Research article

PollenNet: A novel architecture for high precision pollen grain classification through deep learning and explainable AI

F M Javed Mehedi Shamrat [a], Mohd Yamani Idna Idris [a], Xujuan Zhou [b,**],
Majdi Khalid [c], Sharmin Sharmin [a], Zeseya Sharmin [d], Kawsar Ahmed [e,f,g,*],
Mohammad Ali Moni [h,i]

[a] Department of Computer System and Technology, Universiti Malaya, Kuala Lumpur, 50603, Malaysia
[b] School of Business, University of Southern Queensland, Toowoomba, Australia
[c] Department of Computer Science and Artificial Intelligence, College of Computing, Umm Al-Qura University, Makkah, 21955, Saudi Arabia
[d] School of Information Technology, Faculty of Science Engineering and Built Environment, Deakin University, Australia
[e] Department of Electrical and Computer Engineering, University of Saskatchewan, 57 Campus Drive, Saskatoon, SK, Canada, S7N 5A9
[f] Group of Bio-photomatix, Information and Communication Technology, Mawlana Bhashani Science and Technology University, Santosh, Tangail, 1902, Bangladesh
[g] Health Informatics Research Lab, Department of Computer Science and Engineering, Daffodil International University, Daffodil Smart City (DSC), Birulia, Savar, Dhaka, 1216, Bangladesh
[h] AI & Digital Health Technology, Artificial Intelligence & Cyber Future Institute, Charles Stuart University, Bathurst, NSW, 2795, Australia
[i] AI & Digital Health Technology, Rural Health Research Institute, Charles Stuart University, Orange, NSW, 2800, Australia

A R T I C L E   I N F O

A B S T R A C T

Pollen grains play a critical role in environmental, agricultural, and allergy research despite their tiny dimensions. The accurate classification of pollen grains remains a significant challenge, mainly attributable to their intricate structures and the extensive diversity of species. Traditional methods often lack accuracy and effectiveness, prompting the need for advanced solutions. This study introduces a novel deep learning framework, PollenNet, designed to tackle the intricate challenge of pollen grain image classification. The efficiency of PollenNet is thoroughly evaluated through stratified 5-fold cross-validation, comparing it with cutting-edge methods to demonstrate its superior performance. A comprehensive data preparation phase is conducted, including removing duplicates and low-quality images, applying Non-local Means Denoising for noise reduction, and Gamma correction to adjust image brightness. Furthermore, Explainable AI (XAI) is utilized to enhance the interpretability of the model, while Receiver Operating Characteristic (ROC) curve analysis serves as a quantitative method for evaluating the model's capabilities. PollenNet demonstrates superior performance when compared to existing models, with an accuracy of 98.45 %, precision of 98.20 %, specificity of 98.40 %, recall of 98.30 %, and f1-score of 98.25 %. The model also maintains low Mean Squared Error (0.03) and Mean Absolute Error (0.02) rates. The ROC curve analysis, the low False Positive Rate (0.016), and the False Negative Rate (0.017) highlight the reliability and dependability of the model. This study significantly

* Corresponding author. Department of Electrical and Computer Engineering, University of Saskatchewan, 57 Campus Drive, Saskatoon, SK, S7N 5A9, Canada.
** Corresponding author.
E-mail addresses: javedmehedicom@gmail.com (F.M.J.M. Shamrat), yamani@um.edu.my (M.Y. Idna Idris), xujuan.zhou@usq.edu.au (X. Zhou), mknfiai@uqu.edu.sa (M. Khalid), sharmin9889@gmail.com (S. Sharmin), zeseya15@gmail.com (Z. Sharmin), kawsar.ict@mbstu.ac.bd, k.ahmed@usask.ca (K. Ahmed), mmoni@csu.edu.au (M.A. Moni).

improves the efficacy of classifying pollen grains, indicating an important advancement in the application of deep learning for ecological research.

1. Introduction

Pollen grains are tiny fragments produced by plants as part of their reproductive process and have diverse shapes. Pollen morphology often exhibits distinguishing characteristics that may be used to identify the family it belongs to. However, in some cases, unique properties may be seen at the genus or species level. Pollen's extensive array of morphological changes makes it significant in several scientific disciplines since it facilitates the precise identification of the plants from which the pollen grains came [1]. Palynology is the scientific field that specifically studies and analyzes pollen grains, spores, and other minuscule biological particles. This area has many applications, including paleoclimate reconstruction [2], paleoecology [3], biodiversity [4], allergy research [5], studies of flora [6], plant reproductive biology [7], petroleum exploration [8], establishing plant taxonomic connections [9], forensic science [10], and more. The study of pollen grains, including their identification, monitoring, classification, and certification, has attracted considerable attention because of their many applications. These applications include archeology and honey certification [11], predicting allergies caused by airborne pollen [12], and food technology [13]. Traditional pollen tests include the analysis of microscopic images and the identification of pollen grains by comparing them to established species, a process performed by palynologists. However, the analysis of pollen is a laborious and challenging procedure that necessitates the physical preparation of samples, the microscopic identification and localization of pollen grains, and ultimately, the quantification of the abundance of various species within a specific sample. This procedure is distinguished by its lengthy duration, susceptibility to human mistakes, and need for substantial expertise in many tree species [14]. The categorization of pollen grains has developed into a costly analytical process that requires the identification and classification of features by a licensed palynologist. Undoubtedly, it continues to be the most exhaustive, precise, and dependable method. Nevertheless, it impedes scientific progress by wasting significant amounts of time and resources [15].

Computer science, including techniques in image processing and machine learning analysis, has lately seen significant advances in several fields of study. This achievement has facilitated a strong partnership between palynologists and computer scientists to tackle the aforementioned issues. The progress in artificial intelligence, specifically in deep learning, combined with the use of image processing, has enabled the execution of complex analysis with remarkable outcomes in several occupations [16]. Advanced Machine Learning approaches, including deep neural networks, have produced impressive outcomes in addressing many problems across several domains, all while maintaining cost-effectiveness [17]. A crucial determinant in attaining success in deep learning is the existence of a significant volume of annotated data. The efficacy of deep learning techniques is directly proportional to the amount of data, promoting the development of comprehensive datasets across many scientific fields [18]. Research on automated systems for identifying pollen grains will enable the development of tools that botanists and palynologists may use. These tools only need a normal bright-field microscope, an image-capturing camera for data collection, and a computer for data processing. This will obviate the need for specialist equipment. These gadgets possess the capacity to separate pollen grains from the images and assist researchers in classifying them according to species. The contributions of this research are as follows.

1. Development of a novel deep learning model to precisely and effectively classify pollen grains.
2. Implementation of advanced image processing techniques, such as Non-local Means denoising for noise reduction and Gamma correction for image improvement.
3. Employing nine metrics for evaluation, including ROC curves, confusion matrices and stratified 5-fold cross-validation to validate the robustness and broad applicability of PollenNet.
4. Explainable AI (XAI) techniques such as Grad-CAM and Saliency Maps have been applied to enhance the interpretability and transparency of PollenNet's decision-making process.

2. Literature review

Over the last several decades, researchers have made significant advancements in creating automated techniques for classifying pollen grains, a critical aspect of environmental and ecological research [30,31]. Traditional methods, although creative, generally relied heavily on explicit rule-based algorithms, which often encountered challenges when handling the intricate and nuanced aspects of pollen grain classification. The incorporation of machine learning and deep learning has profoundly revolutionized this domain. Through data-driven learning, these advanced algorithms have greatly enhanced the efficiency and accuracy of the classification process. Machine learning has shown exceptional adaptability and precision, particularly in pollen grain analysis. It has shown expertise in identifying and categorizing various types of pollen, which is crucial for understanding changes in the environment, monitoring allergies, and studying plant diversity. This technical innovation has not only accelerated research but also generated potential for deeper comprehension in the realm of ecological and environmental studies.

Rebiai et al. [19] have introduced a system for classifying and automatically recognizing multi-focal microscopic pollen images. The framework is meant to be adaptable and employs a strategy similar to brute force to achieve efficiency for different kinds and sizes of pollen images. First, the best focus is selected using the absolute gradient technique. In the end, a standardized method is used to extract and choose characteristics, and their classification is assessed using Hierarchical Cluster Analysis (HCA). Manikis et al. [20]

acquired a classified dataset from the Hellenic Mediterranean University in another research. The collection included 564 microscopic pollen images. The researchers devised a fully automated analysis procedure with the objective of classifying the images into six distinct pollen categories. The task of image categorization was executed with the Random Forest (RF) classifier. The estimation of generalization performance and avoidance of overfitting was conducted utilizing a repeating nested cross-validation (nested-CV) template. The findings indicated an average accuracy rate of 88.24 %. The dataset supplied by Battiato et al. [21] contains more than 13,000 items. The identification of these items was achieved by a customized segmentation approach used to aerobiological samples to classify pollen. The data has undergone preprocessing to reduce significant background noise in the digital slide scans. The experiment is carried out using conventional Machine Learning techniques, such as Random Forest (RF), Support Vector Machines (SVM), Ada-Boost, and Multilayer Perceptron (MLP). The item's texture was analyzed using the Local Binary Pattern (LBP) and Histogram of Oriented Gradient (HOG) methods. The research determines that the RBF-SVM model with HOG obtains the maximum performance among all the models, with an accuracy of 86.58 %.

Similarly, Grant-Jacob et al. [22] proposed detecting and capturing moving pollen grains by shining three laser beams on them. The inceptionV3 method was used to differentiate between two unique types of pollen grains and a category representing no pollen based on their dispersion patterns. The prediction accuracy of this model reached 86 %. In addition, Xu et al. [23] proposed a collaborative learning strategy to improve the precision of pollen segmentation and classification by using a weakly supervised method. The detection model is used to discern regions of pollen in the unprocessed images. To improve the precision of the classification, the pollen grains were partitioned into sections using a pre-trained U-Net model that leverages unsupervised pollen contour features. The segmented pollen regions were fed into a Convolutional Neural Network (CNN) model to get the activation maps. The categorization accuracy reaches a maximum of 86.6 %. Azari et al. [24] conducted research that employed the optical characteristics of pollen species to identify them. The authors evaluated several categorization algorithms in the Python scikit-learn module, such as Logistic Regression, SVM, Multi-layer Perception, and K-Nearest Neighbors. The authors opted for a Logistic Regression model among the available choices since it is easier to understand. The investigation yielded memory ratings ranging from 70 % to 85 % for the majority of classes, except for Corylus, which only attained a recall of 42 %.

In a comparable manner, Punyasena et al. [25] provide an automated methodology for examining pollen. This procedure involves the automated scanning of slides holding pollen samples, followed by the use of CNN models to detect and classify various kinds of pollen automatically. The researchers trained the data using the ResNet34 model. The average accuracy of the fine-tuned ResNet34 model, determined by computing the mean of the 25 separate per-class accuracies, is 89.5 %. Similarly, Menad et al. [26] showcased the use of deep Convolutional Neural Networks (CNN) in categorizing images of pollen grains for the purpose of identification. The neural network has 8 hidden layers. The first 5 layers comprise convolutional neurons responsible for processing image representations, while the following 3 levels are fully connected layers used for image classification. The model was subjected to a combined process of training and testing that spanned 4000 epochs. The model reached its highest level of accuracy, 85.1 %, during the 200th epoch and consistently maintained this level of performance throughout.

In addition, Romero et al. [27] explain a method for analyzing pollen that utilizes optical super-resolution microscopy and machine learning to create a more accurate and efficient approach. The scientists developed three convolutional neural network (CNN) classification models, namely the Maximum Projection Model (MPM), Multislice Model (MSM), and Fused Model (FM). These models were trained on a dataset including pollen samples from 16 distinct classes. The data was categorized, and the resulting average accuracies for all models ranged from 83 % to 90 %, with the FM model obtaining the greatest accuracy of 90.30 %. Daunys et al. [28] developed an approach to train many models, including as ConvNets, RepVGG, ResNet, EfficientNet, Vision transformer (ViT), ViT Compact Convolutional Transformer (ViT-CCT), and VOLO, for the purpose of classifying distinct kinds of pollen. The efficacy of these models was then assessed by comparing their outcomes on test sub-samples that were not used during the training phase. The fusion of RepVGG, ResNet, and VOLO models achieved a peak classification accuracy of 72.4 %.

