The American Journal of Applied Sciences
ISSN 2689-0992 Volume 08 - 2026

Proceedings of Women in Academia, Research and Management for Work-life Initiatives for Sustainable Health & Empowering Safety (WARM-WISHES 2026)

AI-Driven Hyperspectral Morphotoxic Profiling of Opportunistic Pathogens in Mixed Biofilm Contaminations

Anant Dev Shukla

Amity Institute of Biotechnology, Amity University Uttar Pradesh, Lucknow Campus, Gomti Nagar Extension, Lucknow, India

Sabhyata Ojha

Amity Institute of Biotechnology, Amity University Uttar Pradesh, Lucknow Campus, Gomti Nagar Extension, Lucknow, India

Nikita Basant

Amity Institute of Biotechnology, Amity University Uttar Pradesh, Lucknow Campus, Gomti Nagar Extension, Lucknow, India

Received: 22 Apr 2026 | Received Revised Version: 29 Apr 2026 | Accepted: 07 May 2026 | Published: 16 May 2026

DOI: 10.37547/tajas/warm-08

Abstract

The health care, environmental, and industrial infrastructures encounter challenging issues when dealing with biofilm forming opportunistic pathogens because of their persistence, antimicrobial resistance, and the capacity to coexist in a complex mixed microbial composition. Fast and non-invasion methods of detecting microbial content in biofilms are hence key towards successful monitoring and control. This paper suggests a hyperspectral morphotoxic profiling workflow that uses artificial intelligence and is applicable to analyzing mixed biofilm contaminations of opportunistic pathogens. The method combines hyperspectral imaging and artificial intelligence to identify and make sense of spectral-morphological features of microbial biofilms. The first stage entails the hyperspectral image data acquisition and preprocessing to eliminate noise and standardize spectral content, and the second stage is the dimensionality reduction to retain the most informative spectral data and reduce the complexity of the data. They then train a convolutional neural network to be able to learn discriminative spatial and spectral patterns related to the structures of microbial biofilms and presence of the pathogen in the processed images. The trained model is tested on the basis of automatic delivery of microbial signatures in intricate biofilm settings. To ensure ease of application and illustrate practicality, a simple web based interface will be built with flask framework where one will be able to upload images and get instant AI based predictions through an interactive solution. The current framework of hyperspectral imaging, artificial intelligence, and web-based implementation with a focus on microbial profiling in only a few seconds is discussed as a potential option, and hyperspectral imaging could be used in the context of environmental toxicology, monitoring infections, water quality assessment, and managing biofilm in industries.

Keywords: Artificial Intelligence, Hyperspectral Imaging, Microbial Biofilms, Opportunistic Pathogens, Morphotoxic Profiling, Deep Learning, Convolutional Neural Networks

Cite This Article: Anant Dev Shukla, Sabhyata Ojha, & Nikita Basant. (2026). AI-Driven Hyperspectral Morphotoxic Profiling of Opportunistic Pathogens in Mixed Biofilm Contaminations. The American Journal of Applied Sciences, 75–83. https://doi.org/10.37547/tajas/warm-08

