibration and glare correction techniques. *Food Control*, *66*, 215–226. https://doi.org/https://doi.org/10.1016/j.foodcont.2016.02.007

Khatiwada, B. P., Subedi, P. P., Hayes, C., Carlos, L. C. C., & Walsh, K. B. (2016). Assessment of internal flesh browning in intact apple using visible-short wave near infrared spectroscopy. *Postharvest Biology and Technology*, *120*, 103–111. https://doi.org/https://doi.org/10.1016/j.postharvbio.2016.06.001

Lapuschkin, S., Wäldchen, S., Binder, A., Montavon, G., Samek, W., & Müller, K. R. (2019). Unmasking Clever Hans predictors and assessing what machines really learn. *Nature Communications*, *10*(1), 1–8. https://doi.org/10.1038/s41467-019-08987-4

Li, B., Yin, H., Liu, Y., Zhang, F., Yang, A., Su, C., & Ou-yang, A. (2022). Study on qualitative impact damage of yellow peaches using the combined hyperspectral and physicochemical indicators method. *Journal of Molecular Structure*, *1265*, 133407. https://doi.org/https://doi.org/10.1016/j.molstruc.2022.133407

Li, B., Zhang, F., Liu, Y., Yin, H., Zou, J., & Ou-yang, A. (2023). Quantitative study on impact damage of yellow peach based on hyperspectral image information combined with spectral information. *Journal of Molecular Structure*, *1272*, 134176. https://doi.org/https://doi.org/10.1016/j.molstruc.2022.134176

Li, J., Chen, L., & Huang, W. (2018). Detection of early bruises on peaches (Amygdalus persica L.) using hyperspectral imaging coupled with improved watershed segmentation algorithm. *Postharvest Biology and Technology*, *135*, 104–113. https://doi.org/https://doi.org/10.1016/j.postharvbio.2017.09.007

Li, X., Liu, Y., Jiang, X., & Wang, G. (2021). Supervised classification of slightly bruised peaches with respect to the time after bruising by using hyperspectral imaging technology. *Infrared Physics & Technology*, *113*, 103557. https://doi.org/https://doi.org/10.1016/j.infrared.2020.103557

Liu, Y., Zhou, S., Wu, H., Han, W., Li, C., & Chen, H. (2022). Joint optimization of autoencoder and Self-Supervised Classifier: Anomaly detection of strawberries using hyperspectral imaging. *Computers and Electronics in Agriculture*, *198*, 107007. https://doi.org/https://doi.org/10.1016/j.compag.2022.107007

Loshchilov, I., & Hutter, F. (2019). Decoupled weight decay regularization. *7th International Conference on Learning Representations, ICLR 2019*.

Lü, Q., & Tang, M. (2012). Detection of Hidden Bruise on Kiwi fruit Using Hyperspectral Imaging and Parallelepiped Classification. *Procedia Environmental Sciences*, *12*, 1172–1179. https://doi.org/https://doi.org/10.1016/j.proenv.2012.01.404

Martínez-Romero, D., Castillo, S., & Valero, D. (2003). Forced-air cooling applied before fruit handling to prevent mechanical damage of plums (Prunus salicina Lindl.). *Postharvest Biology and Technology*, *28*(1), 135–142. https://doi.org/https://doi.org/10.1016/S0925-5214(02)00142-4

Munera, S., Blasco, J., Amigo, J. M., Cubero, S., Talens, P., & Aleixos, N. (2019). Use of hyperspectral transmittance imaging to evaluate the internal quality of nectarines. *Biosystems Engineering*, *182*, 54–64. https://doi.org/https://doi.org/10.1016/j.biosystemseng.2019.04.001

Munera, S., Rodríguez-Ortega, A., Aleixos, N., Cubero, S., Gómez-Sanchis, J., & Blasco, J. (2021). Detection of Invisible Damages in 'Rojo Brillante' Persimmon Fruit at Different Stages Using Hyperspectral Imaging and Chemometrics. *Foods*, *10*(9). https://doi.org/10.3390/foods10092170

Qin, J., Chao, K., Kim, M. S., Lu, R., & Burks, T. F. (2013). Hyperspectral and multispectral imaging for evaluating food safety and quality. *Journal of Food Engineering*, *118*(2), 157–171. https://doi.org/https://doi.org/10.1016/j.jfoodeng.2013.04.001

Rodríguez-Ortega, A., Aleixos, N., Blasco, J., Albert, F., & Munera, S. (2023). Study of light penetration depth of a Vis-NIR hyperspectral imaging system for the assessment of fruit quality. A case study in persimmon fruit. *Journal of Food Engineering*, *358*, 111673. https://doi.org/https://doi.org/10.1016/j.jfoodeng.2023.111673

Saeys, W., Nguyen Do Trong, N., Van Beers, R., & Nicolaï, B. M. (2019). Multivariate calibration of spectroscopic sensors for postharvest quality evaluation: A review. *Postharvest Biology and Technology*, *158*, 110981. https://doi.org/https://doi.org/10.1016/j.postharvbio.2019.110981

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2020). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *International Journal of Computer Vision*, *128*(2), 336–359. https://doi.org/10.1007/s11263-019-01228-7

Serrano, M., Martínez-Romero, D., Castillo, S., Guillén, F., & Valero, D. (2004). Role of calcium and heat treatments in alleviating physiological changes induced by mechanical damage in plum. *Postharvest Biology and Technology*, *34*(2), 155–167. https://doi.org/https://doi.org/10.1016/j.postharvbio.2004.05.004

Soliva-Fortuny, R. C., Grigelmo-Miguel, N., Odriozola-Serrano, I., Gorinstein, S., & Martín-Belloso, O. (2001). Browning Evaluation of Ready-to-Eat Apples as Affected by Modified Atmosphere Packaging. *Journal of Agricultural and Food Chemistry*, *49*(8), 3685–3690. https://doi.org/10.1021/JF010190C

Sun, Y., Pessane, I., Pan, L., & Wang, X. (2021). Hyperspectral characteristics of bruised tomatoes as affected by drop height and fruit size. *LWT*, *141*, 110863. https://doi.org/https://doi.org/10.1016/j.lwt.2021.110863

Tessmer, M. A., Besada, C., Hernando, I., Appezzato-da-Glória, B., Quiles, A., & Salvador, A. (2016). Microstructural changes while persimmon fruits mature and ripen. Comparison between astringent and non-astringent cultivars. *Postharvest Biology and Technology*, *120*, 52–60. https://doi.org/https://doi.org/10.1016/j.postharvbio.2016.05.014

Van Zeebroeck, M., Tijskens, E., Liedekerke, P. V., Deli, V., Baerdemaeker, J. D., & Ramon, H. (2003). Determination of the dynamical behaviour of biological materials during impact using a pendulum device. *Journal of Sound and Vibration*, *266*(3), 465–480. https://doi.org/https://doi.org/10.1016/S0022-460X(03)00579-0

Varga, L., Makowski, J., & Zell, A. (2021). *Measuring the Ripeness of Fruit with Hyperspectral Imaging and Deep Learning*. https://doi.org/10.1109/IJCNN52387.2021.9533728

Velásquez C, Prieto F, Palou L, Cubero S, Blasco J, & Aleixos Borrás N. (2023). New Model for the Automatic Detection of Anthracnose in Mango Fruits based on Vis/NIR Hyperspectral Imaging and Discriminant Analysis. *Journal of Food Measurement and Characterization*.

Wellburn, A. R. (1994). The Spectral Determination of Chlorophylls a and b, as well as Total Carotenoids, Using Various Solvents with Spectrophotometers of Different Resolution. *Journal of Plant Physiology*, *144*(3), 307–313. https://doi.org/https://doi.org/10.1016/S0176-1617(11)81192-2

Yang, Y., Wang, L., Huang, M., Zhu, Q., & Wang, R. (2022). Polarization imaging based bruise detection of nectarine by using ResNet-18 and ghost bottleneck. *Postharvest Biology and Technology*, *189*, 111916. https://doi.org/https://doi.org/10.1016/j.postharvbio.2022.111916

Ye, D., Sun, L., Tan, W., Che, W., & Yang, M. (2018). Detecting and classifying minor bruised potato based on hyperspectral imaging. *Chemometrics and Intelligent Laboratory Systems*, *177*, 129–139. https://doi.org/https://doi.org/10.1016/j.chemolab.2018.04.002

Yin, H., Li, B., Liu, Y., Zhang, F., Su, C., & Ou-yang, A. (2022). Detection of early bruises on loquat using hyperspectral imaging technology coupled with band ratio and improved Otsu method. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *283*, 121775. https://doi.org/https://doi.org/10.1016/j.saa.2022.121775

Yuan, R., Liu, G., He, J., Wan, G., Fan, N., Li, Y., & Sun, Y. (2021). Classification of Lingwu long jujube internal bruise over time based on visible near-infrared hyperspectral imaging combined with partial least squares-discriminant analysis. *Computers and Electronics in Agriculture*, *182*, 106043. https://doi.org/https://doi.org/10.1016/j.compag.2021.106043

Yuan, Y., Yang, Z., Liu, H., Wang, H., Li, J., & Zhao, L. (2022). Detection of early bruise in apple using near-infrared camera imaging technology combined with deep learning. *Infrared Physics & Technology*, *127*, 104442. https://doi.org/https://doi.org/10.1016/j.infrared.2022.104442

Zhou, X., Ampatzidis, Y., Lee, W. S., Zhou, C., Agehara, S., & Schueller, J. K. (2022). Deep learning-based postharvest strawberry bruise detection under UV and incandescent light. *Computers and Electronics in Agriculture*, *202*, 107389. https://doi.org/https://doi.org/10.1016/j.compag.2022.107389

Zhu, S., Zhou, L., Gao, P., Bao, Y., He, Y., & Feng, L. (2019). Near-Infrared Hyperspectral Imaging Combined with Deep Learning to Identify Cotton Seed Varieties. *Molecules*, *24*(18). https://doi.org/10.3390/molecules24183268

CHAPTER IV. Raman spectroscopy for multi-label identification of common apple pesticide mixtures using CNNs and Gradient-weighted Class Activation Mapping.

Raman spectroscopy for multi-label identification of common apple pesticide mixtures using CNNs and Gradient-weighted Class Activation Mapping

Salvador Castillo-Girones[a], Quentin Arnould[b], Juan Gómez-Sanchis[c], Jose Blasco[a], Olivier Pigeon[d], Vincent Baeten[b], Juan Antonio Fernández Pierna[b]

[a] Centro de Agroingeniería, Instituto Valenciano de Investigaciones Agrarias (IVIA), CV-315, km 10,7, 46113 Moncada (Valencia), Spain.

[b] Quality and authentication of agricultural products Unit. Centre wallon de Recherches agronomiques (CRA-W). Bâtiment Carson, Rue du Bordia 11, 5030 Gembloux, Belgium.

[c] Departamento de Ingeniería Electrónica, IDAL. Universitat de València, Av. Universitat, s/n, 46100 Burjassot (Valencia), Spain.

[d] Protection, control products and residues Unit. Centre wallon de Recherches agronomiques (CRA-W). Bâtiment Carson, Rue du Bordia 11, 5030 Gembloux, Belgium.

Food Control. 178, 111460

https://doi.org/10.1016/J.FOODCONT.2025.111460

Abstract

Fruits on the market must be free of harmful pesticides, but residues often persist, posing a risk to public health. Detecting these residues efficiently is crucial. However, traditional methods are slow and lack real-time capability. Besides, the detection challenge becomes even more difficult when pesticides are used in mixtures, and Raman spectroscopy offers a promising real-time solution for identifying pesticide residues. This study investigates the detection of commonly used pesticide active compounds on apples in complex mixtures via Raman spectroscopy and a multilabel classification approach, addressing the challenge of identifying active compounds within complex mixtures. Models were also tested on independent, entirely different mixtures to assess their limitations and robustness.