Our study introduces an innovative approach that utilizes deep learning models to tackle the challenges associated with accurately classifying images of pollen grains. Traditional methods sometimes had difficulties in dealing with the intricate and irregular nature of pollen images, resulting in limited accuracy and adaptability. To address these issues, we focused primarily on creating and using a wide variety of sophisticated models, with a particular emphasis on our distinctive and customized development, PollenNet. Among the models considered, including AlexNet [33], MobileNetV2 [34], VGG16 [35], and ResNet50 [36], this model was selected for its exceptional capabilities in image processing and classification. These skills make it particularly suitable for the complex task of categorizing pollen. PollenNet exemplifies our dedication to advancing the field of pollen analysis. The integrated architecture of this system is purposefully designed to use the capabilities of its components, leading to outstanding accuracy and efficiency in classifying various types of pollen. This endeavor is not just focused on improving the precision of classification; it represents significant progress in the capabilities of pollen grain imaging and analysis. We want to set a new standard in the classification of pollen grains by using cutting-edge models. This will substantially contribute to environmental and ecological studies by providing crucial insights.

3. Research methodology and strategy

Our study's methodology involves a methodical approach to data preparation, which ensures the integrity and clarity of information fed into the deep learning framework. We begin by conducting thorough data preprocessing, utilizing techniques such as data extraction and refinement to remove any redundancies and flaws, resulting in a polished dataset. Subsequently, advanced image processing algorithms, such as non-local means denoising and gamma correction, are applied, which are crucial in reducing noise and improving image fidelity. After this extensive preprocessing, our proposed model, PollenNet, was implemented, and its subsequent responsibility was to classify pollen grains. Fig. 1 demonstrates the entire methodology of the proposed study.

3.1. Data managing and processing techniques

The efficacy of deep learning models heavily depends on the quality of the input data. Considering this, our data preprocessing phase was meticulously designed to improve the quality of the raw dataset. This study incorporates a sequence of cutting-edge techniques, commencing with the essential preprocessing steps of data extraction, refinement, denoising, and enhancement. The original dataset underwent a thorough process of data polishing, during which we systematically identified and eliminated duplicates, unclear images, and distorted images. This comprehensive refinement was crucial in building an adequate framework for the succeeding steps of feature extraction, classification, and examining the hidden variables in our data. Furthermore, we employed Non-local Means to reduce image noise while preserving essential structural details, thus refining the visual clarity of our data. Meanwhile, Gamma Correction was applied to adjust the intensity of the images, allowing our model to detect and learn from the nuanced elements required for precise classification. These preprocessing steps are crucial in transforming raw data into a clean and reliable format suitable for sophisticated analytical tasks.

3.1.1. Data extracting

The study used the New Zealand pollen dataset [29], including 19,667 samples of pollen grains. The dataset comprises dark field microscopic images acquired using the Classifynder system, an autonomous system setup specifically tailored for automated pollen analysis. Designed to function independently, the system leverages fundamental shape descriptors to pinpoint pollen grains on standard microscope slides using a low-power objective lens. Upon identification, it engages a high-power objective lens to systematically capture high-resolution images of each pollen entity across multiple focal planes, creating a 'Z-stack'. The most sharply defined sections of the Z-stack are carefully selected and combined into a single, comprehensive composite image. The final image undergoes a segmentation process to separate the pollen grain from its background. Then, fake coloring techniques are applied to highlight the depth of information, which improves the level of detail for future classification tasks. Fig. 2 shows the sample images from the pollen dataset.

Fig. 1. The entire workflow of the proposed model for pollen classification.

3.1.2. Duplicate image removal

This technique uses the MD5 hash function, a widely used cryptographic hash function that produces a 128-bit (16-byte) hash value. For any given image I, the MD5 hash function H maps it to a hash value h, as shown in equation (1):

$$h = H(I) \tag{1}$$

Since MD5 is a deterministic function, identical images will yield the same hash value, facilitating the identification of duplicates.

Algorithm 1.   MD5-Based Duplicate Image Detection.

Input: D = A directory path containing a set of images $\{I_1, I_2, \ldots, I_n\}$, where each image I_i has a unique identifier and corresponding file path fp_i.
Output: Dup = A list of tuples, where each tuple (I_a, I_b) identifies a pair of duplicate images based on their MD5 has values.

START
Step1: Initialization
    1.1 $S \leftarrow \emptyset$: Initialize an empty set for storing hash values with their corresponding image identifiers.
    1.2 $Dup \leftarrow \emptyset$: Initialize an empty list for storing duplicates.
Step2: Hash Computation and Duplicate Detection
    For each image I_i in D, identified by file path fp_i:
    2.1 Compute the MD5 hash h_i of the content at fp_i:
        a. $h_i = H(fp_i)$, where H denotes MD5 hash function.
    2.2 Check if h_i already exists in S:
        a. If $h_i \in S$, a duplicate has been found:
            • Add $(I_i, S[h_i])$ to Dup, indicating that I_i and image associated with h_i in S are duplicates.
            • Else, add h_i to S with its corresponding image identifier I_i.
Step3: Output
    Return Dup, the list of identified duplicate image pairs.
END

3.1.3. Repeating image removal

The Structural Similarity Index (SSIM) is employed for repeating image removal. For two images I_1 and I_2, SSIM is defined as shown in equation (2):

$$SSIM(I_1, I_2) = \frac{(2\mu_{I_1}\mu_{I_2} + C_1)(2\sigma_{I_1 I_2} + C_2)}{(\mu_{I_1}^2 + \mu_{I_2}^2 + C_1)(\sigma_{I_1}^2 + \sigma_{I_2}^2 + C_2)} \tag{2}$$
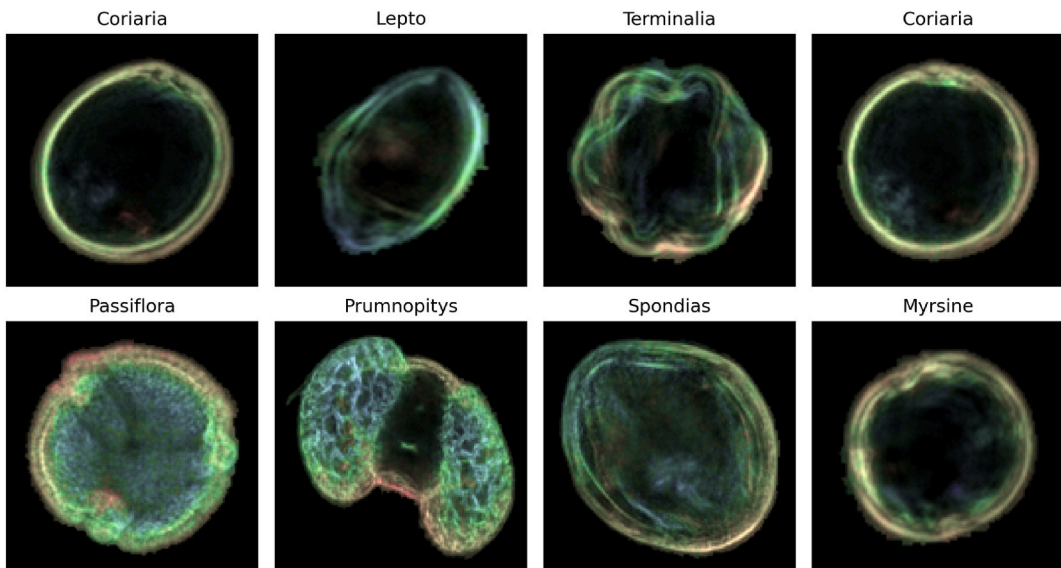

Fig. 2. Samples of the 46 classes retrieved from the selected pollen dataset.

Where μ_{I_1}, μ_{I_2} are the average pixel values, $\sigma_{I_1}^2$, $\sigma_{I_2}^2$ are the variances and $\sigma I_1 I_2$ is the covariance of the images. C_1 and C_2 are constants to stabilize division with weak denominators.

Algorithm 2. Structural Similarity Index (SSIM) Based Repeating Image Removal Algorithm.

Input: D = A directory containing a set of images $\{I_1, I_2, \ldots, I_n\}$.
      T_{SSIM} = A threshold value for SSIM, above which two images are considered similar.
Output: A reduced set of images $\{I'_1, I'_2, \ldots, I'_m\}$ from D.

START
Step1: Initialization
      1.1 R : Empty list to store the non-repeating images after processing.
      1.2 Populate a list L with the paths of all images in D.
Step2: Compute SSIM for Each Image Pair
      2.1 *For* each image I_i in L:
           a. *For* each subsequent image I_j in L where $j > i$, compute $SSIM\ (I_i, I_j)$.
              If $SSIM\ (I_i, I_j) > T_{SSIM}$, mark I_j as repeating.
           b. *If* I_i is not marked as repeating by any previous step, add it to R.
Step3: Output
      Return R, the list of images with visually similar repeats removed.
END

3.1.4. Unclear and distorted image removal

Blurriness is quantified using the variance (σ^2) of the Laplacian (∇^2) of an image. The Laplacian of an image I, denoted as $\nabla^2 I$, measures its second spatial derivative. The variance of the Laplacian is given by equation (3):

$$\sigma^2(\nabla^2 I) \tag{3}$$

Color variance is assessed by calculating the standard deviation (σ) of the pixel intensities across each color channel. For a color channel C of an image I, the standard deviation can be derived using equation (4):

$$\sigma(C) \tag{4}$$

A low mean of these standard deviations across all color channels indicates color uniformity, which is often a sign of distortion. These procedures can be observed in Algorithm 3.

Algorithm 3. Quantitative Assessment of Image Quality Through Blurriness and Color Variance Evaluation.

Input: A collection of images $\{I_1, I_2, \ldots, I_n\}$.
Output: Blurriness score for each image based on $\sigma^2(\nabla^2 I)$.
      Color variance score for each image based on $\sigma(C)$.

START
1: *For Each Image I* in the dataset:
      1.1 Compute the Laplacian $\nabla^2 I$ and then calculate its variance $\sigma^2(\nabla^2 I)$ to assess blurriness.
      1.2 For each color channel C in I, compute the standard deviation $\sigma(C)$ to assess color variance.
      1.3 Calculate the mean of the standard deviations obtained from 1.2 across all color channels
           for I.
      1.4 Store the blurriness score and the mean color variance for each image.
2: *Return* the scores for all images.
END

3.1.5. Non-local means denoising

This study employs the Non-local Means Denoising technique to reduce noise within the pollen image dataset, which is categorized into 46 groups. This method is chosen for its capability to diminish noise while preserving important image details effectively. The underlying principle of Non-local Means Denoising is that the value of a pixel can be more accurately estimated by a weighted average of all other pixel values in the image rather than only considering immediate neighbors. The denoised value $D(x)$ of a pixel at position x in a noisy image I is determined by the relationship presented in equation (5):

$$D(x) = \frac{1}{C(x)} \sum_{y \in I} w(x, y) . I(y) \tag{5}$$

Here, $w(x, y)$ represents the weight determined by the similarity between small patches centered around the pixels at x and y, respectively. Furthermore, $C(x)$ is the normalization factor. Algorithm 4 provides a straightforward overview of the procedural steps required to execute perform of Means Denoising.

Algorithm 4. Efficient Implementation of Non-local Means Denoising for Enhanced Image Clarity.

Input: I = Noisy image from the dataset.
      P = Patch size.
      h = Filtering parameter.
Output: Denoised image D.