1. Introduction

Biofilms are highly structured, three-dimensional communities of bacteria that are embedded in a protective, self-produced extracellular matrix called extracellular polymeric substance (EPS), which is a mixture of polysaccharides, proteins, lipids and extracellular DNA (Yin et al., 2022; Zhao et al., 2023). Biofilm formation is the most common survival strategy of many harmful bacteria, and enables their growth and establishment in a wide range of environments such as hospitals, natural environments and industry (Liu et al., 2023; Tavares-Carreon et al., 2023). In clinical and medical settings, biofilms are found on medical equipment and human tissues, and are the main reason for serious hospital infections (Kranjec et al., 2021; Zhao et al., 2023). Similarly, in food and industrial settings, pathogenic biofilms that form on abiotic surfaces are persistent contaminants, leading to considerable financial losses and reputational damage, as well as significant health risks (Liu et al., 2023; Soni et al., 2022). Perhaps the biggest challenge with biofilms is that they are resistant to typical antimicrobial treatment, rendering the majority of conventional therapies ineffective (Elfadadny et al., 2024; Yin et al., 2022). The EPS serves as a physical barrier that prevents the antimicrobial agent from reaching and killing the microorganisms (Yin et al., 2022). Further to this, in biofilms made up of several species of microorganisms, interactions between the species, horizontal gene transfer and persister cells all contribute to the increased tolerance to antibiotics of the community of microorganisms (Anju et al., 2022; Kranjec et al., 2021). These factors make infections more difficult to eradicate, and require new and improved detection methods. Existing approaches to identifying microorganisms, such as methods that rely on growing the microorganisms in culture, have significant limitations, as they are time-consuming and labour intensive (Soni et al., 2022). Moreover, these traditional approaches may not adequately represent the composition of mixed microbial populations, especially when it comes to identifying slow-growing microorganisms or those that are present in low concentrations (Anju et al., 2022). Therefore, hyperspectral imaging (HSI) has emerged as a rapid, non-invasive technique for analysing biological and microbial samples (Nguyen et al., 2021; Soni et al., 2022). Simultaneously capturing spatial and spectral information over hundreds of narrow spectral bands, HSI reveals the unique physiological and biochemical properties of samples, allowing for the reliable detection and classification of bacteria, fungi and viruses (Alkhatib et al., 2023; Soni et al., 2022). In order to overcome problems with the large dataset and redundancy ("curse of dimensionality") associated with HSI - not to mention the need to analyse these images - researchers have relied increasingly on artificial intelligence and deep learning using Convolutional Neural Networks (CNNs) to interpret these images (Alkhatib et al., 2023). Deep learning tools, such as 1D, 2D, and 3D-CNNs have displayed great potential in extracting complex integrated spatial-spectral features from HSI data (Alkhatib et al., 2023; Nguyen et al., 2021). When applied to microbiology, HSI and deep learning have been useful to classify different microbial colonies, to distinguish between living and dead cells, and to decrease interference from complex background noise (Soni et al., 2022). Despite the significant potential of HSI and deep learning, there is still a major gap when it comes to detecting multi-species biofilm contamination. The current lack of non-destructive and real-time techniques to precisely detect the presence of specific pathogenic strains in complex biofilms with multiple species is evident (Anju et al., 2022; Soni et al., 2022). The application of HSI with deep learning approaches in multi-species biofilms is largely hindered by the ability to extract useful spatial information and spectral features. In particular, the chemical complexity of the food or human tissue substrate, the complex biochemical composition of the EPS layer and the overlapping spectral signals from other microbial species cause non-linear noise and distortions, which mask the signals of the target pathogens (Alkhatib et al., 2023; Soni et al., 2022). And detection of harmful bacterial strains that exist in very low quantities in the presence of the bulk microorganisms is another critical issue that needs to be addressed in the hyperspectral biological analysis (Soni et al., 2022). In this regard, it is important to consider how HSI - and particularly advanced CNN architectures - can be optimally used to extract valid spatial-spectral features for non-invasive, real-time profiling and identification of opportunistic pathogens in mixed biofilm contaminations. Taking insights from the literature into consideration, the aim of this research is to develop a real-time, non-invasive hyperspectral image system and deep learning (CNN) algorithms to detect and profile opportunistic pathogens in polymicrobial communities.

2. Materials and Methods

2.1 Data Acquisition and Preprocessing

The dataset used in this study had a collection of 370 images, each with 236 spectral bands in the visible and near infrared (VNIR) region and each piece of data was stored as a hyperspectral data cube (x, y, 236). Preprocessing consisted of the acquisition of hyperspectral TIFF images, reshaping the hyperspectral cube, Principal Component Analysis (PCA), PCA image reconstruction, z-score normalization, image resizing to 224 × 224 pixels, and a stratified train-test splitting for a balanced model training. Dataset is publicly available on Kaggle (https://www.kaggle.com/datasets/user164919/hyperspectralbiofilm-dataset-of-pathogens). **Table 1** summarizes the different pathogen classes, their ecosystem, and clinical relevance that were incorporated in the hyperspectral biofilm dataset. HSI has become an emerging technology for non-destructive detection of microorganisms because it combines the best of spectroscopy and computer vision to provide spatial and spectral information from a single image (Bonah et al., 2019). Before the images could be used to analyse the data, they were subjected to a preprocessing procedure that involved cleaning the data by eliminating undesirable noise and normalising the spectral data - a process that is an important key in providing quality data to machine learning algorithms. Normalisation and noise reduction are important steps in hyperspectral approaches due to the increased quality of data, which leads to improved classification results at the end of the pipeline (Bonah et al., 2019). The hyperspectral data set for biofilm (data used for analysis and classification) is visualized in **Figure 1.**