Unique spectral fingerprints were established for pure active compounds, which enabled differentiation in mixture analyses. Dimensionality reduction techniques (t-SNE and PCA) effectively distinguished pure compounds from mixtures. Three machine learning models: Partial Least Squares (PLS), Support Vector Machine (SVM), and a 1-D Convolutional Neural Network (1-D CNN) were trained to identify active compounds in mixtures. All models classified key compounds with high F1 scores, with Captan detected with F1 scores higher than 99%, but some compounds, such as Folpet or Mancozeb, showed poor predictions. The 1-D CNN achieved the best performance with the lowest Hamming Loss, and Grad-CAM analysis confirmed it identified relevant spectral regions. This work demonstrates the potential of machine learning-enhanced Raman spectroscopy for effective pesticide monitoring, regulatory compliance, and food safety.

Keywords: Raman; pesticides; apple; multilabel classification; GradCAM

1. Introduction

The presence of pesticide residues is of high concern to consumers, who demand fruit free of pesticide residues. Chemical pesticides have been used for decades to fight against diseases during agricultural product growth, harvest and storage (Kumar Hazra et al., 2019; Yusoff et al., 2016). Although pesticides are designed to offer high specificity of action, their use could lead to unwanted effects, such as the persistence of residual toxins in the environment and contamination of water resources with degradation of flora and fauna (Kaur et al., 2024), and might cause health problems. Exposure to these pesticides through various routes has been linked to neurological, respiratory and cancer diseases. In the last decade, the issue of pesticide residues in agricultural products has attracted intense public attention worldwide (Scorza et al., 2023). The 2021 EU report on pesticide residues in food, prepared by EFSA, provides an overview of the official control activities on pesticide residues carried out in the EU member states, Iceland and Norway (EFSA, 2021). The number of samples analysed for pesticide residues is around 96,000 annually for the entire European Union, and the percentage of Maximum Residue Limits exceedance was around 2.3 % in 2019. EFSA concluded that chronic dietary exposure to the 182 pesticide residues was studied based on current scientific knowledge. is unlikely to pose any health concern to EU consumers. However, it is recommended that member states keep monitoring several food commodities in their national programmes with an analytical scope that is as broad as possible. Considering this situation, it is reasonable for food supply chain operators to develop reliable, non-destructive, and rapid control methods to check the quality of their own products.

A fundamental requirement for pesticide monitoring is a well-established detection method (CIPAC, 2020). The chosen detection technology must possess several key characteristics, such as good sensitivity and reproducibility. Over the last decades, this important requirement has been fulfilled by various chromatographic-based techniques (such as LC/GC-MS, HPLC and TLC). In the literature, pesticides have been studied and analysed by different methods and equipment (Tsagkaris et al., 2021), such as gas chromatography (GC) (Adou et al., 2001), liquid chromatography (HPLC) (Marczewska et al., 2019; McGarvey, 1993; Miszczyk et al., 2015), both with advantages in low detection limit, or mass spectrometry (MS) (Botitsi et al., 2011; Yang et al., 2018), among others. One drawback of these techniques is their slow sample analysis and the need for various sample extraction and preparation methods that require highly skilled professionals. Additionally, certain pesticides can only be analysed using specific equipment or detectors, which are costly to acquire and maintain. While these methods have advanced to enhance detection capabilities, other technologies have emerged that offer additional benefits, such as faster detection, simpler protocols, on-site sampling, reduced costs and portability. These technologies include vibrational spectroscopy methods (Soltani Nazarloo et al., 2021).

Among the vibrational spectroscopic methods, Raman spectroscopy is gaining attention in pesticide residue detection. This technique is based on the inelastic scattering of light. It measures the fundamental vibrations of functional groups, providing well-resolved bands of fundamental vibrational transitions and, thus, helpful information on the molecular structure of compounds.

Although chromatographic methods and Raman spectroscopy are both powerful analytical tools, but they differ in several key aspects. Chromatography is known for its extremely low detection limits, often reaching trace or ultra-trace levels, making it ideal for applications requiring high sensitivity and specificity. However, it is time-consuming, often taking from several minutes to hours per sample due to the need for sample preparation and separation. Chromatographic systems are also expensive to purchase and maintain, requiring specialized consumables and skilled operators (Lewis, 2023; Snyder et al., 2010).

Raman spectroscopy can achieve low detection limits for certain analytes, though generally not as low as chromatography. Its main advantages are speed, where measurements can be completed in seconds to minutes, and practicality, as it often requires little or no sample preparation and can be performed on-site using portable devices. The operational costs are generally lower than those of chromatography, though high-quality instruments can be costly. Challenges remain with reproducibility and interference from complex samples (Agarwal & Atalla, 2020; Cialla-May et al., 2017).

Raman is widely used in chemistry to identify and characterise substances and compounds (Kazlagić & Omanović-Mikličanin, 2020). The main advantage of Raman is its ability to provide molecular fingerprint specificity for each distinct molecule/analyte. This technique could be used to analyse dry extracts or liquids to detect pesticide residues. Raman spectroscopy is fast, the detection limits can be very low, and many different groups of samples can be analysed. Thus, Raman spectroscopy can be an alternative tool for detecting pesticides in agriculture (Méndez-Albores & Hernández, 2018; Reyes et al., 2004).

Previous studies have demonstrated the potential of Raman spectroscopy in the detection of pesticides. For example, Hou et al. (2015) successfully identified pesticides such as isocarbophos, phorate, deltamethrin, and imidacloprid in tea and apples using Surface-Enhanced Raman Spectroscopy (SERS) with gold nanoparticles, with remarkable sensitivity in detecting concentrations as low as 0.005 µg/mL in methanol extracts and 0.02 µg/cm^2 directly in tea leaves and apple peels. Similarly, X. Wang et al. (2014) applied SERS to detect phosmet and disulfoton pesticides, which are commonly used in oranges at concentrations of 8.25 µg/kg and 39.7 µg/kg, respectively, using an Au nanoparticle solution. While these studies focused on individual pesticides, other research has extended the application of Raman to more complex scenarios involving mixtures, as seen in the current study. For instance, Li et al. (2019) demonstrated the ability to predict and identify chlorpyrifos and acetamiprid, both individually and in mixtures, using a silver-enhanced SERS solution after extracting the pesticide from the apples using a water-acetone solution, even at low mixture concentrations (0.1 g/kg). Other advancements include the work of L. Wang et al. (2022), who used silver nanoparticles to detect chlorpyrifos and 2,4-dichloro phenoxy acetic acid simultaneously directly on an extracted skin where the nanoparticles were applied at concentrations below the maximum residue levels set by EU legislation.

This study investigates the detection of common pesticide active compounds in complex mixtures under controlled conditions using Raman spectroscopy combined with a multilabel classification approach. Initially, unique spectral fingerprints of pure active compounds were identified, followed by the analysis of complex mixtures. Two exploratory tools, the t-distributed stochastic neighbour embedding (t-SNE) and the Principal Component Analysis (PCA), were employed and compared to visualise the separation between pure active compounds and mixtures.

Traditionally, pesticide detection studies have focused on identifying specific compounds through standard machine learning (ML) methods. For example, Mandrile et al. (2018) used Partial Least Squares (PLS) regression to detect pyrimethanil residues in apples, achieving a mean error of 18.7 % across six concentration levels (1 mgL^{-1} to 40 mgL^{-1}). Although these methods effectively identify individual compounds, traditional ML models encounter limitations in analysing complex pesticide mixtures, where nonlinear relationships complicate prediction. Artificial Neural Networks (ANNs) have shown promise in addressing this challenge due to their capacity to model complex nonlinear patterns (Elgendy, 2020).

To improve detection capabilities, three machine learning approaches, PLS, a Support Vector Machine (SVM), and a 1-D Convolutional Neural Network (1-D CNN), were trained to perform multilabel classification, aiming to detect nine different active compounds in complex mixtures. A Gradient-weighted Class Activation Mapping (Grad-CAM) analysis was performed using the 1-D CNN to identify specific spectral regions significantly influencing classification decisions. This was particularly important for complex mixtures in the independent test set. The aim was to ensure the model worked correctly and identify the most crucial spectral regions.

This approach facilitates the examination of spectral variations, improves detection capacity, and applies multilabel classification to assess the feasibility of identifying individual active compounds in mixtures. This study challenges conventional methods that typically focus on detecting a single compound by examining nine commonly used apple active compounds, both individually and in complex mixtures. The research aimed to fill a gap in identifying active compounds by using these complex mixtures, which complicate the identification process. Furthermore, it evaluated the models using entirely different independent complex mixtures to assess the limitations of identifying active compounds and to ensure the robustness of the models. The presence of specific compounds in each mixture was determined through multilabel classification, contributing to advances in detection capabilities beyond controlled environments and providing valuable insights for future studies.

2. Materials and methods

2.1. Pesticides and sample preparation

The active compounds used in this work (Table 1) were selected among the most commonly used on apples. Table 1 shows the active compounds utilised, their commercial name, concentration, and the official recommended user concentration for Belgian farmers (Phytoweb, 2024) from SPF Santé publique, Sécurité de la Chaîne Alimentaire et Environnement.

Samples were prepared by mixing the active compounds with water to achieve the maximum residue level shown in Table 1. Later, each formulation was sprayed on a stainless steel plate and left to dry for half an hour for later spectra collection.

First, samples containing individual active compounds were prepared. After this, Three active compound mixtures containing different active compounds (Table 2) were created to assess Raman's ability to detect multiple active compounds simultaneously and train the models with individual active compounds and mixtures. Later, the samples were measured with the Raman microscope.

In order to obtain an independent dataset for testing the models, three new mixtures (A, B, and C) were prepared to evaluate the performance of the models in detecting the active compounds present in each formulation, in data not seen before by the models. The active compounds and corresponding formulations for these mixtures are detailed in Table 3.

A total of 16 measurements were taken for each sample, with 10 samples created for each condition. This resulted in 160 measurements for each active compound and mixture, leading to an overall dataset of 2,400 spectra. Of these, 1,920 spectra comprised the training set, while 480 spectra made up the independent test set, which included mixtures A, B, and C.

Table 8. Characteristics of the pesticides utilised in the experiment.

Active compound	Commercial name	Concentration	Maximum residue level (mg/kg)	Recommended user concentration (L product/L water)
Difenoconazole	REVUS TOP	250 g/mL	0,6	0.0003
Captan	CAPTAN TC	95 %	10	0.0034
Captan	CAPTABELLOS	800 g/kg	10	0.004
Folpet	FOLPAN	80 %	0,3	0.0033
Mancozeb	MANCOPLUS	75 %	5	0.004
Pyrimethanil	PYRUS	400g/L	15	0.0013
Spirotetramat	VSM SPIROTETRAMAT	100 g/L	1	0.005
Tebuconazole	TARCZA	250 g/L	0,3	0.0013
Thiabendazole	TECTO	500 g/L	4	0.002
Captan + Trifloxystrobin	BROCELIAN	600 g/kg 40 g/kg	15	0.0053

Table 9. Characteristics of data of the mixtures of pesticides utilised for training

Mixture	Active compound	Concentration	Formulation
Captan + Trifloxystrobin	Captan +	600 g/kg	WG
	Trifloxystrobin	40 g/kg	
1	Thiabendazole	500 g/L	SC
	Spirotetramat	100 g/L	SC
	Difenoconazole	250 g/mL	SC
	Mancozeb	75 %	WG
	Captan	95 %	TC
	Captan +	600 g/kg	WG
	Trifloxystrobin	40 g/kg	
2	Folpet	80 %	WG
	Tebuconazole	250 g/L	EW
	Captan	95 %	TC
	Captan	800 g/kg	WG
	Captan +	600 g/kg	WG
	Trifloxystrobin	40 g/kg	
	Pyrimethanil	400 g/L	SC

Legend: EW: Oil-in-Water Emulsion (liquid formulation where oil droplets containing the active ingredient are emulsified in water), SC: Suspension Concentrate (liquid formulation containing solid particles suspended in a liquid medium), TC: Technical Concentrate (A technical-grade product containing the active ingredient, typically used to manufacture formulations), WG: Water-Dispersible Granules (Dry granules that disperse in water to form a uniform suspension)

Table 10. Characteristics of the mixtures of pesticides utilised for testing.