START
Step1: Initialization
    for each pixel x in I, set $D(x) = 0$ and $C(x) = 0$.
Step2: Denoising noisy image
    for each pixel x in I, iterate through all pixels y to:
      Calculate $w(x,y) = exp\left(-\frac{\|P(x)-P(y)\|^2}{h^2}\right)$.
      Update $D(x)$ by $D(x) + w(x,y) \times I(y)$.
      Update $C(x)$ by $C(x) + w(x,y)$.
Step3: *Finalize* $D(x)$ for all x by dividing with $C(x)$.
Return the image.
END

3.1.6. Gamma correction

Gamma correction is a crucial non-linear process in image processing used to encode and decode the brightness, also known as tristimulus values in images. The image brightness is adjusted by altering pixel intensity values, as described by the power-law formula in equation (6).

$$I_{corrected} = I_{original}^{\gamma} \tag{6}$$

Here, '$I_{original}$' represents the original intensity of a pixel, γ denotes the gamma value, and $I_{corrected}$ is the gamma-corrected intensity of the pixel. The adjustment of the gamma value enables systematic supervision of image luminance. When the value of γ is less than 1, the image is intensified, thereby enhancing visibility in regions with lower brightness. On the contrary, when the value of γ surpasses 1, the image exhibits a decrease in brightness, hence enhancing the distinction in areas characterized by increased luminosity. Algorithm 5 represents the procedure of gamma correction.

Algorithm 5. Gamma Correction Procedure on pollen images.

Input: I = Original image matrix from the dataset.
      γ = Patch size.
Output: G = Gamma-corrected image.

START
Step1: Preparation
    for each pixel i in I, set $G(i) = 0$.
Step2: Apply Gamma Correction
    for each pixel i in image I:
      Normalize pixel value: $n_i = I(i)/255$.
      Apply gamma correction: $g_i = n_i^{\gamma}$.
      Scale back to range 0-255 and assign to (i) : $G(i) = g_i * 255$.
Step3: *Finalization*
      Ensure all values in G are within the range [0, 255].
      Convert G to the appropriate image data type.
Return the image G.
END

Fig. 3 displays the original image, the image after preprocessed, and a third image that emphasizes the enhanced areas. The 'Enhanced Area' image will augment the visibility of the differentiations, especially in regions where noise has been effectively reduced and enhanced the features. The histogram compares the distribution of pixel intensities across the channels before and after image processing, illustrating the effect of enhancement techniques on image contrast and clarity.

3.1.7. Dataset split

This study initially divided the dataset into a training and validation set (90 %) and a testing set (10 %) to ensure robust model evaluation. The total dataset consisted of 19,612 images after applying various preprocessing steps, including removing unclear and distorted images, duplicate images, and repeating image techniques. This resulted in a final dataset distribution of 17,672 images for training and validation and 1940 for testing.

To further ensure the robustness of the model evaluation, stratified 5-fold cross-validation [42] was applied to the training and validation set. This method involves dividing the dataset into five equally sized folds while maintaining the class distribution in each fold. Each iteration uses four folds for training and one for validation. This process is repeated five times, with each fold being used
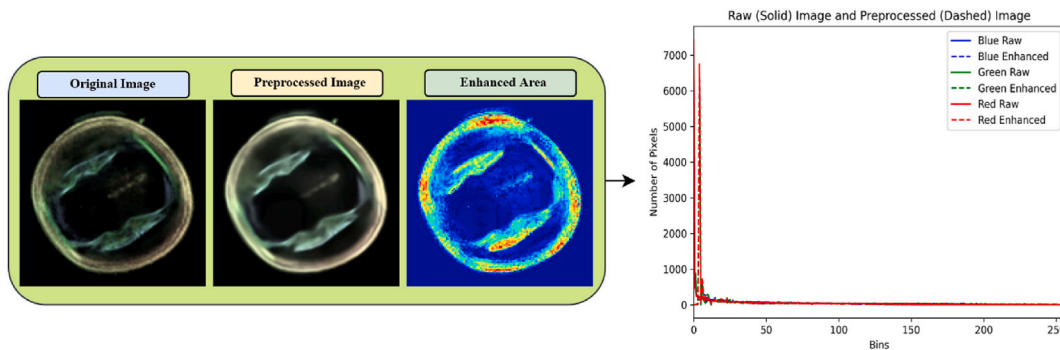
Fig. 3. Demonstrates the original image, its preprocessed version, and the enhancement map alongside a histogram detailing the pixel intensity adjustments across the channels.

once as the validation set. This approach helps rigorously evaluate and tune the model, ensuring that the class distribution remains consistent across all folds and reducing the risk of overfitting. The stratified 5-fold cross-validation ensures the proposed model is tested on all data points while preserving the proportion of classes in each fold. This method comprehensively evaluates the model's performance, allowing for fine-tuning of hyperparameters and improving generalization. Fig. 4 depicts the distribution of the final datasets, illustrating the balanced allocation of images across the training, validation and test sets.

3.2. Proposed model (PollenNet) structure and design

The PollenNet model is a customized Convolutional Neural Network (CNN) particularly developed for classifying pollen grains based on microscopic images. The system architecture is designed to precisely capture the intricate characteristics of pollen grains, which are crucial for proper classification.

The network accepts input images I of shape $128 \times 128 \times 3$. The model includes four convolutional blocks, each comprising a convolutional layer, a max pooling layer, and batch normalization. The first convolutional block, applies 32 filters F_1, each of size 3×3, to the input image I using the convolutional operation, as given in equation (7):

$$C_1 = ReLU \ (I * F_1 + b_1) \tag{7}$$

Where * denotes the convolutional operation, b_1 is the bias vector, and ReLU is the activation function. Furthermore, reduces the spatial dimensions by applying a 2×2 max pooling operation. Then, normalizes the output of the pooling layer.

Subsequent convolutional Blocks (2nd, 3rd and 4th), follow the similar pattern with the increasing number of filters F_2, F_3, F_4 and include dropout layers with a rate of 0.5 after the 2nd and 4th blocks. The output and final convolutional block is flattened into a single vector V, transforming the 2D feature maps into 1D feature vector.
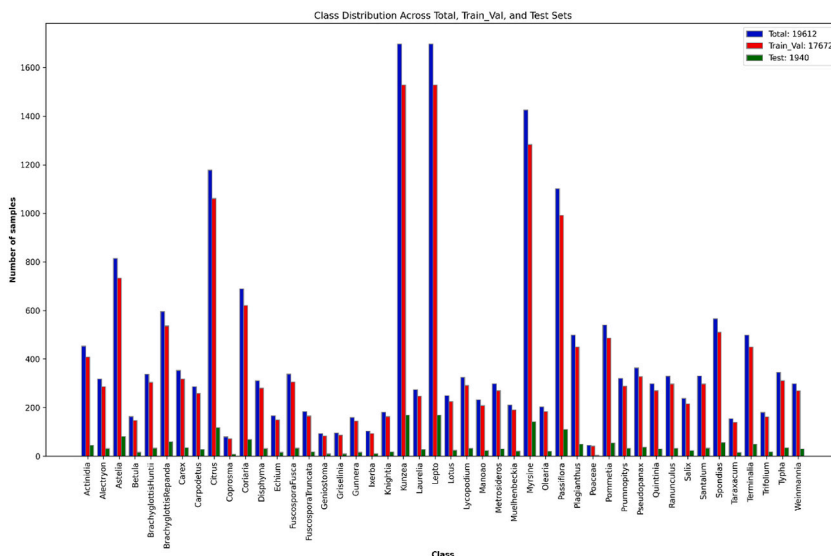

Fig. 4. Detailed overview of the final pollen dataset.

Then, the flattened vector *V* is passed through dense layers. The first dense layer transforms the vector, as represented in equation (8):

$$D_1 = ReLU(W_{d1}V + b_{d1}) \tag{8}$$

Where W_{d1} and b_{d1} are the weights and biases of the dense layer, respectively.

The final dense layer with the SoftMax activation function classifies the image into 46 categories. To convert the linear output (logits) into a probability distribution, the logits are computed by applying a linear transformation to D_1, as indicated in equation (9):

$$Logits = W_o D_1 + b_o \tag{9}$$

Where W_o and b_o are the weights and biases of the output layer.

The SoftMax function is then applied to these logits to obtain the probability distribution, as illustrated in equation (10):

$$O = Softmax\,(Logits) \tag{10}$$

The SoftMax function itself is mathematically defined by equation (11):

$$Softmax\,(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{N} e^{x_j}} \tag{11}$$

Where x_i is the ith logit corresponding to the ith class, N is the total number of classes, and the index j iterates over each class in the output layer, ensuring that the dominator is the sum of the exponential values of all class logits. Algorithm 6 illustrates the detailed architecture of the proposed model.

Algorithm 6. PollenNet for Pollen Grain Classification.

```
Input: I = Input image matrix of shape 128×128×3.
Output: O = Output probability distribution over 46 pollen categories.

START
Step1: Convolutional Blocks
       set C_O = I.
       for each k in {1, 2, 3, 4} do
           F_k = 2^(k+4) (32, 64, 128, 256 filters respectively)
           Apply convolution: C_k = ReLU (Conv2D (C_(k-1), F_k, 3 × 3) + b_k)
           Apply max pooling: C_k = MaxPooling2D (C_k, 2 × 2)
           Apply batch normalization: C_k = BatchNormalization(C_k)
           If k in {2, 4} then
               Apply dropout: C_k = Dropout (C_k, 0.5)
               end if
               end for
Step2: Dense Layers and Classification
       Flatten output: V = Flaten (C_4)
       Apply first dense layer: D_1 = ReLU(Dense (V, 512, W_d1) + b_d1)
       Apply dropout: D_1 = Dropout(D_1, 0.5)
       Compute logits: Logits = Dense(D_1, 46, W_o) + b_o
        Apply softmax: O = softmax(logits)
Step3: Finalization
          Ensure all values in G within the range [0, 255].
          Convert G to the appropriate image data type.
Return O
END
```

In the Pollen Net model, each convolutional block is characterized by an increasing number of filters, denoted by F_k, where k represents the block index. The quantity of filters doubles with each subsequent convolutional block to capture more complex features. The output of each block is represented by C_k, with the subscript indicating the block's sequential position in the network. The convolutional layers are together with the ReLU activation function, max-pooling operations for spatial dimension reduction, and batch normalization to standardize the inputs to each layer. Additionally, dropout is strategically applied after 2nd and 4th blocks to prevent overfitting by randomly zeroing a subset of features.

Following the progression through the convolutional blocks, the final output is flattened into a one-dimensional vector *V*, providing a format suitable for dense layer processing. The first dense layer, characterized by the weights W_{d1} and biases b_{d1}, transforms the vector with a ReLU activation function to create D_1, a representation that encapsulates the learned high-level features. The model then computes the logits, a set of raw prediction values that are not yet normalized, using the weights W_o and biases b_o of the output layer.

The final step in classification is the application of the softmax function, which converts the logits into a normalized probability distribution *O* over the possible pollen categories, ensuring that the sum of probabilities for all classes equals one. This is achieved by exponentiating each logit and dividing by the sum of all exponentiated logits, as encapsulated by the softmax operation in the network. Table 1 illustrates the detailed specifications of the proposed model PollenNet.

Fig. 5 portrays the detailed architecture of the proposed model. The model comprised four convolutional blocks, each comprising a Conv2D layer with a ReLU activation function. Following the convolutional operations, max-pooling layers are employed to reduce

Table 1

Detailed layer-by-layer architecture and parameter specification of the PollenNet convolutional neural network model for pollen grain image classification.