Table 1: Pathogen Classes Used in the Hyperspectral Biofilm Dataset

Pathogen Class	Type	Common Habitat	Clinical Relevance
E. coli	Bacteria	Gut, water	UTIs, infections
E. faecalis	Bacteria	GI tract	Nosocomial infections
S. aureus	Bacteria	Skin	MRSA, wound infections
K. pneumoniae	Bacteria	Respiratory tract	Pneumonia
P. aeruginosa	Bacteria	Soil, water	Drug-resistant infections
C. albicans	Fungus	Human microbiota	Candidiasis

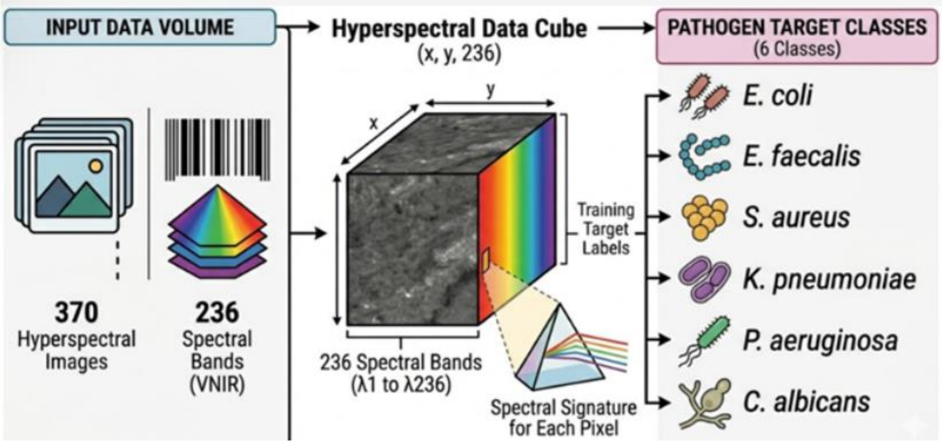

Figure 1: Hyperspectral Biofilm Dataset Visualisation

2.2 Dimensionality Reduction

Because hyperspectral data inherently has a lot of spectral dimensions, Principal Component Analysis (PCA) was used to extract only the meaningful spectral features and ignore the rest. This reduces the data's complexity by decreasing its dimensionality while retaining important spatial and spectral features to distinguish between the different pathogen classes. PCA is a data reduction technique commonly used for processing hyperspectral data to eliminate noise in the data, simplify the computational requirements and preserve the most relevant information required for classification (Bonah et al., 2019).

2.3 Convolutional Neural Network (CNN)

A pathway custom CNN was designed to classify the pathogens. An input layer was used to input 224×224×3 images, while three convolutional blocks containing 32, 64 and 128, 3×3 filters were added to learn and reduce the features. Every convolutional block was provided with a ReLU activation function and a 2×2 max-pooling operation. This was closed by a flatten, dense (256 neurons) and SoftMax layer which classified the input image into six pathogen classes: E. coli, E. faecalis, S.aureus, K. pneumoniae, P. aeruginosa and C. albicans. The application of CNNs to microbial classification has been reported widely as they are capable of automatically learning important features from complex images data and therefore are the network of choice for classifying images (Wu & Gadsden, 2023; Kotwal et al. 2021). The proposed model of the convolutional neural network (CNN) for pathogen classification is described in the **Figure 2.**