Mixture	Active compound	Concentration	Formulation
A	Thiabendazole	500 g/L	SC
	Difenoconazole	250 g/mL	SC
	Folpet	80 %	WG
	Captan	95 %	TC
	Pyrimethanil	400 g/L	SC
	Captan +	600 g/kg	WG
	Trifloxystrobin	40 g/kg	
B	Spirotetramat	100 g/L	SC
	Mancozeb	75 %	WG
	Tebuconazole	250 g/L	EW
	Captan	95 %	TC
	Captan	800 g/kg	WG
	Captan +	600 g/kg	WG
	Trifloxystrobin	40 g/kg	
C	Folpet	80 %	WG
	Mancozeb	75 %	WG
	Captan	95 %	TC
	Pyrimethanil	400 g/L	SC

Legend: EW: Oil-in-Water Emulsion (liquid formulation where oil droplets containing the active ingredient are emulsified in water), SC: Suspension Concentrate (liquid formulation containing solid particles suspended in a liquid medium), TC: Technical Concentrate (A technical-grade product containing the active ingredient, typically used to manufacture formulations), WG: Water-Dispersible Granules (Dry granules that disperse in water to form a uniform suspension)

2.2. Raman spectra acquisition

The pure active compounds, mixtures 1 and 2, and the Captan + Trifloxystrobin mixture were measured in a first experiment. 160 measurements of each active compound and of each mixture (1 and 2) and of Captan + Trifloxystrobin were taken (a total of 1,920 measurements). Mixtures A, B and C were created in a second experiment in order to test the already-created models with completely independent data, and 160 measurements per mixture were taken (a total of 480 spectral measurements). In total, 2400 spectral measurements were made, 75 % for training and the remaining 25 % as an independent test set.

Raman spectra were acquired on a Confocal Raman Microscope Senterra II spectrometer (Bruker Optics, Ettlingen, Germany – Figure 1) with a 785 nm diode array laser and a thermoelectrically cooled CCD detector, operating at -65 °C. Each pesticide sample was manually placed on a stainless steel plate for spectra collection. After letting the samples dry for 10 minutes at room temperature (~20 °C), spectra were acquired with an integration time of 10 s and 5 co-additions using a 100 mW laser. Raman intensity was recorded 3,650 to 50 cm^{-1} with a spectral resolution of 4 cm^{-1}. OPUS 7.8 Software (Bruker Optics, Ettlingen, Germany) was used for spectral data acquisition.

Figure 7. Raman microscope utilised in the experiment.

2.3. Data processing and analysis

After data acquisition, Standard Normal Variate (SNV) spectral pre-processing was performed. It normalises each spectrum by subtracting its mean and dividing by its standard deviation, reducing light scattering effects and making spectra more comparable, bringing out the true chemical information in the data by minimising variations caused by scattering. (Fearn et al., 2009).

To visualise if there are differences between the spectra of each pesticide, Principal Components Analysis (PCA) and the algorithm for data visualisation “t-distributed Stochastic Neighbor Embedding” (t-SNE) were used in combination as described by Stevens et al. (2024). PCA transforms the original Raman spectra data into a set of orthogonal principal components that capture the maximum variance in the data. The upper principal components were then used to project the data into a lower-dimensional space. By plotting the data against these principal components, PCA allows the differences between the Raman spectra of different pesticides to be visualised, revealing patterns or clusters that indicate variations between them (Greenacre et al., 2022). t-SNE allows representing in a low-dimensional vectorial space (usually two dimensions) a set of data points located in a high-dimensional space. In this case, each of these dimensions corresponds to one spectral variable. The more similar the two data points in the high-dimensional space, the more likely they will be represented close to each other in the t-SNE space. When plotting the t-SNE representation, the separation between groups indicates differences between those groups and, thus, spectral differences (Stevens et al., 2024).

After SNV correction, PCA was applied to the dataset, capturing substantial variance across a comprehensive set of principal components (PCs). Subsequently, the dominant but irrelevant factors, typically observed among the initial PCs, were identified by plotting the different PCA loadings for each wavelength. This was then used to analyse all PCs' scores, except those related to the irrelevant factors. This technique allows the t-SNE map to highlight the most relevant factors.

All data processing and analysis were performed using Python 3.9 with NumPy, Pandas, Sci-kit learn, Keras, Math and Scipy libraries.

2.4. Multilabel classification model creation

For training the model, the active compounds, mixtures 1 and 2, and the Captan + Trifloxystrobin mixture were used for model creation, whereas mixtures A, B and C were used for testing the models as an independent test set.

Two commonly used classical Machine Learning (ML) models and a Deep Learning (DL) model (a 1-D CNN) were created: a multilabel Partial Least Squares Discriminant Analysis (PLS-DA) model, a Support Vector Machine (SVM) model and a 1-D Convolutional Neural Network (1-D CNN). PLS-DA is a simple model that was used for its ability to reduce dimensionality and maximise class separation, making it particularly suitable for high-dimensional and noisy data. SVM was chosen for its ability to find the optimal hyperplane for class separation, demonstrating its effectiveness in multiclass classification. Finally, the deep learning 1-D CNN was used to analyse complex data using convolutional filters, excelling in capturing non-linear patterns. Each model was selected based on its strengths and in order to experiment with different model complexities.

To optimise the parameters and hyperparameters of the ML models, all the possible hyperparameters available were tested (for that, GridsSearchCV from the Scikit-learn library was used), and the parameters that minimised the prediction error were chosen. In order to perform multilabel classification, the PLS and SVM models were trained using scikitlearn MultiOutputClassifier and scikit-multilearn, which trains a separate binary classifier for each label, treating the problem as multiple independent binary classification tasks. In contrast, Keras was used to build the 1-D CNN, where different architectures were tested (from 3 to 10 convolutional layers and from 1 to 3 final dense layers, and also testing different numbers of neurons per layer using GridsearchCV until a suitable one was found for this task based on the test results. The structure of the 1-D CNN model created for multilabel pesticide prediction comprises nine convolutional and max pooling layers with 1024, 512, 256, 128, 64, 32, 32, 16 and 16 filters, respectively, with kernel size 2 and Rectified Linear Unit (ReLu) as activation function, and two fully connected dense layers. The first one has 32 neurons and a ReLu activation function, and the output layer has 9 neurons in the last layer, one for each pesticide and a sigmoid activation function. Models were trained using Adam optimiser and binary crossentropy as loss function, which allows, as in the case of the other ML models to have binary classification for each label, treating the problem as multiple independent binary classification tasks (Elgendy, 2020).

For PLS-DA, the lowest RMSE was achieved with 19 latent variables. For SVM, the best parameters were found using a "balanced" class weight, gamma of 1, C of 0.1, and a "linear" kernel for all labels except for Folpet, Mancozeb, Tebuconazole and Trifloxystrobin predictions, where a polynomial kernel of degree 3 was used.

To assess the performance of the multilabel classification results, the Hamming Loss was calculated for each model. Hamming Loss is a metric used in multilabel classification to determine the performance of a model. It quantifies the proportion of incorrectly predicted labels (Herrera et al., 2016). Additionally, F1 scores were calculated for all the different compounds present in the mixtures for each model to measure the model's performance. This was done to identify which active compounds are easier to distinguish in the mixtures and determine which model best predicts each active compound.

F1 score combines precision and recall to capture problems better with an uneven class distribution. F1 scores better measure the incorrectly classified classes than accuracy by considering both measures of precision and recall (Equation 1).

$$F1\ score = \frac{2 * Precision * Recall}{Precision + Recall}$$

Equation 4. F1 score calculation

Precision is the number of positives correctly predicted relative to the number of positive predictions. Recall is the number of positives correctly predicted relative to the number of known positives.

Besides, Gradient-weighted Class Activation Mapping (Grad-CAM) was performed on the CNN. It is a heatmap visualisation commonly used to provide explainability in CNNs when working with images to see which part of the image is activating the last convolutional layer by focusing on it. “Gradient-weighted” refers to the use of gradients to determine the importance of each feature map in the last convolutional layer of a neural network. These gradients are used to produce a heatmap that highlights the regions of an input image that are most important for network prediction.

In this work, the visualisation was conducted with the 1-D CNN model to identify the spectral bands that contribute the most to the classification decision and provide valuable information on model performance, such as specific bands associated with pesticides rather than non-useful parts of the spectra. GradCAM creates heatmaps that represent the activation patterns of the input data. The analysis targets the last convolution layer's feature map before the fully connected layers (Selvaraju et al., 2020). Some authors, such as (Tewes et al., 2023), have used GradCAM on Raman spectra of fungal spores and carotenoid-pigmented microorganisms using a 1-D CNN to determine which spectral regions are responsible for accurate classification.

3. Results

3.1. Pesticide spectra characterisation

Figure 3 displays the spectra of the pure active compounds analysed in this study. Each active compound has a unique fingerprint, making them easily distinguishable from each other due to their characteristic scattering bands. However, as shown in Figure 4, in the case of the mixtures, their characteristic band scattering signals appeared weaker than the pure signals. This observation aligns with the conclusions of Li et al. (2019), who are working on detecting chlorpyrifos and acetamiprid pesticides in apples using a silver-enhanced solution. Despite this, the characteristic scattering bands of the active compounds are still present in the spectra of the mixtures, as observed by L. Wang et al. (2022) when using Raman to identify Chlorpyrifos and 2,4-dichlorophenoxyacetic acid and mixtures of them.

Captan shows characteristic scattering bands in the spectral range between 800 to 50 cm^{-1}. Weak scattering bands corresponding to Captan can be observed in all mixtures (Figure 4), particularly in mixtures 1 and 2, as well as in the mix of Captan + Trifloxystrobin. In addition, Thiabendazole displays an intense scattering band around 1,600 cm^{-1}. A similar scattering of bands on a smaller scale can be observed in mixtures A and 1. Similarly, Pyrimethanil shows a strong scattering band at 12,000 cm^{-1}, which can be faintly observed in mixtures A and 2 due to its low concentration in the final mix. Furthermore, Folpet shows distinct high scattering bands between 700 to 500 cm^{-1}, at 1,000 cm^{-1}, and around 1,800 cm^{-1}. These scattering bands are also present in mixture 2. In mixtures A and C, in which this compound is also present, the scattering bands are present but almost imperceptible.

When comparing the different mixtures, minor differences are found between mixtures B and C. Mixture B shows a higher number of scattering bands between 1000-50 cm^{-1}, probably due to the higher number of pesticides present in the mixture.