Layer (Name)	Stride & Padding	Output Shape	Karnel	Activation Function	Dropout	Parameters
Input Layer	–	$128 \times 128 \times 3$	–	–	–	0
Conv 1 (Conv2D)	Stride: 1, Padding: Same	$128 \times 128 \times 32$	3×3	ReLU	–	896
Max Pooling 1	Stride: 2, Padding: Valid	$64 \times 64 \times 32$	2×2	–	–	0
Conv 2 (Conv2D)	Stride: 1, Padding: Same	$64 \times 64 \times 64$	3×3	ReLU	–	18,496
Max Pooling 2	Stride: 2, Padding: Valid	$32 \times 32 \times 64$	2×2	–	–	0
Conv 3 (Conv2D)	Stride: 1, Padding: Valid	$30 \times 30 \times 128$	3×3	ReLU	0.5	73,856
Max Pooling 3	Stride: 2, Padding: Valid	$15 \times 15 \times 128$	2×2	–	–	0
Conv 4 (Conv2D)	Stride: 1, Padding: Same	$15 \times 15 \times 256$	3×3	ReLU	0.5	295,168
Max Pooling 4	Stride: 2, Padding: Valid	$7 \times 7 \times 256$	2×2	–	0.5	0
Flatten Layer	–	12544	–	–	–	0
Dense (Fully Connected)	–	512	–	ReLU	0.5	6423040
Dense (Output)	–	46	–	softmax	–	23598

spatial dimensions and parameter counts, thus increasing the computational efficiency. Furthermore, Dropout layers are strategically placed after specific convolutional layers to reduce overfitting by randomly deactivating a fraction of neurons during training. Then, flattening the multidimensional convolutional features into a one-dimensional vector is a crucial step before the classification process. The flattened vector is then fed into a dense layer with ReLU activation followed by a dropout layer to enforce the regularization further. The final dense layer uses the softmax activation function to distribute the probabilities across 46 distinct pollen categories where each neuron corresponds to one category.

4. Result and analysis

This section presents a comprehensive overview of the study's findings. It begins with a performance analysis of the applied models, highlighting their effectiveness and comparative advantages. The complexity of the model's predictive abilities is further explored through confusion matrices and ROC curves, serving as illustrative measures of classifier performance. The assessment approach is supported by a robust stratified 5-fold cross-validation process, ensuring the results' reliability and generalizability. Explainable AI (XAI) is employed to provide clarity on the interpretative aspects of model predictions, offering a clear perspective on the underlying decision-making processes. We deliberately selected four pre-trained distinct models in our analysis: AlexNet, MobileNetV2, VGG16, and ResNet50. The models were chosen based on their distinct abilities and proficiencies in the domains of image processing and classification. The proposed PollenNet model, specifically developed for this work, was intended to tackle the unique challenges of categorizing pollen grain images. The selection of AlexNet, an early deep learning model, was based on its well-documented effectiveness in image recognition tasks, which serves as a dependable benchmark for comparison. The MobileNetV2 model, known for its effectiveness and speed, was chosen to assess performance in a resource-constrained setting. The decision to include VGG16 was based on its renowned characteristics of simplicity and depth and its remarkable capacity to record visual textures and patterns. Ultimately, ResNet50, which incorporates a residual learning framework, was selected to evaluate the effectiveness of more complex network architectures in this particular domain. The integration of these models provided a comprehensive analysis of several architectural techniques, extending from traditional to innovative, in the field of pollen grain image classification [32]. By implementing a varied
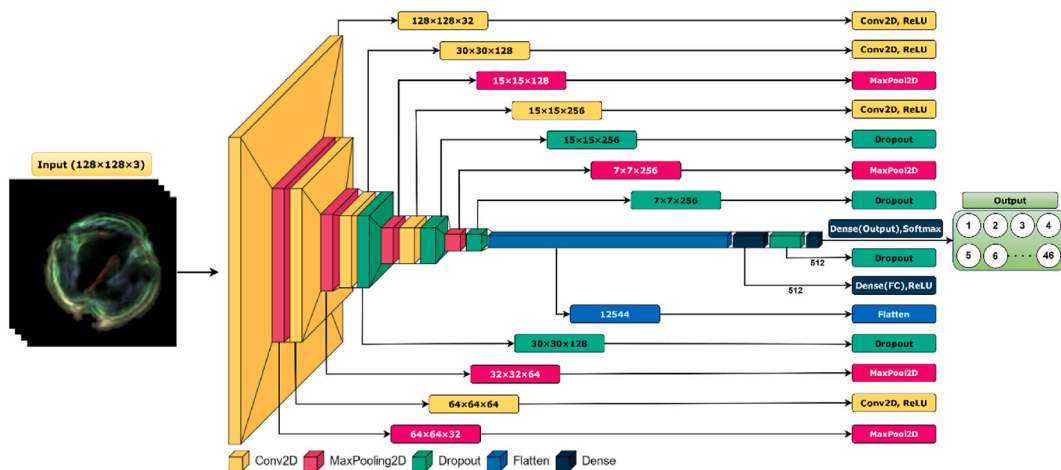

Fig. 5. Schematic representation of the PollenNet architecture models for grain classification, detailing the convolutional, pooling, dropout and dense layer.

and comprehensive method, we successfully evaluated and compared many techniques, augmenting the scope and comprehensiveness of our study's findings and conclusions.

The research encompassed five algorithms, including the proposed PollenNet model and four existing transfer learning models. The algorithms were implemented to carry out the crucial task of detecting and classifying images of pollen grains. The classification of pollen grain images is essential in several scientific and environmental contexts, as it significantly influences the understanding of pollen-related allergies, the monitoring of ecological changes, and the progress of agriculture and climatology. Accurately identifying and categorizing pollen grains is essential for monitoring plant diversity and may aid in predicting allergic reactions, significantly impacting public health and environmental studies. To fully evaluate these models, each one underwent intensive training for 100 epochs, and their performance was meticulously documented at every stage. The rigorous training procedure was essential for accurately evaluating the ability of each model to analyze and classify complex pollen grain images, which display diverse morphological characteristics. The effectiveness and efficiency of these models in classifying pollen grains were evaluated using certain performance measures, calculated using equations (12)–(20). This approach provides a comprehensive comprehension of the performance of each algorithm when used for the complex task of analyzing pollen grain images. This work is crucial for advancing our understanding of environmental and ecological processes.

4.1. Environmental setup

The deep learning models used in our study were principally developed using the TensorFlow (version 2.0) frameworks and the Keras library (version 2.10.0), both of which were coded in Python version 3.7. We used the Seaborn and Matplotlib tools to enhance data visualization, which is renowned for creating detailed and intuitive graphical representations of our results. The computational evaluation of these models was conducted utilizing a system equipped with high-performance computing capabilities. The machine used an AMD Ryzen 7 CPU operating at a clock speed of 3.90 GHz. In addition, it had 32 GB of RAM, providing enough capacity for processing large datasets and complex calculations. We used an AMD Radeon RX 580 series GPU for graphic processing, which confers significant benefits for deep learning tasks. To ensure the stability and compatibility of all software components used in our investigation, we conducted the whole operation inside a Windows 10 environment.
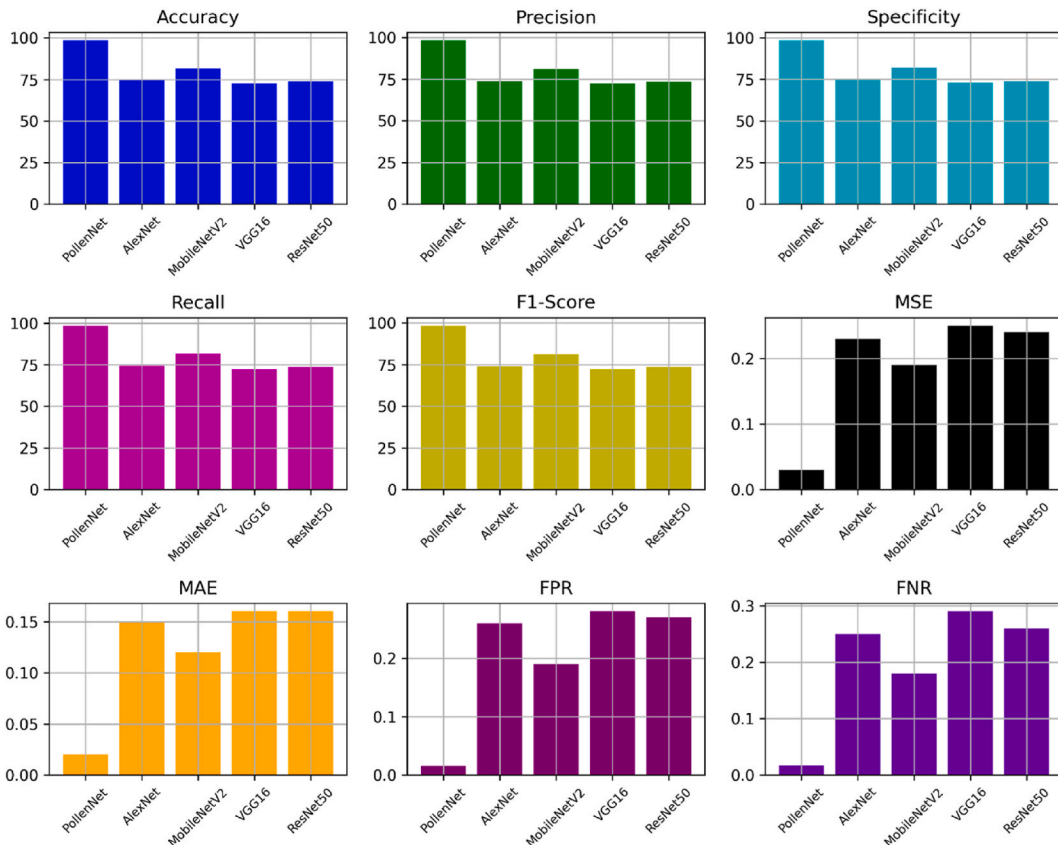

Fig. 6. Performance comparison among all applied models for pollen grain classification.

4.2. Evaluation metrics

We used a range of statistical measures, including Accuracy, Specificity, Recall, Precision, False Positive Rate (FPR), F1-score, Mean Squared Error (MSE), and Mean Absolute Error (MAE), to thoroughly evaluate our model's performance in classifying pollen images [37].

The model's accuracy is defined as the ratio of accurate predictions to the total number of predictions produced.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{12}$$

Here, TP = True Positives, TN = True Negatives, FP = False Positives and FN = False Negatives.

Denotes the proportion of pertinent instances out of the total instances that were predicted as positive by the model. It offers valuable insight regarding the accuracy of positive predictions.

$$Precision = \frac{TP}{TP + FP} \tag{13}$$

The parameter denotes the percentage of true positive instances that the model successfully predicted, thereby showcasing its capability to identify positive cases.

$$Recall = \frac{TP}{TP + FN} \tag{14}$$

This metric quantifies the true negative rate, which serves as an indicator of the model's efficacy in precisely classifying the negative instance among all other instances.

$$Specificty = \frac{TN}{TN + FP} \tag{15}$$

The percentage of negative instances that are erroneously categorized as positive.

$$FPR = \frac{FP}{FP + TN} \tag{16}$$

Achieves harmony between Precision and Recall through their integration into a single metric. Additionally, this metric evaluates the capability of a model to distinguish between positive and negative data.

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{17}$$

The FNR is a metric that measures the proportion of actual positive instances that were incorrectly classified as negative by the model.

$$FNR = \frac{FN}{FN + TP} \tag{18}$$

The difference between the estimator and what is estimated is calculated as the average of the squares of the errors or deviations.

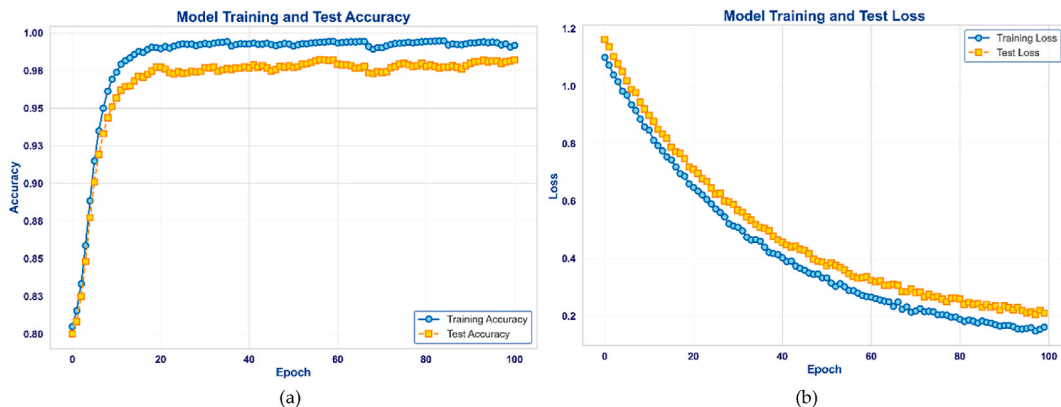$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2 \tag{19}$$

Fig. 7. Demonstrates the (a) accuracy and (b) loss of the proposed PollenNet model.