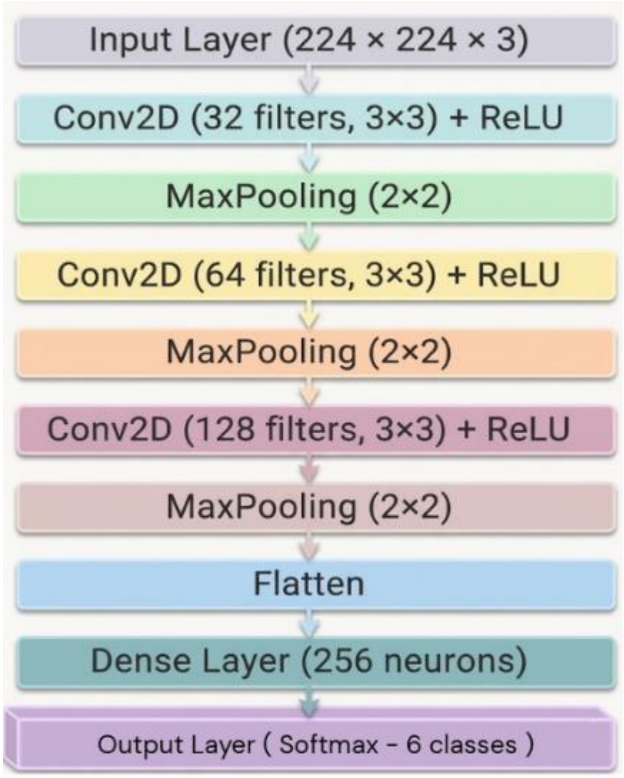

Figure 2: CNN Architecture

2.4 Model Training and Evaluation

The model was trained to learn certain types of visual and spectral patterns from microbial biofilms and the pathogens of interest. The model's performance was monitored as it was trained to make sure it was capable of dealing with new data, a critical aspect of any diagnostic system. In the case of machine learning approaches for microbial classification training and validation are important steps to ensure that the model accurately predicts new, previously unseen data and is effective in different environments (Kotwal et al., 2021).

2.5 Web-Based Deployment

For easy practical use, a user-friendly web interface was designed using the Flask framework. This system enables users to upload their images and get the predictions of the model in an automated manner. Online platforms are increasingly being used in the analysis of hyperspectral imaging systems as they make the use of deep learning systems available to people who are not experts in coding or technology, and can interact with these complex systems via accessible and easy to use interfaces (Dhaene et al., 2023). The overall workflow of the hyperspectral biofilm pathogen detection system is shown in the following figure (**Fig. 3**).

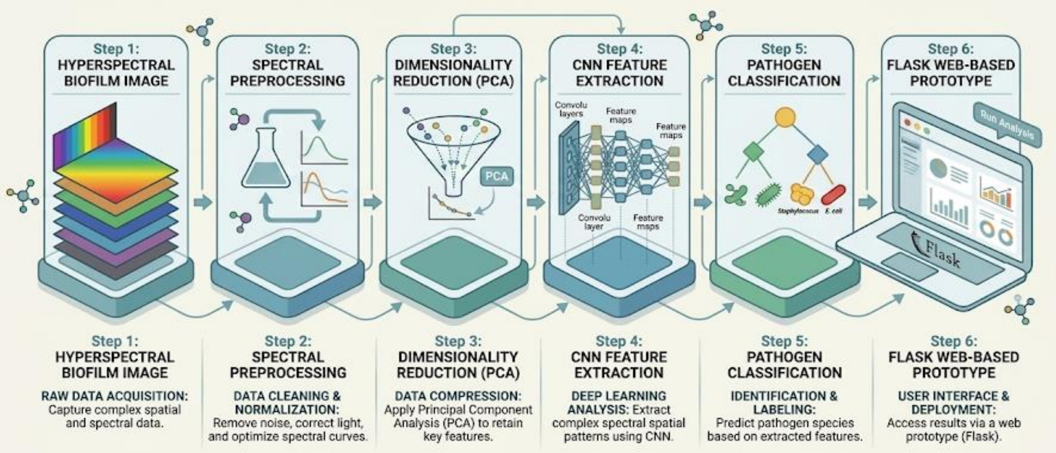

Figure 3: Hyperspectral Biofilm pathogen Detection Workflow

3. Results and Discussion

The results of the experiments demonstrate the effectiveness of the hyperspectral morphotoxic profiling approach in the classification of pathogens from a mixture of biofilm. The CNN model performed well with a training accuracy of 97.6% and a validation accuracy of 85.3%. This demonstrates the model successfully learned the various spatial and spectral patterns from hyperspectral data cubes that decided what type of pathogen was present in each pixel, and was able to effectively detect all six species of interest: *E. coli, E. faecalis, S. aureus, K. pneumoniae, P. aeruginosa,* and *C. albicans*.