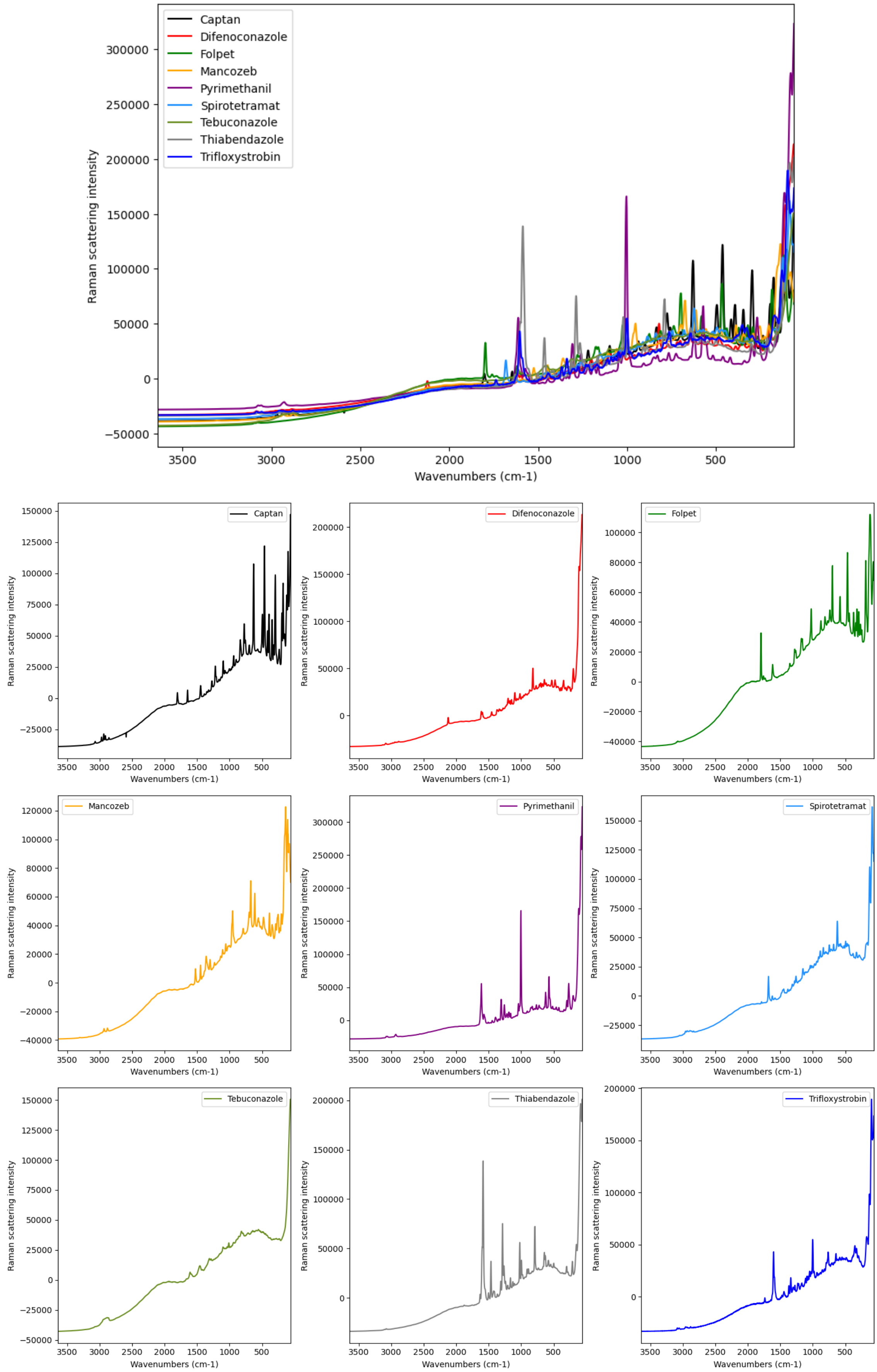


Figure 8. Mean spectra with SNV pretreatment for each of the nine pure pesticides together (Figure 3 top) and independently (Figure 3 bottom).

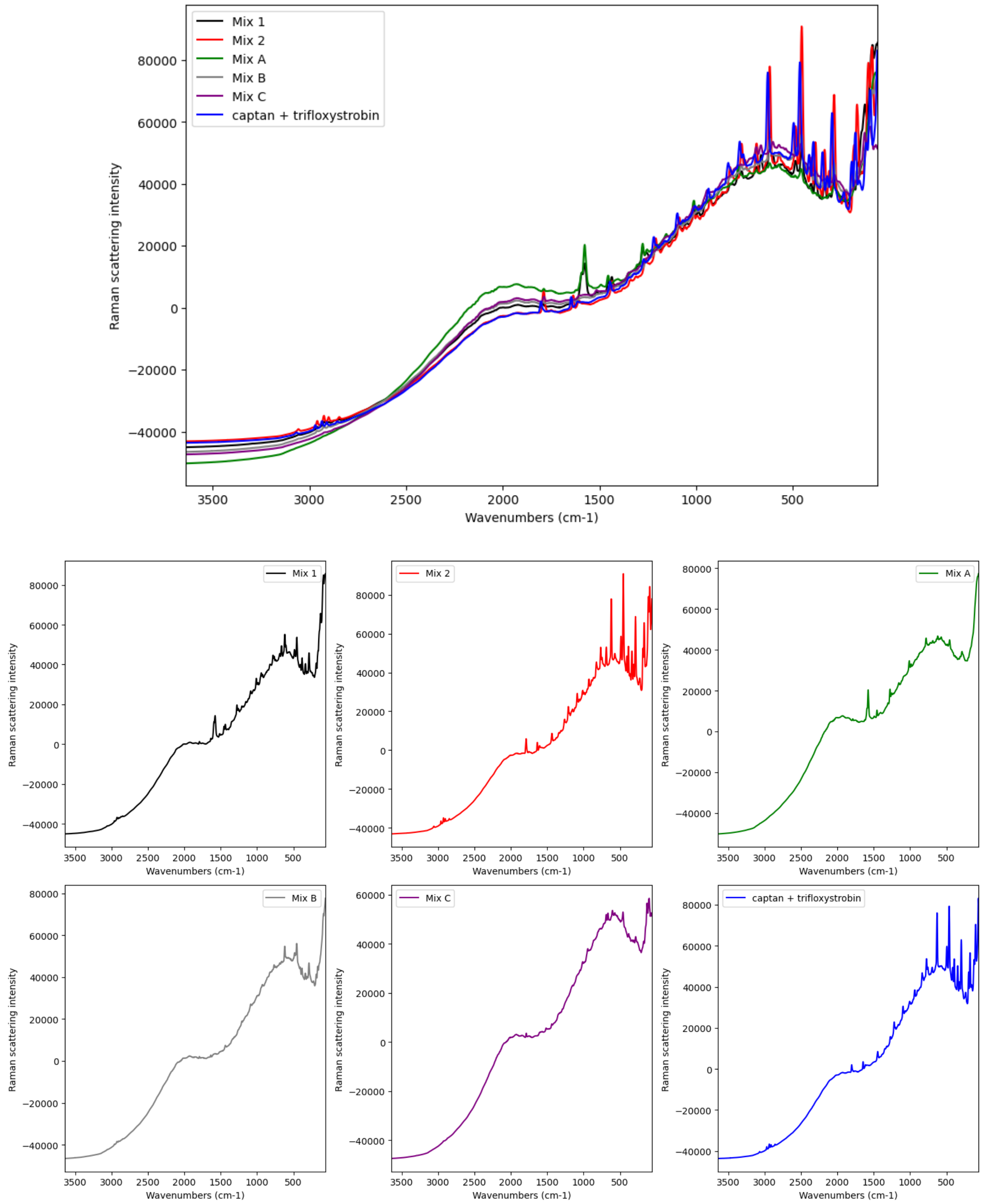


Figure 9. Mean spectra for each of the six pesticide mixtures with SNV pretreatment together (Figure 3 top) and independently (Figure 3 bottom).

The Raman spectral region from 2500 to 50 cm^{-1} reveals important details about the chemical bonds and functional groups found in the active ingredients used in this study. Each part of this region highlights different types of molecular vibrations, which help in identifying and analysing these compounds. The higher wavenumber region (2500–2000 cm^{-1}) typically shows vibrations from triple bonds, but these are uncommon in most of these pesticides, which mostly contain aromatic rings, heterocycles, and various functional groups rather than triple bonds. More relevant is the 2000–1500 cm^{-1} range, where stretching vibrations of C=O and C=N bonds appear. For example, captan and folpet, both phthalimide fungicides, contain imide groups that show strong C=O stretching bands in this region. Similarly, triazole fungicides such as difenoconazole and tebuconazole display characteristic C=N and aromatic C=C stretching vibrations between 1600 and 1500 cm^{-1} due to their heterocyclic structures (Movasaghi et al., 2007; Turner, 2024).

The fingerprint region (1500–500 cm^{-1}) is especially useful for distinguishing these pesticides. It contains unique patterns from ring vibrations, skeletal motions, and bending of bonds. The aromatic and heterocyclic rings in difenoconazole, tebuconazole, thiabendazole, and pyrimethanil produce distinct bands between 1600 and 1200 cm^{-1} for ring stretching and 800–600 cm^{-1} for out-of-plane deformations. Sulfur-containing fungicides like mancozeb exhibit strong S–S and C–S stretching, below 600 cm^{-1}, while spirotetramat, with its spirocyclic core and ester functionalities, shows characteristic C–O and C–C stretching bands in the 1300–1000 cm^{-1} range. Finally, the lowest region (500–50 cm^{-1}) is linked to vibrations involving heavier atoms and metal–ligand bonds, which is important for pesticides like mancozeb that contain metal ions and for identifying different crystalline forms of these chemicals. This low-frequency information adds another layer to the structural analysis of these compounds (Hayes, 2010; Movasaghi et al., 2007; Rolfe et al., 2016; Turner, 2024).

3.2. Spectral analysis

To visualise high-dimensional spectral datasets and to establish differences between pure active compounds and mixtures, which appear very similar in some cases when examining the spectral plots, as described in the previous section, t-SNE and PCA were applied to transform the spectral data of pure active compounds and mixtures into a 2-dimensional space and compared. The results are shown in Figures 5 and 6. In the figures, each point represents a sample, with its location determined by its new representation given by both PCA and t-SNE. Different colours were used to represent each pesticide category.

In the case of PCA, the first 10 PCs explained 99.9 % of the variance. However, after plotting each PC, the first 3 PCs were irrelevant. One of the irrelevant factors to remove in Raman is fluorescence, which is explained in the first two PCs (96.5 % of the total variance). The light absorbed and emitted through fluorescence does not correspond to vibrations, so it is not beneficial in Raman spectroscopy. However, the detector in a Raman spectrometer is unable to differentiate between light emitted from fluorescence phenomena and light from scattered Raman phenomena (Panczer et al., 2012). On the contrary, t-SNE is suitable to highlight the most relevant factors and enhance the different pesticides and mixtures on the basis of their Raman scattering spectra (Stevens et al., 2024).

In the PCA plots, the best separation between pure active compounds was found when plotting PC 4 and PC 5, and between PC 6 and PC 7 in the case of the mixtures. However, the separation is more apparent in the t-SNE plot, especially in the case of pure active compound samples. In the case of the mixtures, PCA does not seem to separate the samples, whereas t-SNE shows slightly better separation for each mixture. Overall, it seems that the mixtures are concentrated in a specific plot region in the t-SNE representation, except for the Captan and Trifloxystrobin mixture, which were the most distinguishable. Similar results were discussed by Luo et al. (2021) regarding the visualisation of vibrational spectroscopy for agri-food samples using t-SNE and PCA, where t-SNE demonstrated the best visual differentiation among samples.