Where y_i is the actual value, $\widehat{y}_i$ is the predicted value and n is the number of observations.

Evaluates the average magnitude of the errors in a set of predictions without considering their direction. It's the mean over the test sample of the absolute differences between prediction and actual observation, where all individual differences have equal weight.

$$MAE = \frac{1}{n}\sum\nolimits_{i=1}^{n} |y_i - \widehat{y}_i| \tag{20}$$

4.3. Performance evaluation of all applied models

This study thoroughly evaluated five distinct models to determine their efficacy in precisely classifying pollen images with high dependability. Fig. 6 displays the performance of all the models. After analyzing the models 'PollenNet,' 'AlexNet,' 'MobileNetV2,' 'VGG16,' and 'ResNet50,' we have identified significant attributes and benefits in their performance metrics. PollenNet demonstrates outstanding accuracy, earning an impressive test accuracy of 98.45 %, validation accuracy of 98.60 % and a precision of 98.20 %. PollenNet's exceptional precision is shown in its ability to classify positive instances when accurately categorizing pollen grains. Furthermore, the impressive specificity of 98.40 % showcases its strong ability to precisely identify negative samples, which is essential for reducing the occurrence of false positives in critical classification tasks.

In comparison, the other models, such as AlexNet, MobileNetV2, VGG6 and ResNet50, demonstrated varying but typically lower performance levels. AlexNet, despite showing a decent recall of 74.30 %, lacked test accuracy at 74.80 %, indicating problems in consistently detecting true positive instances. Furthermore, MobileNetV2 and VGG6, with test accuracies of 81.60 % and 72.68 %, respectively and moderate F-scores, exhibited a balance between precision and recall but failed to match PollenNet's balanced and high-performance results. ResNet50, achieving test accuracy of 73.85 % and specificity of 74.10 %, demonstrated the lack of capacity during pollen grain classification. All models have F1 scores, demonstrating an appropriate compromise between accuracy and recall. Among them, PollenNet has the highest score of 98.25 %. The models demonstrate different performance levels when assessed using Mean Squared Error (MSE) and Mean Absolute Error (MAE) as measures of error rates. PollenNet has a Mean Squared Error (MSE) of 0.03 and a Mean Absolute Error (MAE) of 0.02, indicating a significant degree of precision in its predictive capabilities. PollenNet excels in the crucial aspect of the False Positive Rate, with a very low FPR of 0.016 and False Negative Rate (FNR) of 0.017. This demonstrates its effectiveness in decreasing incorrect positive classifications.

PollenNet stands out because of its remarkable accuracy, precision, and low false positive rate. These specific features make it ideal for the complex and precise task of classifying pollen grains. While the different models have distinct advantages, they also have areas that might be improved, such as reaching an acceptable balance between accuracy, recall, and reducing error rates. Fig. 7 illustrates the performance metrics of the proposed PollenNet model during the training and testing phase. Subfigure (a) depicts the model's training and test accuracy over 100 epochs, displaying an upward trend in accuracy that eventually levels off at a high level, indicating the model's effective learning from the data. Subfigure (b) illustrates the training and test loss. The training loss decreases rapidly and maintains a low value as the epochs of number increase, indicating effective learning by the model. The test loss also decreases initially but begins to fluctuate as the epoch progresses, a typical occurrence during performance on new, unseen data.

These results provide valuable insights into each model's pros and cons in identifying pollen grain images. By analyzing these patterns, we may get a more thorough knowledge of the model's ability to distinguish distinct types of pollen. This component is crucial in environmental monitoring, allergy prediction, and plant biodiversity studies. Furthermore, Table 2 summarizes the performance metrics of the five models used in the investigation. Here, the training accuracy is denoted as 'T_Acc', validation accuracy is denoted as 'V_Acc' and test accuracy is denoted as 'T_Ac'.

The results of our study demonstrate that the PollenNet model showed significant promise, as it successfully categorized images of pollen grains. PollenNet had an impressive test accuracy rate of 98.45 %, combined with a precision of 98.20 %, indicating its strong ability to categorize various types of pollen precisely. The model demonstrated a remarkable degree of specificity, achieving a rate of 98.40 %, and a recall rate of 98.30 %, indicating its ability to reliably classify positive instances while minimizing the incidence of false negatives. The F1-score, a metric that captures the compromise between accuracy and recall, was calculated to be 98.25 %, showing the model's overall robustness. When analyzing error measurements, PollenNet demonstrated outstanding accuracy with a low Mean Squared Error (MSE) of 0.03 and Mean Absolute Error (MAE) of 0.02. The model exhibited an impressively low False Positive Rate (FPR) of 0.016 and a False Negative Rate (FNR) of 0.017, indicating its exceptional accuracy in accurately classifying and differentiating various pollen types. Throughout the 5-fold cross-validation process, PollenNet excelled consistently, showcasing its robustness and reliability. Eventually, the model outperformed all applied pre-trained models in the comparison and demonstrated its superiority in the given classification task.

Table 2
Final overview of findings of all applied models for pollen image classification.

Model	T_Acc	V_Acc	T_Ac	Precision	SPE	Recall	F1-S	MSE	MAE	FPR	FNR
PollenNet	99.50 %	98.60 %	98.45 %	98.20 %	98.40 %	98.30 %	98.25 %	0.03	0.02	0.016	0.017
AlexNet	80.47 %	75.20 %	74.80 %	73.90 %	74.80 %	74.30 %	74.10 %	0.23	0.15	0.26	0.25
MobileNetV2	84.91 %	82.10 %	81.60 %	81.10 %	82.00 %	81.60 %	81.35 %	0.19	0.12	0.19	0.18
VGG16	77.35 %	73.00 %	72.68 %	72.25 %	73.10 %	72.50 %	72.37 %	0.25	0.16	0.28	0.29
ResNet50	79.50 %	74.00 %	73.85 %	73.50 %	74.10 %	73.70 %	73.60 %	0.24	0.16	0.27	0.26

Fig. 8. Confusion Matrix for PollenNet model Classification. The matrix illustrates the prediction distribution across actual classes with a pronounced diagonal indicating a high rate of correct classifications and minimal off-diagonal entries signifying few misclassifications.

Fig. 9. Analysis of ROC curve and AUC scores for multi-class classification of the proposed model where each curve represents one of the classes.

4.4. Confusion matrix and ROC curve

During the testing phase of our research, we evaluated the effectiveness of the models by utilizing confusion matrices to analyze the patterns of accurate categorization and inaccurate categorization for images of pollen grains. The models were assessed using a comprehensive test dataset comprising 1940 samples across 46 distinct classes of pollen. Fig. 8 presents the confusion matrix of the PollenNet model, which provides a detailed summary of its classification performance for the various types of pollen grains.

The axes of the confusion matrix correspond to the actual and predicted class labels. Diagonal entries in the matrix represent correctly classified samples, highlighted by darker shades, while off-diagonal cells represent instances of misclassification, with lighter shades indicating a lower frequency of occurrences. A prominent dark diagonal line is observed in the matrix, reflecting PollenNet's
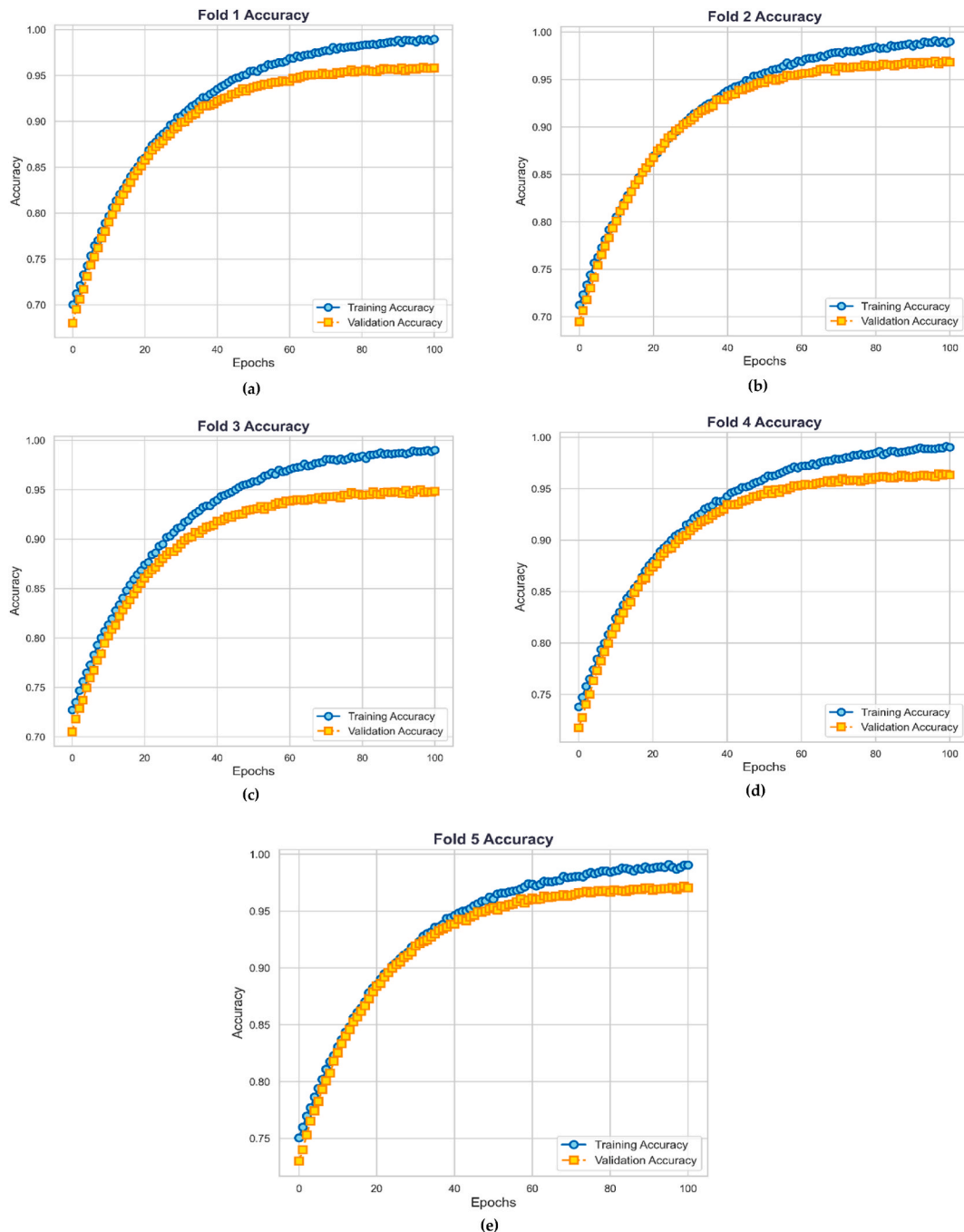
Fig. 10. Each graph (a–e) represents one of the 5-folds, showing training and validation accuracy over epochs.

high accuracy in correctly classifying the majority of samples.