Further insights into the performance of the model were obtained from diagnostic visualisations. The spectrum variability of the biofilm difference between the various biofilm pixels is shown in **Figure 4.**

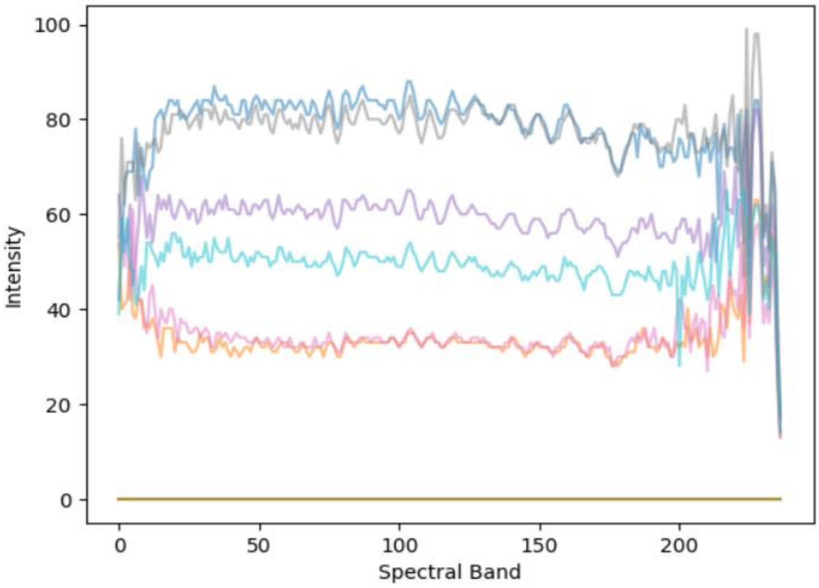

Figure 4: Spectral Variability Across Biofilm Pixels

The confusion matrix developed in the testing phase additionally demonstrated the ability to accurately predict the different individual pathogen classes. The system was shown to be reliable even when using features to predict multiple mixed microbial classes as pathogens. These test results indicate the overall effectiveness of an approach that integrates deep learning with hyperspectral imaging for rapid and non-destructive diagnostics. The results of the classification performance of the proposed model are represented with Confusion matrix as shown in **Figure 5.**

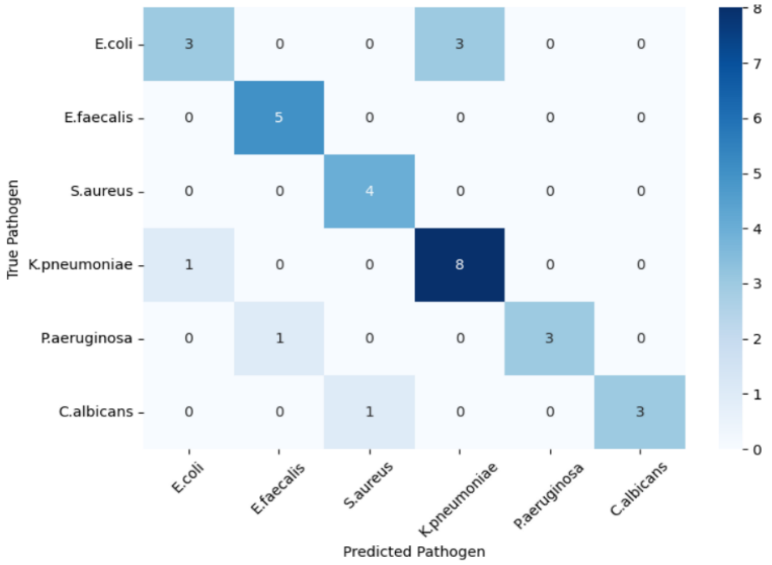

Figure 5: Confusion Matrix

The PCA-reduced hyperspectral image showed that there were distinct spectral variations among the pixels with biofilm, and this helped the model distinguish the spectral variations of different microbial types. **Figure 6** shows the dimensionality reduction of hyperspectral data via Principal Component Analysis (PCA).