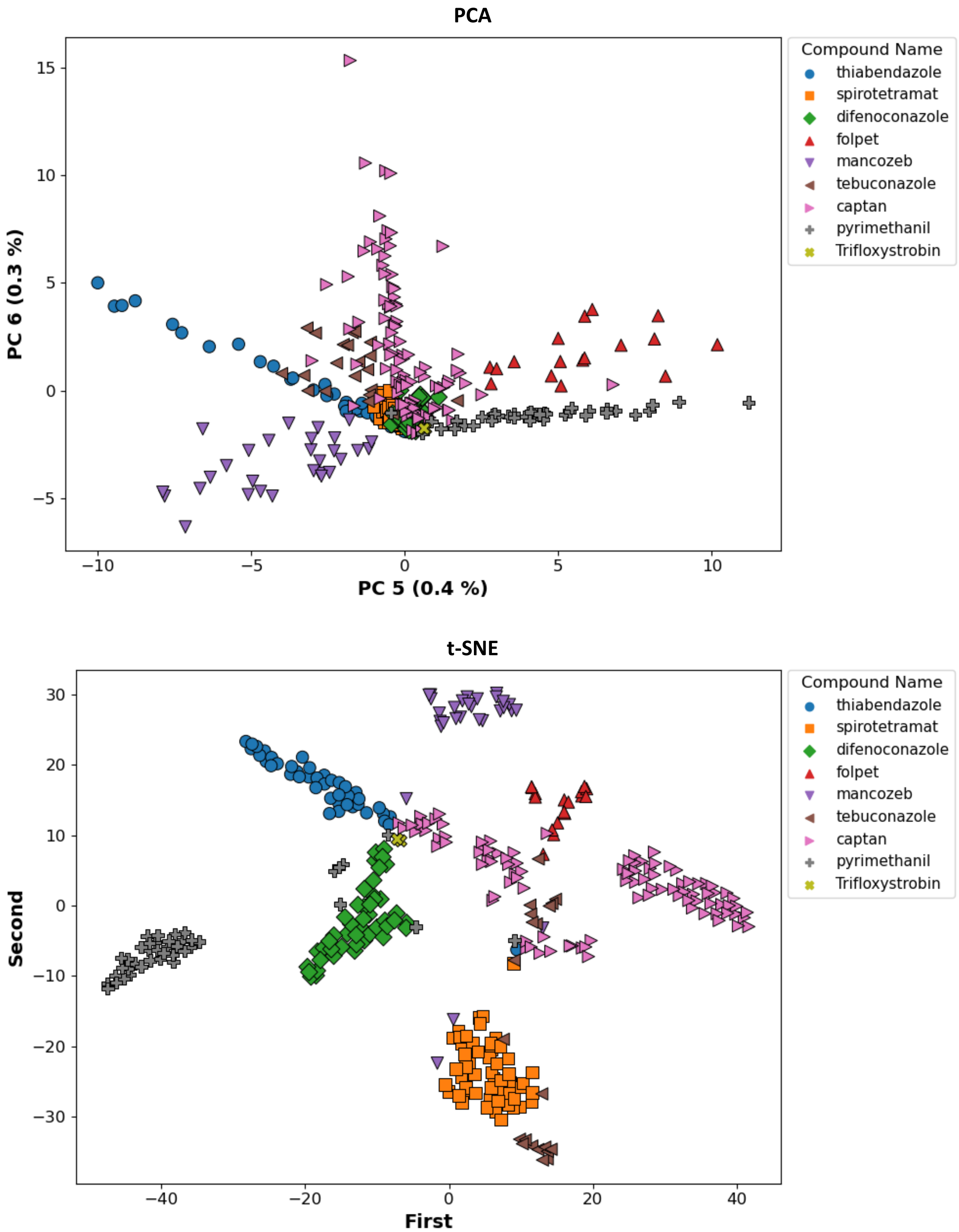


Figure 5. 2-D visualisation of PCA and t-SNE representation of the pure and mixture Raman scattering spectra.

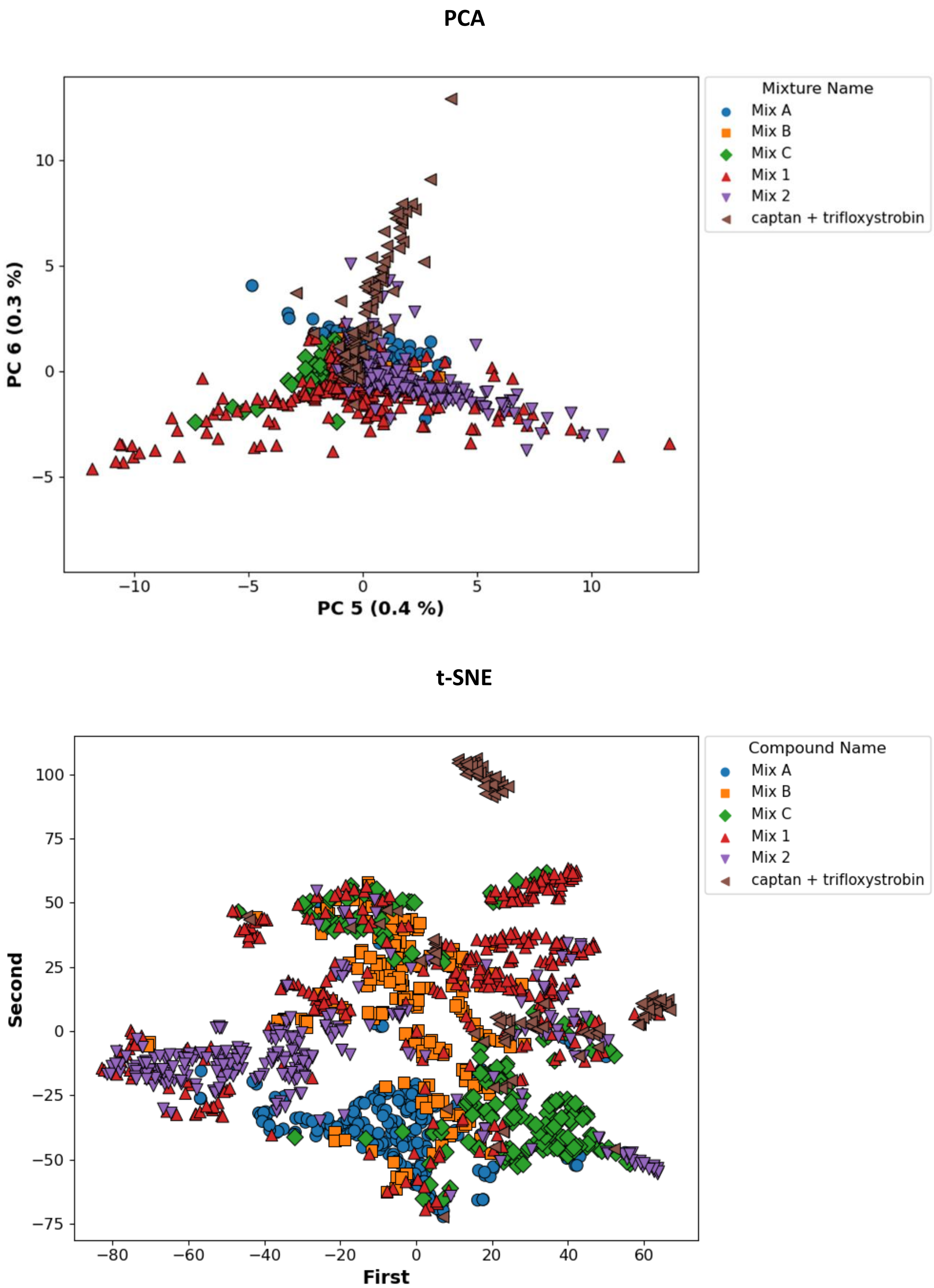


Figure 6. 2-D visualisation of PCA and t-SNE representation of the pure and mixture Raman scattering spectra.

3.3. Multilabel classification of pesticides

After establishing that distinctions exist between pure active compounds and mixtures, as demonstrated by the t-SNE analysis, a PLS-DA, an SVM and a 1-D CNN machine learning models were trained to predict the presence of various active compounds in each sample. This analysis included 'challenging samples' such as Mixtures A, B, and C, which are combinations of different compounds (Table 3), making it more difficult for the models to identify each pesticide correctly. As a fruit can be treated with various pesticides, sometimes in combination, it is crucial to determine which pesticides are present in a sample to ensure compliance with EU regulations on safety limits. A multilabel classification approach was employed to address this, with the results presented in Table 4.

The Hamming loss measures the fraction of incorrectly predicted labels and is commonly used in multilabel classification. This approach was applied by Pan et al. (2021) in their study on identifying complex mixtures in Raman spectroscopy. According to the results presented in this study (Table 4), the 1-D CNN model outperformed both PLS and SVM in minimising errors across the labels, achieving a Hamming loss of 0.33. This was slightly better than the other models, which both had a Hamming loss of 0.41. Despite the overall Hamming loss appearing high across all cases and the challenges associated with predicting the presence of specific compounds in mixtures with very similar spectra, the models exhibited high F1 scores when evaluating the classification of certain pesticides. This is illustrated in Table 4, which highlights the specific compound predictions.

The results in Table 4 reveal distinct patterns in model performance across different active compound identification. For Captan, all three models achieved the highest F1 scores, with SVM and the CNN reaching 100 %, indicating an exceptionally reliable detection of this active compound. Thiabendazole and Tebuconazole showed high F1 scores as well, with the CNN obtaining the best predictive results, with F1 scores of 95 % and 77 % for Thiabendazole and Tebuconazole, respectively. In contrast, Folpet and Mancozeb showed low F1 scores across all models, although the 1-D CNN showed a moderate improvement over PLS and SVM in all those cases. Folpet presents a challenging classification with the lowest performance across all models, with the highest score of 43 % with the 1-D CNN, followed by Mancozed with a 50 %.

Difenoconazole, Pyrimethanil, Trifloxystrobin and Spirotetramat showed moderate performance, with all models yielding similar F1 scores. The 1-D CNN slightly outperformed the other models in Difenoconazole and Pyrimethanil identification, with F1 scores of 64 % and 62 %, respectively. In contrast, in the case of Trifloxystrobin and Spirotetramat, the SVM outperformed the other models, with F1 scores of 63 % and 60 %, respectively.

High F1 scores for certain active compounds, such as Captan, Thiabendazole, and Trifloxystrobin, can be attributed to their distinctive spectral fingerprints, which enable more accurate identification. Active compounds with highly differentiated and characteristic spectral features achieved higher F1 scores in the analysis. Captan, in particular, was predicted with greater accuracy, likely because it appeared in more samples since it is a widely used fungicide in apple production. To further advance the applicability of these findings under real-world conditions, it is essential to increase the number of samples and incorporate real samples into future studies. This will enhance the robustness and reliability of the detection models for practical implementation.

These results indicate that, while the models show variable performance depending on the specific pesticide, the 1-D CNN demonstrates more consistent results overall, especially in more complex classifications. This suggests the potential of deep learning techniques to improve the pesticide active compound identification when using Raman spectra.

These findings are in agreement with other studies. X. Wang et al. (2024) detected up to 10 different pesticides, including glyphosate, propoxur, and chlorpyrifos. They sprayed the samples with the different pesticides, extracted the pesticides from the surface and added silver nanoparticles to the extracted solution with an accuracy of 99 %. Ma et al. (2023) further demonstrated the utility of handheld Raman devices combined with Ag and Au nanoparticles to detect acetamiprid and thiram in apple juice at 1.22 µM and 0.076 µM, respectively, both individually and in combination. Alternatively, Zhai et al. (2017), who employed a self-modelling mixture analysis instead of multilabel classification as in the present study, extracting spectra of individual pesticides from mixtures and achieving high coefficients of determination (0.89 for acetamiprid, 0.93 for chlorpyrifos, and 0.94 for carbendazim). Although promising, this method was limited to simpler mixtures and fewer pesticides.

Table 4. F1 scores of the multilabel classification on the test set (mixtures A, B and C) using the PLS-DA, SVM and 1-D CNN models.

Active compound	**F1 score**		
	PLS	**SVM**	**1-D CNN**
Difenoconazole	51 %	50 %	64 %
Captan	99 %	100 %	100 %
Folpet	23 %	22 %	43 %
Mancozeb	22 %	24 %	50 %
Pyrimethanil	47 %	34 %	62 %
Spirotetramat	41 %	63 %	46 %
Tebuconazole	63 %	72 %	77 %
Thiabendazole	70 %	52 %	95 %
Trifloxystrobin	51 %	60 %	54 %
Hamming Loss	**0.41**	**0.41**	**0.33**

A GradCAM analysis was also performed to assess the predictive identification and performance of the 1-D CNNs and explain how the neural network classifies complex mixtures. It indicates the specific parts of the spectra that the last convolutional layer emphasises during classification. The model accurately identifies the active compounds by focusing on the characteristic scattering bands rather than the bulk of the spectra. This can be seen in Figure 7 with pure Captan and Thiabendazole GradCAM predictions, which show characteristic spectral fingerprints and differentiated scattering bands. The darker the red, the more the network focuses on that specific spectral area. The model primarily emphasises the distinguishing bands of the two example active compounds. For Captan, it focuses on the spectral region from 800 to 50 cm^{-1}, which corresponds to the specific areas where Captan significantly differs from other active compounds. In the case of Thiabendazole, which exhibits a strong characteristic scattering band around 1,600 cm^{-1}, the model notably concentrates on that particular band.

The aim of the GradCAM application was to analyse the model's performance with complex test set mixtures: mixtures A, B, and C (Figure 8). In this context, the model encountered difficulties in detecting and focusing on specific wavelengths corresponding to the scattering bands due to the intricacy of the pesticide mixtures. However, it successfully concentrated on the spectral regions between 1,600 and 50 cm-1, where the characteristic scattering spectral bands are located. This focus can be attributed to the fact that the characteristic signals of the mixtures were weaker than those of the pure signals, as previously demonstrated in Figures 3 and 4. Consequently, the model overlooked the spectral range beyond 2500 cm^{-1}, focusing on the areas where the scattering bands from various active compounds were present. Notably, in some cases, such as Mixture A at the 1,600 cm^{-1} scattering band, the model also identified this particular band, corresponding to the characteristic scattering band of Thiabendazole. Thus, even when faced with predicted complex mixtures, the model consistently maintained its focus on the region of interest.