For instance, the class "Kunzea" had 169 test samples, of which 166 were correctly classified, reflecting the model's high precision for this class. Similarly, "Lepto" had 169 test samples, with 166 correctly classified. The "Citrus" class had 117 test samples, and the proposed model accurately identified 115. Classes such as "Alectryon", with 31 samples, and "Astelia", with 81 samples, also show high correct classification rates, further emphasizing the model's precision and recall capabilities. These high numbers and the diagonal for most classes signify PollenNet's robustness in correctly identifying pollen grains. The lighter shades in off-diagonal cells indicate fewer misclassifications, such as in the "FuscosporaFusca" and "Pseudopanax" classes, where the misclassification was negligible. Overall, the confusion matrix shows that PollenNet achieved an impressive accuracy rate of 98.45 %. This high accuracy is crucial for applications requiring precise identification of pollen grains, underscoring PollenNet's effectiveness and reliability.

Fig. 9 shows the comprehensive ROC curves for the PollenNet model, evaluating its classification performance across 46 different pollen grain classes. Each curve represents the trade-off between the True Positive Rate (Sensitivity) and False Positive Rate (1-
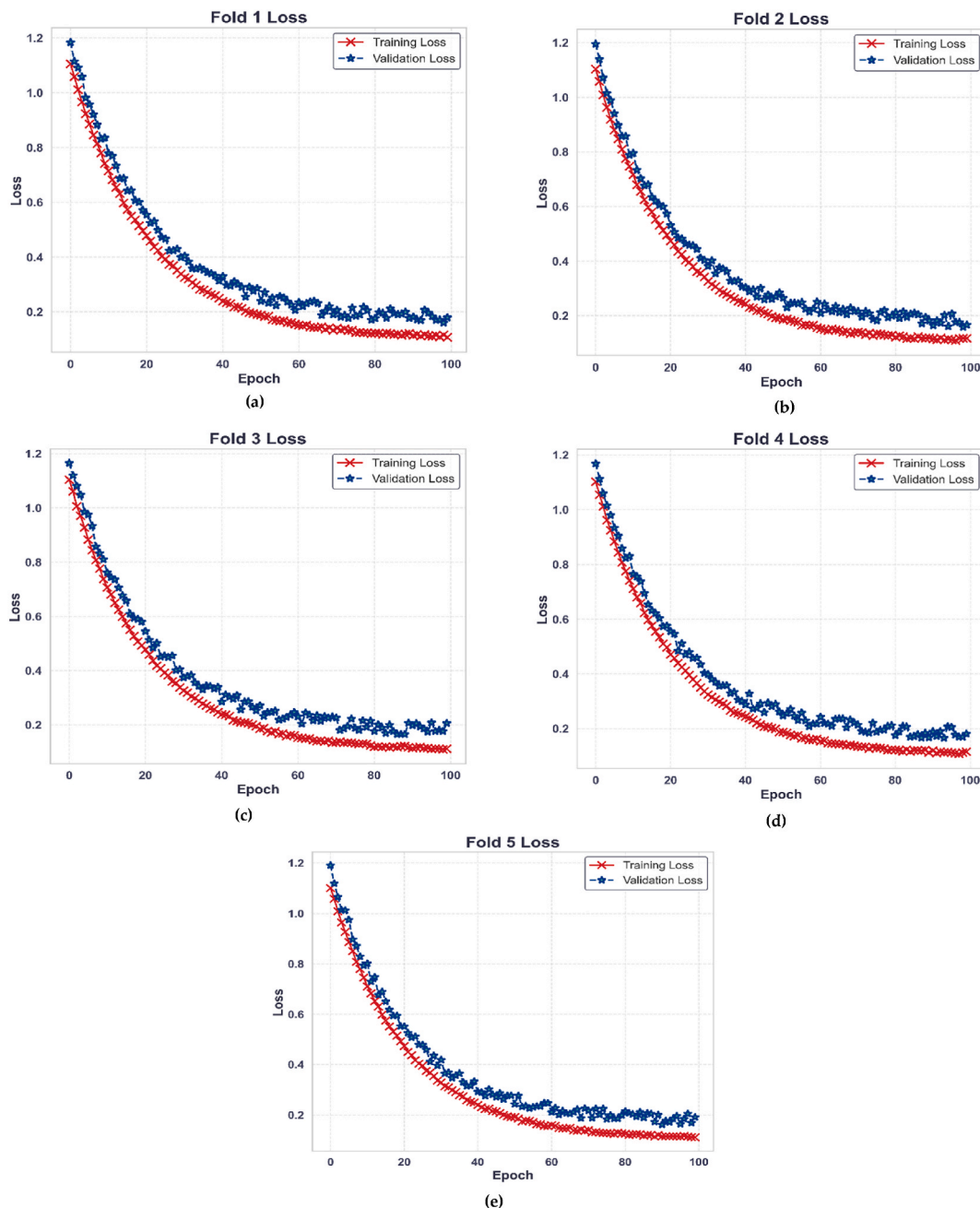

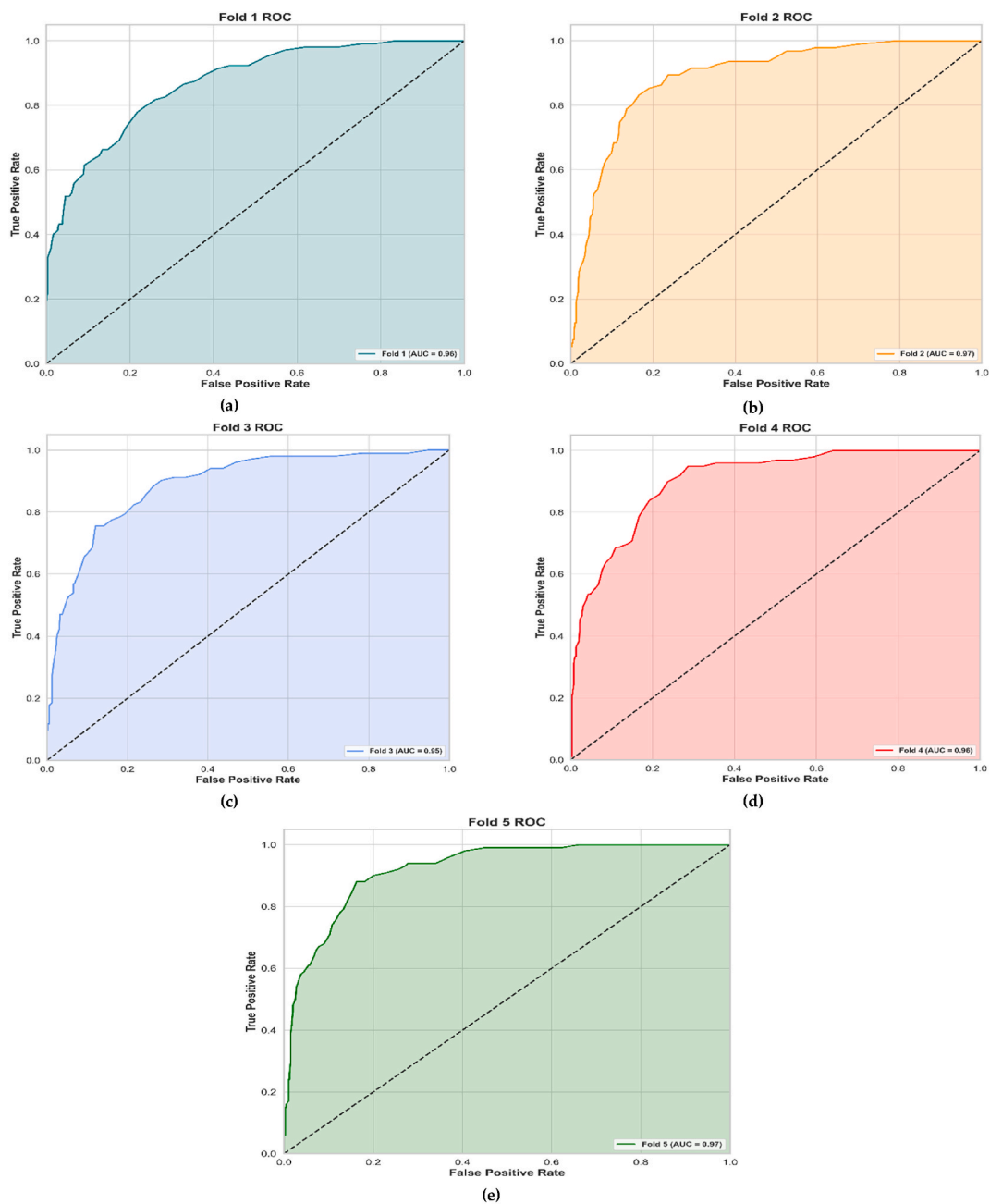Fig. 11. Training and validation loss across 5-fold (a–e) in PollenNet's cross-validation.

Fig. 12. ROC curves and AUC scores across stratified 5-fold (a–e) cross-validation for PollenNet.

Table 3
Performance metrics for proposed PollenNet across five validation folds.

Model	T_Acc	V_Acc	Precision	SPE	Recall	F1-S	FPR	FNR
Fold 1	98.68 %	96.00 %	95.50 %	95.70 %	95.60 %	95.55 %	0.05	0.04
Fold 2	98.90 %	97.00 %	96.50 %	96.60 %	96.60 %	96.55 %	0.04	0.03
Fold 3	98.00 %	95.00 %	94.10 %	94.30 %	94.20 %	94.15 %	0.06	0.05
Fold 4	98.75 %	96.50 %	95.85 %	96.00 %	95.90 %	95.88 %	0.04	0.04
Fold 5	98.99 %	97.20 %	96.70 %	96.80 %	96.75 %	96.72 %	0.03	0.03

specificity) for a specific class. Particularly, classes such as Betula, Manoao, and Metrosideros exhibit exceptionally high AUC values of 1.00, indicating perfect or near-perfect discrimination between positive and negative instances. Other classes, including Actinidia and Citrus, display slightly lower AUC values, such as 0.99, reflecting strong but slightly less than perfect classification performance.

The AUC values are remarkably consistent, with most classes having AUCs close to 1.00, showcasing the model's robustness. The lower AUC values, such as 0.98 for FuscosporaFusca and Santalum, were observed, still indicating a strong performance. The diagonal dashed line in the plot represents the performance of a random classifier, with a 50 % chance of correct classification. The distance of the PollenNet curves from this line underscores its superior performance. The ROC analysis in this figure further confirms PollenNet's ability to accurately distinguish between various pollen grains accurately, maintaining high precision and recall across a diverse set of classes. This detailed evaluation of class-specific AUC values highlights the model's effectiveness in multi-class classification tasks, ensuring reliable identification across all categories.

4.5. Stratified 5-fold cross-validation

Fig. 10 represents the Stratified 5-fold cross-validation accuracy of PollenNet. The figure illustrates the training and validation accuracy of PollenNet across five distinct cross-validation folds. Recent studies have employed the method of monitoring each fold during k-fold cross-validations by tracking training and validation accuracy and losses across epochs [43,44]. This approach is widely used in deep learning research to gain deeper insights into the model's learning dynamics and ensure robust performance evaluation across different data splits. Each subplot (a-e) corresponds to a separate fold, depicting the model's performance over 100 training epochs. The blue lines represent the training accuracy, while the orange line denotes validation accuracy for each fold. The plots collectively indicate a consistent achievement of high accuracy, with training and validation accuracies maintaining above 98 % and 95 %, respectively, across all folds. This performance pattern underscores the robustness and generalizability of PollenNet, demonstrating its capacity to maintain high predictive accuracy even with varying subsets of data.