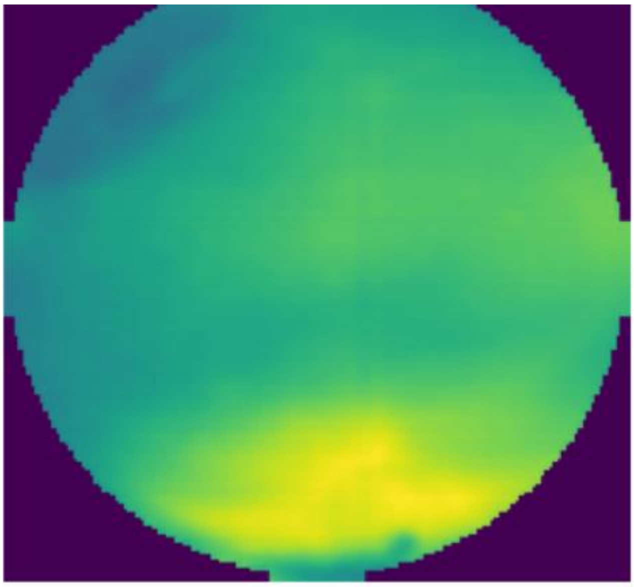

Figure 6: PCA Reduced Hyperspectral Image

4. Conclusion

In this paper we tested an early version of the AI system for automatic detection of biofilm contaminations caused by a group of microorganisms. The results suggest that this approach is capable of quickly and non-destructively detect these contaminations in data-driven approaches. The workflow comprises these key steps - collection of hyperspectral image data, pre-processing, data reduction (using principal component analysis), model development (classification model of CNN type developed with Keras library) and a minimalist interface featuring Flask framework. This all makes up an integrated process from collecting hyperspectral data through to useful information on the presence of contaminants. The model performed well, with training and validation accuracy scores of 97.6% and 85.3% respectively - this means the model can learn to recognise important patterns of spectral and spatial mapping of six types of microorganisms, which ultimately allow the automated identification and characterisation of microorganisms. Although the results here show it's early stages of development, there are a few aspects of the system that should be enhanced to ensure better identification and detection. A key aspect of future work will be to collect more samples with more types of pathogens in order to train a robust model, and test the model using samples with different conditions in different clinics, industries and environments. It's also critical the system is tested with other datasets and biofilm samples to make sure it performs consistently in real world conditions (beyond the lab) and does not overfit the data. Future studies could experiment with a more advanced deep learning approach (such as a CNN-transformer, attention-based continuous spectral-spatial networks and multimodal learning) that includes other types of data (morphological, biochemical, microscopic images etc.) to enhance classification performance. It could also have explainable AI capabilities by showing which wavelengths and portion of the image the model decides the biofilm is of and proving the science behind it - so the user can understand the science behind the model (as well as increasing model trust). As for use cases, there are many possibilities.

Future research could build on this work to develop a mobile device that could be used in hospitals, food processing units, water treatment plants and the environment. And by connecting the system to lightweight hyperspectral sensors, edge computing and alarm systems, microbial contamination could be on-line monitored - as close to real time as possible - which could potentially enhance its detection and response. In summary, this research has shown that combining hyperspectral imaging, artificial intelligence, and an easy-to-use interface has the potential to be a steppingstone towards futuristic microbial pathogen detection systems. The system described in this report is a good starting block, and there is much potential for this work to be continued for more sophisticated, advanced and user-friendly systems for biofilm pathogen detection for many of these applications.

Author Declaration Statements

Declaration: The authors hereby declare that the manuscript submitted for consideration is an original work and has not been published or submitted elsewhere for publication. The authors take full responsibility for the integrity, accuracy, and ethical compliance of the work presented in the manuscript.

AI Usage Statement: Authors declare that AI tools, if used, were solely employed to improve the clarity, grammar, and language of the manuscript (as indicated in the reviewer’s comments). No data, results, or scientific content were generated or altered using AI.

Conflict of Interest: All authors confirm that:

Any potential conflicts of interest, whether financial or non-financial, have been fully disclosed. **– ~~Yes~~ / Not Applicable√**

All sources of funding and financial support received for the conduct of the study have been appropriately acknowledged. **– ~~Yes~~ / Not Applicable√**

Necessary ethical approvals have been obtained from the relevant institutional or regulatory bodies for studies involving human participants, animals, or sensitive data, wherever applicable. **– ~~Yes~~ / Not Applicable√**

AI Usage Statement: Authors declare that AI tools, if used, were solely employed to improve the clarity, grammar, and language of the manuscript (as indicated in the reviewer’s comments). No data, results, or scientific content were generated or altered using AI.