Such insights enhance confidence in the model, demonstrating that it knows where to concentrate for predictions and that the outcomes are not influenced by external factors or overfitting. This assurance indicates that the model will respond consistently with new samples. GradCAM images display the spectra of the active compounds or mixtures, highlighting the regions of focus that correspond to the characteristic scattering bands. Ultimately, this serves as a visual assessment of the model's behavior rather than a chemical interpretation.

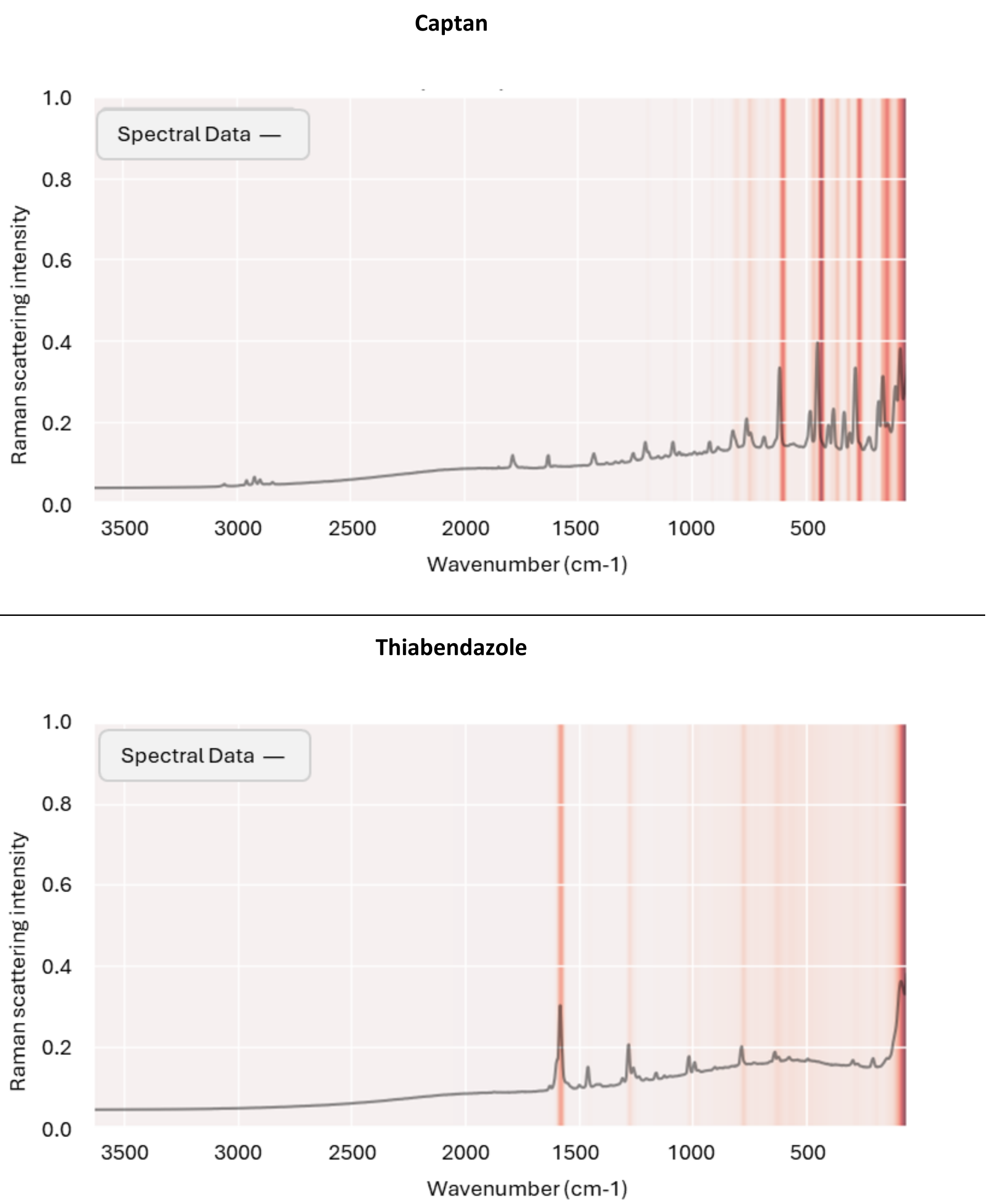
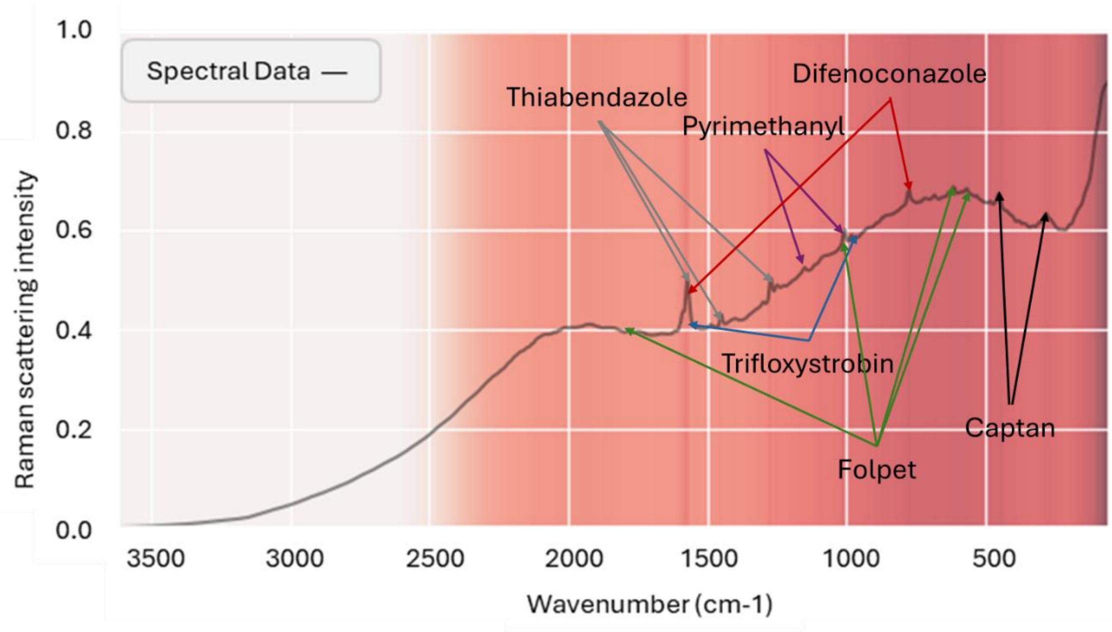


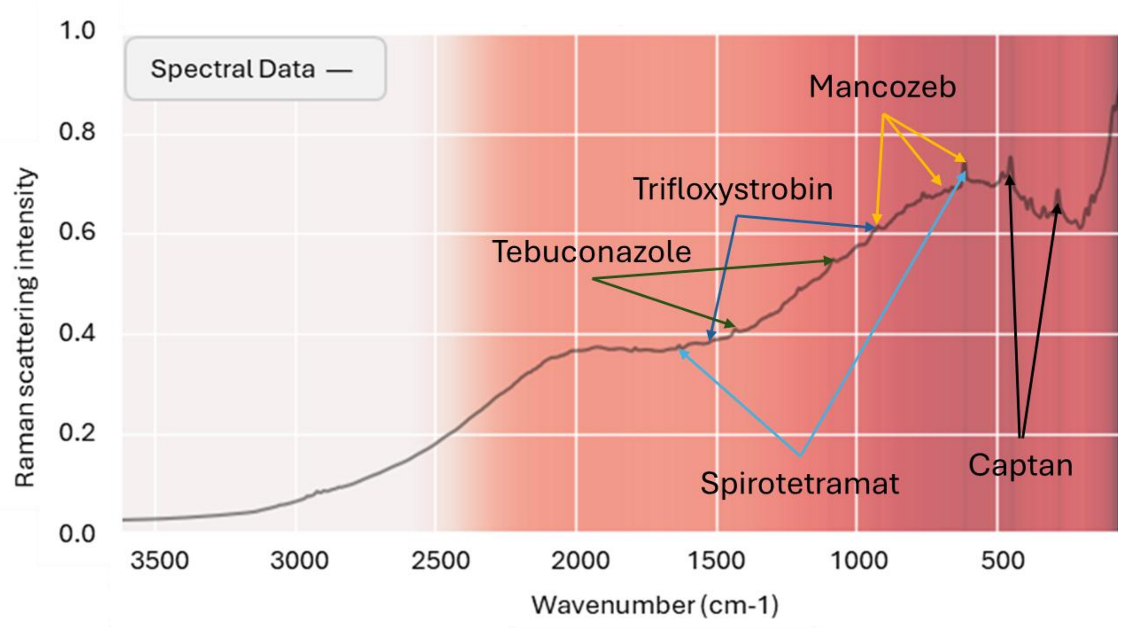
Figure 7. Example of GradCAM in two pure pesticides (Captan and Thiabendazole).

Mix A



Spectral Data
Thiabendazole
Difenoconazole
Pyrimethanyl
Trifloxystrobin
Folpet
Captan
1.0
0.8
0.6
0.4
0.2
0.0
Raman scattering intensity
3500
3000
2500
2000
1500
1000
500
Wavenumber (cm-1)

Mix B



Spectral Data
Mancozeb
Trifloxystrobin
Tebuconazole
Spirotetramat
Captan
1.0
0.8
0.6
0.4
0.2
0.0
Raman scattering intensity
3500
3000
2500
2000
1500
1000
500
Wavenumber (cm-1)

Mix C

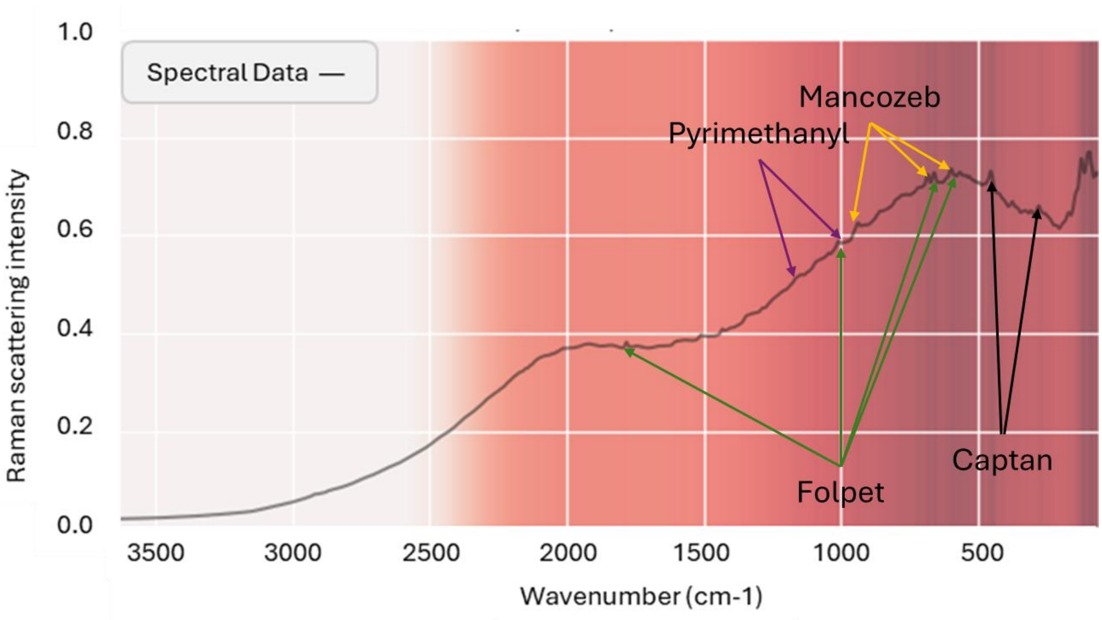


Figure 8. GradCAM on the test set samples (Mixtures A, B and C).