Fig. 11 depicts loss metrics during the stratified 5-fold cross-validation of PollenNet. Subfigures (a-e) illustrate the training and validation loss as a function of epochs for each fold. The red line represents the training loss, decreasing steadily as the model learns from the data. The blue line depicting the validation loss shows an overall drop with some fluctuation, indicating that the model is encountering and adjusting to new patterns within the validation set. The consistency of this pattern across all 5-folds demonstrates
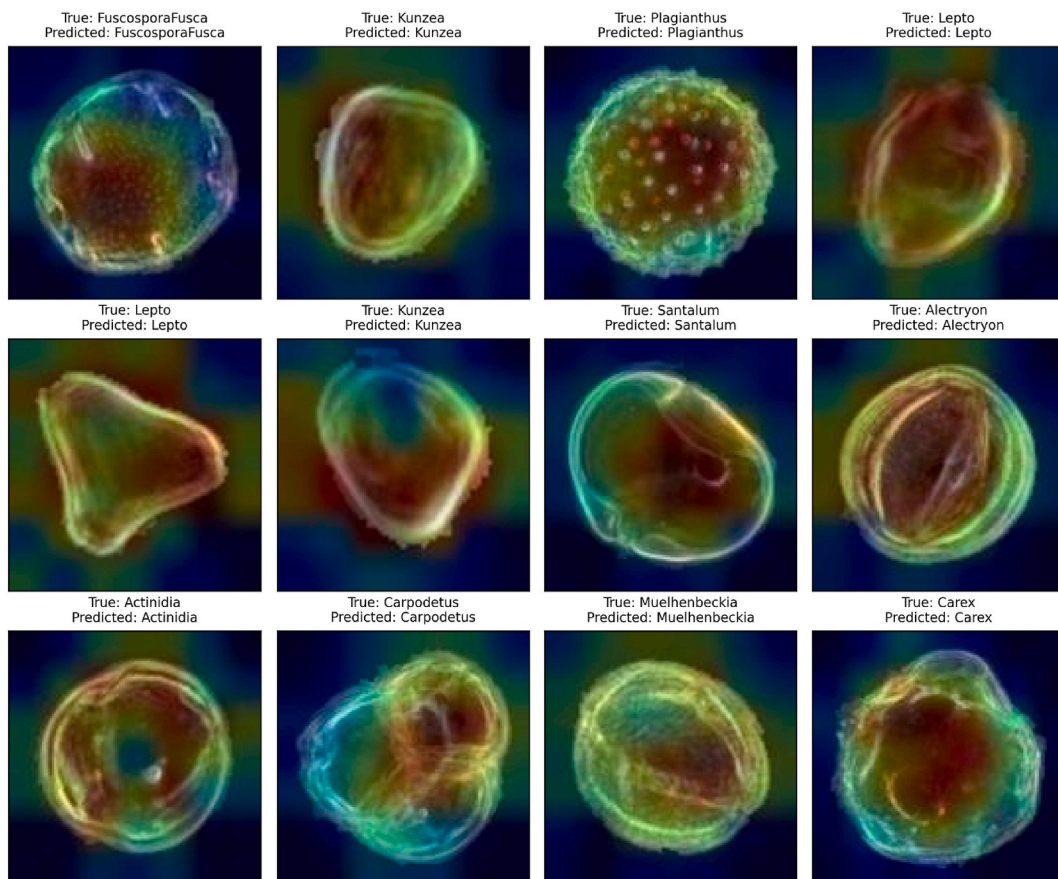

Fig. 13. Grad-CAM heatmaps demonstrate PollenNet's accurate feature detection across various correctly classified pollen grains.

PollenNet's effective learning and generalization capability across folds under the cross-validation framework.

The Receiver Operating Characteristic (ROC) curves and Area Under the Curve (AUC) scores serve as essential indicators in the attempt to gain a deep comprehension of a model's performance across different classifications. Fig. 12(a–e) exhibits the ROC curves and AUC scores for each fold in the 5-fold cross-validation of PollenNet. The ROC curves for Fold 1 through Fold 5 are displayed, with AUC values of 0.96, 0.97, 0.95, 0.96, and 0.97, respectively. Each curve, accompanied by its shaded area, indicates the model's ability to classify different classes across all folds correctly. The shaded areas under each curve visually represent the AUC, providing an intuitive understanding of the model's performance. The curves trending towards the upper left corner confirms the higher true positive rates and low false positive rates, signifying the model's reliable performance. The consistency in the AUC scores across all folds demonstrates PollenNet's robustness and effectiveness in accurately handling the diverse and complex nature of the dataset despite any class imbalances or morphological variations. These results affirm the model's generalizability and ability to maintain high predictive accuracy across varying validation sets.

Furthermore, Table 3 provides a comprehensive overview of PollenNet's performance across five validation folds, illustrating the model's consistency and robustness. Each fold demonstrates high training accuracy (T_Acc), ranging from 98.00 % to 98.99 %, indicating the model's effectiveness in learning from the training data. The validation accuracy (V_Acc) range of 95.00 %–97.20 % demonstrates the model's robust generalization capabilities on unseen data. Precision values, which measure the proportion of relevant instances correctly identified, remain consistently high between 94.10 % and 96.70 %, reflecting the model's accuracy in retrieving relevant data. Specificity (SPE), which evaluates the model's ability to identify negative results correctly, is uniformly high across all folds, underscoring the model's efficacy in minimizing false positives. Recall values ranging from 94.20 % to 96.75 % further indicate the model's strong capability to identify relevant instances. The F1-Score (F1-S), which balances precision and recall, is consistently high across all folds. Additionally, the False Positive Rates (FPR) and False Negative Rates (FNR) are low, reinforcing the model's effectiveness in minimizing errors. Overall, the results demonstrate that PollenNet performs reliably across different folds, maintaining high accuracy, precision, and low error rates, which underscores the model's robustness in classification tasks.

5. Explainable AI (XAI) for model prediction analysis

In the field of Explainable AI (XAI) [39], visualization techniques such as Gradient-weighted Class Activation (Grad-CAM), saliency maps, and misclassification analysis play a pivotal role in demonstrating the internal mechanisms of neural networks such as
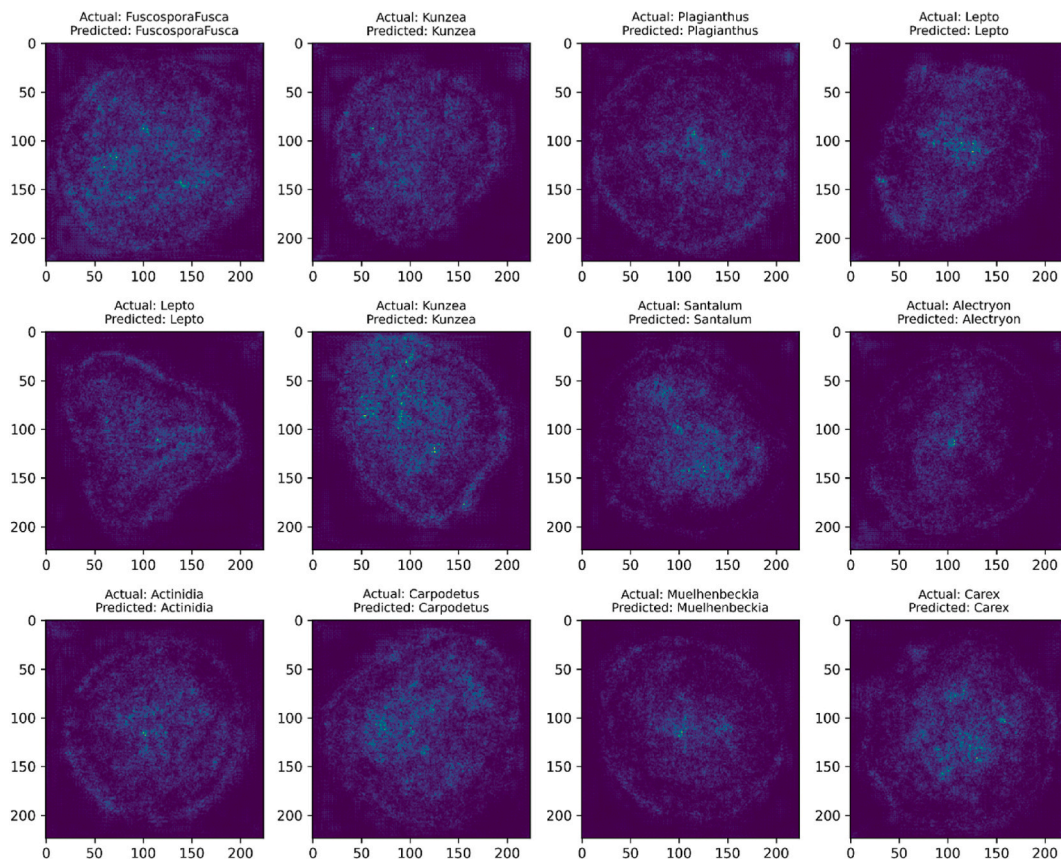

Fig. 14. Saliency maps highlight the focus area utilized by PollenNet to successfully classify different pollen grain classes.
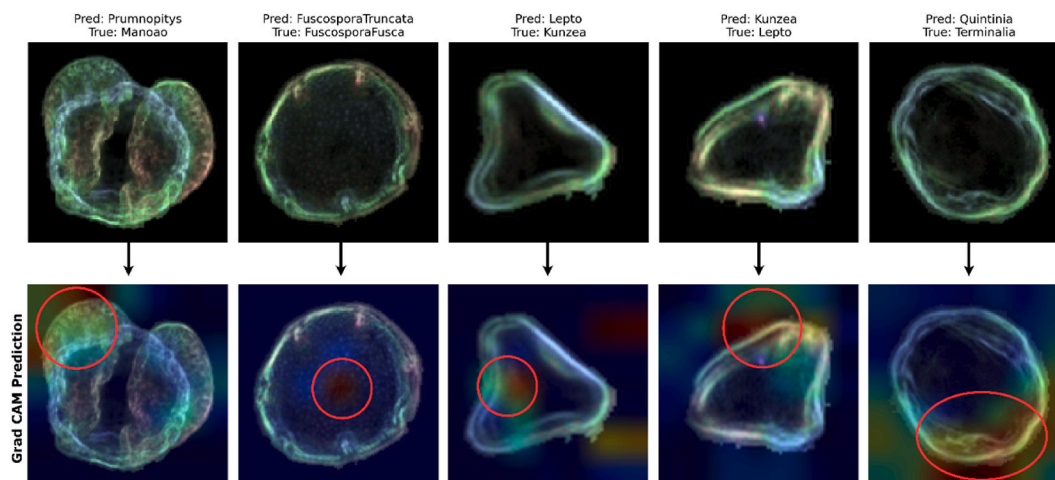
Fig. 15. Grad-CAM heat maps on specific misclassified images by PollenNet showing regions impacting its predictions.

PollenNet. Grad-CAM generates heatmaps that reveal the region's most influence on the model's predictions, while saliency maps detail the specific pixels that capture the model's attention. In addition, carefully examining misclassified images may indicate subtle anomalies in the dataset or identify areas where the model may benefit from further training. Together, these methods enhance our understanding of PollenNet's decision-making process.

5.1. Grad-CAM analysis for visualizing decision-making patterns

Fig. 13 depicts Grad-CAM [38,40] visualization of PollenNet's classifications. Displayed heatmaps for correctly classified pollen grains by PollenNet, indicating the image's most influential areas for the classification. The levels from blue to red on the heatmaps represent the model's focus from least to most influential, with red areas being key to the correct class identification. These visualizations confirm PollenNet's capability to successfully recognize and leverage distinct features for each class, such as the central pollen structure in 'FucosporaFusca' and the defining edges in 'Kunzea'. The consistency in PollenNet's attention to relevant features across various classes underlines its exceptional performance in comprehensive feature extraction for accurate pollen image classification.

5.2. Saliency map analysis for feature visualization

Fig. 14 displays the saliency maps [41] generated by PollenNet for the purpose of classifying pollen grain images. The presented saliency maps highlight the precise areas inside each image that were identified as the most significant by PollenNet during the categorization procedure. Saliency maps are a form of visual explanation that illuminates the gradient of the model's attention, where brighter regions indicate a higher contribution to the model's predictive conclusion. These particular maps correspond to various classes of pollen grains where the model's focus is clearly on identifying specific textural and structural features that are crucial for recognizing each category. The saliency maps produced by PollenNet effectively highlight important regions, demonstrating its advanced ability to discover distinctive features and accurately distinguish across intricate categories in the dataset.