References:

1. Alkhatib, M. Q., Al-Saad, M., Aburaed, N., Almansoori, S., Zabalza, J., Marshall, S., & Al-Ahmad, H. (2023). Tri-CNN: A three branch model for hyperspectral image classification. *Remote Sensing, 15*(2), 316. https://doi.org/10.3390/rs15020316
2. Anju, V. T., Busi, S., Imchen, M., Kumavath, R., Mohan, M. S., Salim, S. A., Subhaswaraj, P., & Dyavaiah, M. (2022). Polymicrobial infections and biofilms: Clinical significance and eradication strategies. *Antibiotics, 11*(12), 1731. https://doi.org/10.3390/antibiotics11121731

3. Dhaene, Z., Žižakić, N., Huang, S., Li, X., & Pižurica, A. (2023). HSIToolbox: A web-based application for the classification of hyperspectral images. *SoftwareX, 22*, 101340. https://doi.org/10.1016/j.softx.2023.101340
4. Elfadadny, A., Ragab, R. F., AlHarbi, M., Badshah, F., Ibáñez-Arancibia, E., Farag, A., Hendawy, A. O., De los Ríos-Escalante, P. R., Aboubakr, M., Zakai, S. A., & Nageeb, W. M. (2024). Antimicrobial resistance of *Pseudomonas aeruginosa*: Navigating clinical impacts, current resistance trends, and innovations in breaking therapies. *Frontiers in Microbiology, 15*, 1374466. https://doi.org/10.3389/fmicb.2024.1374466
5. Kotwal, S., Rani, P., Arif, T., Manhas, J., & Sharma, S. (2021). Automated bacterial classifications using machine learning based computational techniques: Architectures, challenges and open research issues. *Archives of Computational Methods in Engineering, 29*, 2469–2490. https://doi.org/10.1007/s11831-021-09660-0
6. Kranjec, C., Morales Angeles, D., Torrissen Mårli, M., Fernández, L., García, P., Kjos, M., & Diep, D. B. (2021). Staphylococcal biofilms: Challenges and novel therapeutic perspectives. *Antibiotics, 10*(2), 131. https://doi.org/10.3390/antibiotics10020131
7. Liu, X., Yao, H., Zhao, X., & Ge, C. (2023). Biofilm formation and control of foodborne pathogenic bacteria. *Molecules, 28*(6), 2432. https://doi.org/10.3390/molecules28062432
8. Nguyen, C., Sagan, V., Maimaitiyiming, M., Maimaitijiang, M., Bhadra, S., & Kwasniewski, M. T. (2021). Early detection of plant viral disease using hyperspectral imaging and deep learning. *Sensors, 21*(3), 742. https://doi.org/10.3390/s21030742
9. Soni, A., Dixit, Y., Reis, M. M., & Brightwell, G. (2022). Hyperspectral imaging and machine learning in food microbiology: Developments and challenges in detection of bacterial, fungal, and viral contaminants. *Comprehensive Reviews in Food Science and Food Safety, 21*(4), 3717–3745. https://doi.org/10.1111/1541-4337.12983
10. Tavares-Carreon, F., De Anda-Mora, K., Rojas-Barrera, I. C., & Andrade, A. (2023). *Serratia marcescens* antibiotic resistance mechanisms of an opportunistic pathogen: A literature review. *PeerJ, 11*, e14399. https://doi.org/10.7717/peerj.14399
11. Wu, Y., & Gadsden, S. A. (2023). Machine learning algorithms in microbial classification: A comparative analysis. *Frontiers in Artificial Intelligence, 6*, 1200994. https://doi.org/10.3389/frai.2023.1200994
12. Yin, R., Cheng, J., Wang, J., Li, P., & Lin, J. (2022). Treatment of *Pseudomonas aeruginosa* infectious biofilms: Challenges and strategies. *Frontiers in Microbiology, 13*, 955286. https://doi.org/10.3389/fmicb.2022.955286
13. Zhao, A., Sun, J., & Liu, Y. (2023). Understanding bacterial biofilms: From definition to treatment strategies. *Frontiers in Cellular and Infection Microbiology, 13*, 1137947. https://doi.org/10.3389/fcimb.2023.1137947