4. Conclusions

This study employed Raman spectroscopy to analyse pure and mixed pesticide active compounds commonly used on apples. Pure compounds exhibited distinct spectral fingerprints, enabling clear identification, while these features became less pronounced in mixtures. Dimensionality reduction techniques, PCA and t-SNE, were applied to assess sample differentiation, with t-SNE providing superior visualization for complex spectral data. Multilabel classification using three machine learning algorithms was then used to detect specific pesticides within mixtures. Compounds such as Captan, Thiabendazole, and Tebuconazole were identified with high accuracy, achieving F1 scores up to 100%, 95%, and 77%, respectively. Difenoconazole, Pyrimethanil, Trifloxystrobin, and Spirotetramat showed moderate detection performance, with F1 scores above 60% across models. Folpet and Mancozeb were more difficult to detect, resulting in lower F1 scores. The 1-D CNN achieved the best results among the models, with the lowest Hamming Loss. GradCAM analysis further validated the CNN by highlighting spectral regions corresponding to the unique scattering bands of each compound. This approach demonstrates strong potential for accurately detecting active compounds in complex samples, which was not previously done before. While chromatography can detect lower concentrations, Raman spectroscopy effectively identifies pesticides at EU maximum residue levels, supporting regulatory compliance and food safety monitoring. Model robustness was confirmed using an independent test set of complex mixtures.

Overall, this work establishes a foundation for multilabel identification of pesticide active compounds in complex mixtures using Raman spectroscopy and machine learning. Future efforts should focus on expanding the dataset, testing on real-world samples, and further optimising the models to ensure reliability and practical application in the field.

Acknowledgements

This work is funded by GVA-IVIA and FEDER by project 52204, project PID2023-150192OR-C31 and C33, project AEI TED2021-130117B-C31 and C33, by the project GVA-PROMETEO CIPROM/2021/014 and by the project PID2021-127975OR-C21. Salvador Castillo thanks INIA for the FPI-INIA grant number PRE2020-094491, partially supported by European Union FSE funds. We thank the Colruyt group for financial support during the project entitled 'Detection of synthetic chemical pesticides by Raman spectroscopy'. We also thank the Protection, Control Products and Residues Unit of the Walloon Agricultural Research Centre (CRA-W) for supplying the samples used in this work.

References

Adou, K., Bontoyan, W. R., & Sweeney, P. J. (2001). Multiresidue Method for the Analysis of Pesticide Residues in Fruits and Vegetables by Accelerated Solvent Extraction and Capillary Gas Chromatography. *Journal of Agricultural and Food Chemistry*, *49*(9), 4153–4160. https://doi.org/10.1021/jf001528q

Botitsi, H. V., Garbis, S. D., Economou, A., & Tsipi, D. F. (2011). Current mass spectrometry strategies for the analysis of pesticides and their metabolites in food and water matrices. *Mass Spectrometry Reviews*, *30*(5), 907–939. https://doi.org/10.1002/MAS.20307

CIPAC. (2020). MULTI-ACTIVE METHOD FOR THE ANALYSIS OF ACTIVE SUBSTANCES IN FORMULATED PRODUCTS TO SUPPORT QUALITY CONTROL SCOPE. In *ESPAC*. https://www.cipac.org/images/pdf/MAM_Method_Rev_21_June_2020.pdf

EFSA. (2021). The 2019 European Union report on pesticide residues in food. *EFSA Journal*, *19*(4), e06491. https://doi.org/10.2903/J.EFSA.2021.6491

Elgendy, M. (2020). Deep Learning for Vision Systems. *Manning Publications Co.*, 480. https://www.manning.com/books/deep-learning-for-vision-systems

Fearn, T., Riccioli, C., Garrido-Varo, A., & Guerrero-Ginel, J. E. (2009). On the geometry of SNV and MSC. *Chemometrics and Intelligent Laboratory Systems*, *96*(1), 22–26. https://doi.org/https://doi.org/10.1016/j.chemolab.2008.11.006

Greenacre, M., Groenen, P. J. F., Hastie, T., D'Enza, A. I., Markos, A., & Tuzhilina, E. (2022). Principal component analysis. *Nature Reviews Methods Primers 2022 2:1*, *2*(1), 1–21. https://doi.org/10.1038/s43586-022-00184-w

Herrera, F., Charte, F., Rivera, A. J., & del Jesus, M. J. (2016). Multilabel Classification. *Multilabel Classification*, 17–31. https://doi.org/10.1007/978-3-319-41111-8_2

Homepage | Phytoweb. (n.d.). Retrieved December 13, 2024, from https://fytoweb.be/en

Hou, R., Pang, S., & He, L. (2015). In situ SERS detection of multi-class insecticides on plant surfaces. *Analytical Methods*, *7*(15), 6325–6330. https://doi.org/10.1039/C5AY01058F

Kaur, R., Choudhary, D., Bali, S., Bandral, S. S., Singh, V., Ahmad, M. A., Rani, N., Singh, T. G., & Chandrasekaran, B. (2024). Pesticides: An alarming detrimental to health and environment. *Science of The Total Environment*, *915*, 170113. https://doi.org/10.1016/J.SCITOTENV.2024.170113

Kazlagić, A., & Omanović-Mikličanin, E. (2020). Application of Raman Spectroscopy in Food Forensics: A Review. *IFMBE Proceedings*, *73*, 257–263. https://doi.org/10.1007/978-3-030-17971-7_40

Kumar Hazra, D., Purkait, A., & Dipak Kumar Hazra, C. (2019). Role of pesticide formulations for sustainable crop protection and environment management: A review. *Journal of Pharmacognosy and Phytochemistry*, *8*(2), 686–693.

https://www.phytojournal.com/archives/2019.v8.i2.7684/role-of-pesticide-formulations-for-sustainable-crop-protection-and-environment-management-a-review

Li, Y., Peng, Y., Qin, J., & Chao, K. (2019). A Correction Method of Mixed Pesticide Content Prediction in Apple by Using Raman Spectra. *Applied Sciences*, *9*(8). https://doi.org/10.3390/app9081699

Luo, N., Yang, X., Sun, C., Xing, B., Han, J., & Zhao, C. (2021). Visualization of vibrational spectroscopy for agro-food samples using t-Distributed Stochastic Neighbor Embedding. *Food Control*, *126*, 107812. https://doi.org/https://doi.org/10.1016/j.foodcont.2020.107812

Ma, L., Han, E., Yin, L., Xu, Q., Zou, C., Bai, J., Wu, W., & Cai, J. (2023). Simultaneous detection of mixed pesticide residues based on portable Raman spectrometer and Au@Ag nanoparticles SERS substrate. *Food Control*, *153*, 109951. https://doi.org/https://doi.org/10.1016/j.foodcont.2023.109951

Mandrile, L., Giovannozzi, A. M., Durbiano, F., Martra, G., & Rossi, A. M. (2018). Rapid and sensitive detection of pyrimethanil residues on pome fruits by Surface Enhanced Raman Scattering. *Food Chemistry*, *244*, 16–24. https://doi.org/https://doi.org/10.1016/j.foodchem.2017.10.003

Marczewska, P., Miszczyk, M., Płonka, M., Kronenbach-Dylong, D., Szeremeta, D., & Sajewicz, M. (2019). Application of different chromatographic techniques and chemometric analysis in authenticity testing of plant protection products containing azoxystrobin as an active substance. *Journal of Environmental Science and Health, Part B*, *54*(7), 590–597. https://doi.org/10.1080/03601234.2019.1610298

McGarvey, B. D. (1993). High-performance liquid chromatographic methods for the determination of N-methylcarbamate pesticides in water. soil, plants and air. *Journal of Chromatography A*, *642*(1), 89–105. https://doi.org/https://doi.org/10.1016/0021-9673(93)80079-N

Méndez-Albores, A., & Hernández, I. (2018). *DETECCIÓN DE RESIDUOS DE GLIFOSATO EN MAÍZ FRENTE A LOS RIESGOS A LA SALUD HUMANA Y A ABIÓTICOS, MEDIANTE ESPECTROSCOPIA RAMAN AUMENTADA POR LA SUPERFICIE.*

Miszczyk, M., Płonka, M., Bober, K., Dołowy, M., Pyka, A., & Pszczolińska, K. (2015). Application of chemometric analysis based on physicochemical and chromatographic data for the differentiation origin of plant protection products containing chlorpyrifos. *Journal of Environmental Science and Health, Part B*, *50*(10), 744–751. https://doi.org/10.1080/03601234.2015.1048108

Pan, L., Pipitsunthonsan, P., Daengngam, C., Channumsin, S., Sreesawet, S., & Chongcheawchamnan, M. (2021). Identification of Complex Mixtures for Raman Spectroscopy Using a Novel Scheme Based on a New Multi-Label Deep Neural Network. *IEEE Sensors Journal*, *21*(9), 10834–10843. https://doi.org/10.1109/JSEN.2021.3059849

Panczer, G., De Ligny, D., Mendoza, C., Gaft, M., Seydoux-Guillaume, A. M., & Wang, X. (2012). Raman and fluorescence. *European Mineralogical Union Notes in Mineralogy*, *12*(1), 61–82. https://doi.org/10.1180/EMU-NOTES.12.2

Reyes, C., Medina-Gutiérrez, C., Medina-Valtierra, J., & Berrú, R. (2004). Aplicación de la espectroscopia Raman para la caracterización de pesticidas orgánicos. *Revista Internacional de Contaminación Ambiental*.

Scorza, F. A., Beltramim, L., & Bombardi, L. M. (2023). Pesticide exposure and human health: Toxic legacy. *Clinics*, *78*, 100249. https://doi.org/10.1016/J.CLINSP.2023.100249

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2020). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *International Journal of Computer Vision*, *128*(2), 336–359. https://doi.org/10.1007/s11263-019-01228-7

Soltani Nazarloo, A., Rasooli Sharabiani, V., Abbaspour Gilandeh, Y., Taghinezhad, E., Szymanek, M., & Sprawka, M. (2021). Feasibility of Using VIS/NIR Spectroscopy and Multivariate Analysis for Pesticide Residue Detection in Tomatoes. *Processes*, *9*(2). https://doi.org/10.3390/pr9020196

Stevens, F., Carrasco, B., Baeten, V., & Fernández Pierna, J. A. (2024). Use of t-distributed stochastic neighbour embedding in vibrational spectroscopy. *Journal of Chemometrics*, *38*(4), e3544. https://doi.org/10.1002/CEM.3544

Tewes, T. J., Welle, M. C., Hetjens, B. T., Tipatet, K. S., Pavlov, S., Platte, F., & Bockmühl, D. P. (2023). Understanding Raman Spectral Based Classifications with Convolutional Neural Networks Using Practical Examples of Fungal Spores and Carotenoid-Pigmented Microorganisms. *AI (Switzerland)*, *4*(1), 114–127. https://doi.org/10.3390/AI4010006/S1

Tsagkaris, A. S., Pulkrabova, J., & Hajslova, J. (2021). Optical Screening Methods for Pesticide Residue Detection in Food Matrices: Advances and Emerging Analytical Trends. *Foods*, *10*(1). https://doi.org/10.3390/foods10010088

Wang, L., Ma, P., Chen, H., Chang, M., Lu, P., Chen, N., Yuan, Y., Chen, N., & Zhang, X. (2022). Rapid Determination of Mixed Pesticide Residues on Apple Surfaces by Surface-Enhanced Raman Spectroscopy. *Foods*, *11*(8). https://doi.org/10.3390/foods11081089

Wang, X., Du, Y., Zhang, H., Xu, Y., Pan, Y., Wu, T., & Hu, H. (2014). Fast enrichment and ultrasensitive in-situ detection of pesticide residues on oranges with surface-enhanced Raman spectroscopy based on Au nanoparticles decorated glycidyl methacrylate–ethylene dimethacrylate material. *Food Control*, *46*, 108–114. https://doi.org/https://doi.org/10.1016/j.foodcont.2014.04.035