6. Study limitations (interpreting model misclassifications with Grad-CAM)

Fig. 15 illustrates the use of Grad-CAM to identify PollenNet's areas of focus that resulted in misclassification. Despite achieving an impressive accuracy in training, testing and validation the PollenNet occasionally misinterprets images. Each subfigure, with a red circle, depicts an input image superimposed with a heatmap that outlines the regions deemed critical by PollenNet during the prediction phase. The warmer colors highlight segments that significantly influenced PollenNet's decision-making process. The model's significant accuracy across a diverse dataset with 46 classes and more than 17000 images indicates that the misclassification shown may be attributable to nuances within the dataset. Such nuances could intrinsically class ambiguities, subtle variations within class instances or imbalances in representative features. These visual examinations imply that, while PollenNet demonstrates exceptional performance, there remains an opportunity for further fine-tuning and dataset augmentation to address these edge cases, thereby enhancing the boundaries of its classification capabilities.

7. Conclusions

The study presents 'PollenNet' as a novel approach for categorizing images of pollen grains. It exhibits exceptional accuracy (98.45 %) and effectiveness. The instrument has substantial promise in ecological research, particularly in environmental studies and

agricultural applications, as shown by its accuracy, specificity, recall, and F1 score. The minimal error statistics provide further validation of its reliability. PollenNet's future possibilities include augmenting the dataset to encompass a broader spectrum of pollen varieties and exploring novel image processing techniques. These efforts have significant potential to enhance PollenNet's dependability and usefulness. Additional investigation might also focus on the use of PollenNet in real-world scenarios to aid with agricultural planning and allergy forecasting. Furthermore, the methodologies used in this study have the potential to be adapted for similar challenges in several scientific disciplines, hence promoting interdisciplinary progress. PollenNet is a significant advance in the classification of pollen, facilitating broader scientific inquiries and practical applications. This underscores the pivotal significance of machine learning in advancing environmental and botanical studies.

Data availability

All data generated or analyzed during this study are included in this published article. It also available in- https://doi.org/10.6084/m9.figshare.12370307.v1.

CRediT authorship contribution statement

F M Javed Mehedi Shamrat: Writing – review & editing, Writing – original draft, Visualization, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Mohd Yamani Idna Idris:** Writing – review & editing, Supervision, Project administration. **Xujuan Zhou:** Writing – review & editing, Project administration, Funding acquisition. **Majdi Khalid:** Writing – review & editing, Project administration. **Sharmin Sharmin:** Writing – original draft, Visualization, Software, Resources, Methodology, Investigation, Formal analysis, Data curation. **Zeseya Sharmin:** Writing – review & editing, Validation, Project administration. **Kawsar Ahmed:** Writing – review & editing, Validation, Supervision, Project administration, Conceptualization. **Mohammad Ali Moni:** Writing – review & editing, Validation, Supervision, Software, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] H. Halbritter, S. Ulrich, F. Grímsson, M. Weber, R. Zetter, M. Hesse, R. Buchner, M. Svojtka, A. Frosch-Radivo, Illustrated Pollen Terminology, Springer, 2018.
[2] J.S. Salonen, M. Korpela, J.W. Williams, M. Luoto, Machine-learning based reconstructions of primary and secondary climate variables from North American and European fossil pollen data, Sci. Rep. 9 (1) (2019) 15805.
[3] C.M. Barnes, A.L. Power, D.G. Barber, R.K. Tennant, R.T. Jones, G.R. Lee, J. Love, Deductive automated pollen classification in environmental samples via exploratory deep learning and imaging flow cytometry, New Phytol. 240 (3) (2023) 1305–1326.
[4] B. Balmaki, M.A. Rostami, T. Christensen, E.A. Leger, J.M. Allen, C.R. Feldman, L.A. Dyer, Modern approaches for leveraging biodiversity collections to understand change in plant-insect interactions, Frontiers in Ecology and Evolution 10 (2022) 924941.
[5] N. Khanzhina, E. Putin, Pollen recognition for allergy and asthma management using gist features, in: Digital Transformation and Global Society: First International Conference, DTGS 2016, St. Petersburg, Russia, June 22-24, 2016, Revised Selected Papers 1, Springer International Publishing, 2016, pp. 515–525.
[6] Z. Tan, J. Yang, Q. Li, F. Su, T. Yang, W. Wang, L. Min, PollenDetect: an open-source pollen viability status recognition system based on deep learning neural networks, Int. J. Mol. Sci. 23 (21) (2022) 13469.
[7] S. Dunker, E. Motivans, D. Rakosy, D. Boho, P. Maeder, T. Hornick, T.M. Knight, Pollen analysis using multispectral imaging flow cytometry and deep learning, New Phytol. 229 (1) (2021) 593–606.
[8] A. Daood, E. Ribeiro, M. Bush, Pollen grain recognition using deep learning, in: International Symposium on Visual Computing, Springer International Publishing, Cham, 2016, December, pp. 321–330.
[9] M.K. Sobol, S.A. Finkelstein, Predictive pollen-based biome modeling using machine learning, PLoS One 13 (8) (2018) e0202214.
[10] K.C. Riley, J.P. Woodard, G.M. Hwang, S.W. Punyasena, Progress towards establishing collection standards for semi-automated pollen classification in forensic geo-historical location applications, Rev. Palaeobot. Palynol. 221 (2015) 117–127.
[11] T.N. Ngo, D.J.A. Rustia, E.C. Yang, T.T. Lin, Automated monitoring and analyses of honey bee pollen foraging behavior using a deep learning-based imaging system, Comput. Electron. Agric. 187 (2021) 106239.
[12] J. Schiele, F. Rabe, M. Schmitt, M. Glaser, F. Häring, J.O. Brunner, A. Damialis, Automated classification of airborne pollen using neural networks, in: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2019, July, pp. 4474–4478.
[13] V. Sevillano, J.L. Aznarte, Improving classification of pollen grain images of the POLEN23E dataset through three different applications of deep learning convolutional neural networks, PLoS One 13 (9) (2018) e0201807.
[14] G.L. Vega, Image-based Detection and Classification of Allergenic Pollen (Doctoral Dissertation, Université de Bourgogne, 2015.
[15] M. del Pozo-Baños, J.R. Ticay-Rivas, J. Cabrera-Falcón, J. Arroyo, C.M. Travieso-González, L. Sánchez-Chavez, S.T. Pérez, J.B. Alonso, M. Ramírez-Bogantes, Image Processing for Pollen Classification, 2012.
[16] M. Sonka, V. Hlavac, R. Boyle, Image Processing, Analysis and Machine Vision, Springer, 2013.
[17] R. Shakil, B. Akter, F.J.M. Shamrat, S.R.H. Noori, A novel automated feature selection based approach to recognize cauliflower disease, Bulletin of Electrical Engineering and Informatics 12 (6) (2023) 3541–3551.
[18] F.J.M. Shamrat, S. Azam, A. Karim, K. Ahmed, F.M. Bui, F. De Boer, High-precision multiclass classification of lung disease through customized MobileNetV2 from chest X-ray images, Comput. Biol. Med. 155 (2023) 106646.
[19] A. Rebiai, B.B. Seghir, H. Hemmami, S. Zeghoud, T. Siham, I. Kouadri, H. Terea, F. Brahmia, Clustering and Discernment of bee pollen using an image analysis system, Algerian Journal of Chemical Engineering AJCE 1 (2) (2021) 41–48.
[20] G.C. Manikis, K. Marias, E. Alissandrakis, L. Perrotto, E. Savvidaki, N. Vidakis, Pollen grain classification using geometrical and textural features, in: 2019 IEEE International Conference on Imaging Systems and Techniques (IST), IEEE, 2019, December, pp. 1–6.
[21] S. Battiato, A. Ortis, F. Trenta, L. Ascari, M. Politi, C. Siniscalco, Detection and classification of pollen grain microscope images, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 980–981.

[22] J.A. Grant-Jacob, M. Praeger, R.W. Eason, B. Mills, In-flight sensing of pollen grains via laser scattering and deep learning, Engineering Research Express 3 (2) (2021) 025021.

[23] J. Li, Q. Xu, W. Cheng, L. Zhao, S. Liu, Z. Gao, H. You, Weakly supervised collaborative learning for airborne pollen segmentation and classification from SEM images, Life 13 (1) (2023) 247.

[24] H. El Azari, J.B. Renard, J. Lauthier, T. Dudok de Wit, A laboratory evaluation of the new automated pollen sensor beenose: pollen discrimination using machine learning techniques, Sensors 23 (6) (2023) 2964.

[25] S.W. Punyasena, D.S. Haselhorst, S. Kong, C.C. Fowlkes, J.E. Moreno, Automated identification of diverse Neotropical pollen samples using convolutional neural networks, Methods Ecol. Evol. 13 (9) (2022) 2049–2064.

[26] H. Menad, F. Ben-Naoum, A. Amine, Deep convolutional neural network for pollen grains classification, in: JERI, 2019, April.

[27] I.C. Romero, S. Kong, C.C. Fowlkes, C. Jaramillo, M.A. Urban, F. Oboh-Ikuenobe, S.W. Punyasena, Improving the taxonomy of fossil pollen using convolutional neural networks and superresolution microscopy, Proc. Natl. Acad. Sci. USA 117 (45) (2020) 28496–28505.

[28] G. Daunys, L. Šukienė, L. Vaitkevičius, G. Valiulis, M. Sofiev, I. Šaulienė, Comparison of computer vision models in application to pollen classification using light scattering, Aerobiologia (2022) 1–13.

[29] Katherine Holt, New Zealand pollen, figshare. Dataset. (2020), https://doi.org/10.6084/m9.figshare.12370307.v1. (Accessed 10 February 2024).

[30] S. Battiato, A. Ortis, F. Trenta, L. Ascari, M. Politi, C. Siniscalco, Pollen13k: a large scale microscope pollen grain image dataset, in: 2020 IEEE International Conference on Image Processing (ICIP), IEEE, 2020, October, pp. 2456–2460.

[31] S. Battiato, A. Ortis, F. Trenta, L. Ascari, M. Politi, C. Siniscalco, Detection and classification of pollen grain microscope images, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 980–981.

[32] C. Mondol, F.J.M. Shamrat, M.R. Hasan, S. Alam, P. Ghosh, Z. Tasnim, K. Ahmed, F.M. Bui, S.M. Ibrahim, Early prediction of chronic kidney disease: a comprehensive performance analysis of deep learning models, Algorithms 15 (9) (2022) 308.

[33] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Adv. Neural Inf. Process. Syst. 25 (2012).

[34] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.C. Chen, Mobilenetv2: inverted residuals and linear bottlenecks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 4510–4520.

[35] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556 (2014).

[36] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[37] A. Sutradhar, M. Al Rafi, P. Ghosh, F.J.M. Shamrat, M. Moniruzzaman, K. Ahmed, M.A. Moni, An intelligent thyroid diagnosis system utilising multiple ensemble and explainable algorithms with medical supported attributes, IEEE Transactions on Artificial Intelligence (2023).

[38] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba, Learning deep features for discriminative localization, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2921–2929.

[39] A.B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, F. Herrera, Explainable Artificial Intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI, Inf. Fusion 58 (2020) 82–115.

[40] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-cam: visual explanations from deep networks via gradient-based localization, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 618–626.

[41] K. Simonyan, A. Vedaldi, A. Zisserman, Deep inside convolutional networks: visualising image classification models and saliency maps, arXiv preprint arXiv: 1312.6034 (2013).

[42] S. Szeghalmy, A. Fazekas, A comparative study of the use of stratified cross-validation and distribution-balanced stratified cross-validation in imbalanced learning, Sensors 23 (4) (2023) 2333.

[43] D. Rastogi, P. Johri, V. Tiwari, A.A. Elngar, Multi-class classification of brain tumour magnetic resonance images using multi-branch network with inception block and five-fold cross validation deep learning framework, Biomed. Signal Process Control 88 (2024) 105602.

[44] P. Bashivan, I. Rish, M. Yeasin, N. Codella, Learning representations from EEG with deep recurrent-convolutional neural networks, arXiv preprint arXiv: 1511.06448 (2015).