Wang, X., Jiang, S., Liu, Z., Sun, X., Zhang, Z., Quan, X., Zhang, T., Kong, W., Yang, X., & Li, Y. (2024). Integrated surface-enhanced Raman spectroscopy and convolutional neural network for quantitative and qualitative analysis of pesticide residues on pericarp. *Food Chemistry*, *440*, 138214. https://doi.org/https://doi.org/10.1016/j.foodchem.2023.138214

Yang, X., Imasaka, T., & Imasaka, T. (2018). Determination of Pesticides by Gas Chromatography Combined with Mass Spectrometry Using Femtosecond Lasers Emitting at 267, 400, and 800 nm as the Ionization Source. *Analytical Chemistry*, *90*(7), 4886–4893. https://doi.org/10.1021/acs.analchem.8b00537

Yusoff, S. N. M., Kamari, A., & Aljafree, N. F. A. (2016). A review of materials used as carrier agents in pesticide formulations. *International Journal of Environmental Science and Technology*, *13*(12), 2977–2994. https://doi.org/10.1007/S13762-016-1096-Y/METRICS

Zhai, C., Peng, Y., Li, Y., & Chao, K. (2017). Extraction and identification of mixed pesticides' Raman signal and establishment of their prediction models. *Journal of Raman Spectroscopy*, *48*(3), 494–500. https://doi.org/10.1002/JRS.5049

GENERAL DISCUSSION

This doctoral thesis addressed several challenges commonly found in fruit production and commercialisation, emphasising the importance of advanced non-destructive technologies that can substitute destructive analysis combined with machine learning techniques, focusing on neural networks. The research presented here explored innovative approaches to solve important issues related to fruit quality, safety, and consumer acceptance, demonstrating the effectiveness of hyperspectral imaging and Raman spectroscopy combined with different neural network approaches. Experiments conducted on strawberries, persimmons, plums, and pesticides commonly used in apples provided evidence of the robustness, accuracy, and applicability of these technologies and machine learning methods.

Innovation in strawberry quality and acceptance thanks to ANNs and HSI.

Hyperspectral imaging and ANNs represent a powerful combination for assessing fruit quality properties such as TSS, TA, and texture. This technological synergy not only excels in predicting internal quality and consumer acceptance but also helps avoid food fraud by enabling accurate identification of closely related cultivars, as shown in Chapter I. A key novelty of the work carried on with strawberries was its scale. No prior research has investigated consumer preferences using such a large number of samples and as many as 17 different strawberry cultivars. This extensive sample size and diversity of varieties ensure broad variability, increasing both the robustness of the experimental results and the reliability of the predictive models. While numerous studies have explored the prediction of internal quality or consumer acceptance in strawberries, none have brought together such a wide range of samples and varieties to predict consumer preference with comparable confidence. The results reinforce this approach: in strawberries (Chapter I), the ANN models predicted internal quality parameters with high accuracy ($R^2 = 0.85$ for TSS, 0.81 for TA, and 0.76 for texture), while the classification of 17 cultivars achieved an F1 score of 0.84, illustrating the potential of HSI and ANNs to help prevent food fraud. Most notably, consumer acceptance prediction reached a mean cultivar R^2 of 0.78, demonstrating the models' capacity to capture the balance between sweetness, acidity, and other traits that affect sensory preferences.

The scale and cross-cultivar approach of this study established a strong foundation for future industrial integration of these advanced technologies. It paves the way for improved market targeting, enhanced product quality, and reductions in food waste, as fruits most likely to satisfy consumer preferences can be prioritised. Nevertheless, further research should expand the study of consumer acceptance to include people from additional countries, samples from multiple seasons, and an even broader selection of strawberry varieties. Such future work would help validate and generalise these findings and further revolutionise strawberry sorting and marketing strategies.

Innovative dimensionality reduction using VAEs for flesh firmness prediction using HSI.

In persimmon fruit (Chapter II), this study presents a novel approach that has never been done before: the application of VAEs for both hyperspectral image and texture curve analysis. What differentiates this work is the use of VAEs, which enable a reduction in data dimensionality of over 99 %, obtaining the most important and relevant information. Besides, in the case of hyperspectral imaging, dimensionality reduction allowed for the selection of just six wavelengths (440, 460, 520, 580, 660, and 1000 nm), significantly improving computational efficiency and making industrial integration, such as cost-effective, real-time sorting lines with multispectral cameras, much more feasible for in-line estimation.

Moreover, VAEs demonstrated their capability of effective feature extraction not only from images but also from texture curves. Unlike traditional studies that rely solely on specific parameters, such as the area under the curve, maximum breaking force, or the slope, the autoencoders captured information from the entire curve, yielding a richer and more representative feature set. By using these VAE-extracted features and classifying with SVM, firmness classification reached an accuracy of 81.3%, successfully distinguishing fruit suitable for storage (> 40 N), commercialisation (40 > 20 N), or those deemed too soft (< 20 N).

Looking ahead, future studies should apply this promising approach across multiple seasons and with a wider range of persimmon varieties to further validate its effectiveness and robustness. This will help ensure the generalizability of the method and accelerate its adoption in real-world industrial conditions.

Detecting Invisible damages in fruits with Interpretable CNNs.

Damage and bruise detection is a critical step in reducing food losses across the agro-industry. Chapter III addresses this issue in plums, presenting an approach that, while similar methods have been described in the literature, has never before been implemented in this particular variety of plum (Presenta). This variety's dark skin makes visual identification of internal damage especially challenging, as conventional inspection methods often fail to detect bruises beneath the surface. However, HSI excels at revealing spectral differences associated with bruising, such as water release, chlorophyll degradation, or cell wall rupture-markers that go unnoticed in visual assessments.

Once these spectral variations were captured, CNNs could process the data and classify bruised fruits with high accuracy, achieving overall F1 scores of 90% and up to 100% for severe bruising. Further dimensionality reduction identified ten critical wavelengths, with three (985 nm, 822 nm, and 705 nm) being the most important. Remarkably, even when the CNNs were trained using just these three wavelengths, the models reached F1 scores of up to 88%. This demonstrates that effective bruise detection is possible with multispectral imaging, significantly improving the practicality and cost-efficiency of commercial implementation.

A key novelty of this study was the use of GradCAM, which provides visual insight into the model's decision-making process by highlighting the regions of the image the model focuses on when predicting bruising. This interpretability allows us to verify the effectiveness of the model and provides a rational basis for selecting between models with similar quantitative performance. For instance, GradCAM analysis showed that among the three CNN architectures tested, ResNet most consistently and accurately focused on the bruised areas, confirming its suitability for this task.

Future studies should aim to validate and extend these findings across multiple seasons and a wider range of dark-skinned plum cultivars to ensure robustness and broader applicability of these methods.

Safety assessment by complex Pesticide Mixture Identification with Raman and Interpretable CNNs.

Food safety is critical for public health, consumer confidence, and regulatory compliance. Pesticide residues, particularly in complex mixtures, remain a pressing concern due to their potential health risks. While previous studies have addressed pesticide identification using Raman spectroscopy, none have tackled the challenge of identifying active compounds within the complex pesticide mixtures commonly applied in apple orchards. This study distinguished itself by implementing a multilabel classification approach to determine multiple active compounds in such mixtures, a scenario much closer to real-world orchard practices.

The integration of 1-D CNNs enabled highly accurate identification of certain active compounds, such as Captan, Thiabendazole, and Tebuconazole, which were identified with high accuracy, achieving F1 scores up to 100 %, 95 %, and 77 %, respectively, even in the presence of overlapping spectral signals. However, low accuracies were observed for compounds like Mancozeb or Folpet, highlighting areas for future refinement. A further innovation of this work was the use of GradCAM, which allows visualisation of the spectral regions the CNNs focus on during classification. This demonstrated that the models rely on the characteristic wavelengths of the target compounds, thereby providing greater confidence in the reliability of predictions. This approach set this study apart from prior work, both in its capacity to tackle complex, real-world mixtures and in the transparency it offers to model decision-making.

Looking forward, future studies should extend this framework to in-situ detection and validation with real-world samples, applying pesticides directly to apples and acquiring Raman spectra under field conditions. Additionally, exploring alternative acquisition strategies and continually updating models with new data will be essential to ensure robustness and practical applicability in commercial settings. This will help bridge the gap between laboratory advancements and operational, real-time food safety monitoring.

Both hyperspectral imaging and Raman spectroscopy can reveal critical insights from complex data, providing scalable and non-destructive solutions for the food industry. As these technologies continue to advance, they have the potential to improve consumer safety and more effectively address industry challenges. All studies faced shared challenges in resolving signal complexity. Overlapping spectral features, whether from pesticide mixtures or early-stage bruises, limited detection accuracy for certain compounds or conditions. These challenges highlight the need for continuous optimisation, including improved preprocessing techniques, enhanced spectral resolution, and hybrid approaches.

CONCLUSIONS

The research presented in this work doctoral thesis demonstrates the significant potential of combining advanced non-destructive analytical equipment, specifically hyperspectral imaging and Raman spectroscopy, with state-of-the-art ANN methodologies, including MLPs, different types of CNNs and AEs, for addressing critical challenges in the agri-food sector. While the studies focused on fruits such as strawberries, persimmons, plums, and apples, the broader implications of these findings extend across the entire industry, demonstrating how these approaches can revolutionise quality assessment, safety monitoring, and process optimisation for a wide range of agri-food products.

The application of HSI, particularly in the visible to near-infrared (VIS-NIR) range, has proven to be a powerful tool for non-destructive evaluation of internal and external quality attributes. In the case of strawberries, for instance, integrating HSI with ANNs enabled accurate prediction of key quality parameters such as TSS, TA, and texture. Notably, the research showed that the ratio of these parameters (TSS/TA) was a strong indicator of consumer acceptance, a finding that underscores the complexity of sensory perception. The ANN models consistently outperformed traditional machine learning methods such as SVM or decision tree-based models, highlighting their superior ability to capture hidden patterns and non-linear relationships within high-dimensional spectral data. Moreover, large-scale cultivar classification was achieved, opening new possibilities for automated sorting and traceability in the supply chain and avoiding food fraud.

The use of autoencoders together with HSI, specifically VAEs, further demonstrated the potential of advanced neural networks for dimensionality reduction and feature extraction. In the assessment of persimmon flesh texture, VAEs enabled the condensation of hyperspectral data into a minimal set of informative features, achieving a 99 % data reduction, facilitating rapid and accurate classification of fruit according to industrial standards for storage and commercialisation. Besides, six key wavelengths were identified. This approach not only simplified the analysis but also paved the way for practical, real-time implementation of multispectral imaging-based sorting systems in packing houses, where speed and scalability are essential.

CNNs, with their exceptional capacity for image analysis and pattern recognition, were key in addressing the challenge of detecting invisible damage in plums. By taking advantage of the spatial and spectral richness of HSI data, CNNs were able to identify early-stage bruises in dark-skinned cultivars, an achievement that is particularly valuable given the limitations of visual inspection. The use of interpretability tools such as GradCAM provided insights into the decision-making process of the models, ensuring that predictions were based on relevant and meaningful features. Furthermore, the identification of three key wavelengths for bruise detection demonstrated the feasibility of developing efficient multispectral systems for industrial use, with the potential to reduce food loss due to undetected internal damage significantly.

Raman spectroscopy, when combined with one-dimensional CNNs, offered a robust solution for the detection of pesticide residues in apples. The research demonstrated that even in complex mixtures where spectral bands overlap, advanced neural networks could successfully identify specific active compounds, supporting regulatory compliance and enhancing food safety monitoring. The application of GradCAM in this context validated the model's ability to focus on characteristic spectral regions, further increasing confidence in the results and providing a foundation for the development of rapid, on-site screening tools.

Thus, these findings illustrate the profound impact that the integration of HSI, Raman spectroscopy, and neural network-based analytics can have on the agri-food industry. These technologies enable rapid, objective, and non-destructive assessment of quality and safety attributes, overcoming many of the limitations associated with traditional methods. Their flexibility allows for adaptation to a wide variety of products and challenges, from predicting consumer preferences and classifying cultivars to detecting invisible defects and ensuring chemical safety. This research addresses practical problems and lays the foundation for future innovations in agri-food analytics.