

# AI for Mycetoma Diagnosis in Histopathological Images: The MICCAI 2024 Challenge

Hyam Omar Ali<sup>1,2,3\*</sup>, Sahar Alhesseen<sup>2</sup>, Lamis Elkhair<sup>4,3</sup>,  
Adrian Galdran<sup>5†</sup>, Ming Feng<sup>6†</sup>, Zhixiang Xiong<sup>6†</sup>,  
Zengming Lin<sup>7†</sup>, Kele Xu<sup>7†</sup>, Liang Hu<sup>6†</sup>, Benjamin Keel<sup>8†</sup>,  
Oliver Mills<sup>8†</sup>, James Battye<sup>8†</sup>, Akshay Kumar<sup>8†</sup>,  
Asra Aslam<sup>8,9†</sup>, Prasad Dutande<sup>10†</sup>, Ujjwal Baid<sup>11†</sup>,  
Bhakti Baheti<sup>11†</sup>, Suhas Gajre<sup>10†</sup>, Aravind Shrenivas Murali<sup>12†</sup>,  
Eung-Joo Lee<sup>12†</sup>, Ahmed Fahal<sup>3</sup>, Rachid Jennane<sup>1</sup>

<sup>1</sup>\*Institute Denis Poisson, Orleans University, Orleans, 45067, France.

<sup>2</sup>The Faculty of Mathematical Sciences, University of Khartoum,  
Khartoum, 11111, Sudan.

<sup>3</sup>Mycetoma Research Centre, University of Khartoum, Khartoum,  
11111, Sudan.

<sup>4</sup>The Faculty of Pharmacy, University of Khartoum, Khartoum, 11111,  
Sudan.

<sup>5</sup>Computer Vision Center, UAB, Barcelona, 08193, Spain.

<sup>6</sup>School of Computer Science and Technology, Tongji University,  
Shanghai, 201804, China.

<sup>7</sup>College of Computer Science and Technology, National University of  
Defense Technology, Changsha, 410073, China.

<sup>8</sup>University of Leeds, Leeds, LS2 9JT, United Kingdom.

<sup>9</sup>University of Sheffield, Sheffield, S10 2TN, United Kingdom.

<sup>10</sup>Shri Guru Gobind Singhji Institute of Engineering and Technology,  
Nanded, 431606, India.

<sup>11</sup>Indiana University, Bloomington, 47405, United States.

<sup>12</sup>University of Arizona, Tucson, 85721, United States.

\*Corresponding author(s). E-mail(s):

[hyam.omar-abbass-ali@univ-orleans.fr](mailto:hyam.omar-abbass-ali@univ-orleans.fr);

†These authors contributed equally to this work.

## Abstract

Mycetoma is a neglected tropical disease caused by fungi or bacteria leading to severe tissue damage and disabilities. It affects poor and rural communities and presents medical challenges and socioeconomic burdens on patients and healthcare systems in endemic regions worldwide. Mycetoma diagnosis is a major challenge in mycetoma management, particularly in low-resource settings where expert pathologists are limited. To address this challenge, this paper presents an overview of the Mycetoma MicroImage: Detect and Classify Challenge (mAIcetoma) which was organized to advance mycetoma diagnosis through AI solutions. mAIcetoma focused on developing automated models for segmenting mycetoma grains and classifying mycetoma types from histopathological images. The challenge attracted the attention of several teams worldwide to participate and five finalist teams fulfilled the challenge objectives. The teams proposed various deep learning architectures for the ultimate goal of this challenge. Mycetoma database (MyData) was provided to participants as a standardized dataset to run the proposed models. Those models were evaluated using evaluation metrics. Results showed that all the models achieved high segmentation accuracy, emphasizing the necessity of grain detection as a critical step in mycetoma diagnosis. In addition, the top-performing models show a significant performance in classifying mycetoma types.

**Keywords:** Artificial intelligence, Histopathology image analysis, Segmentation, Classification, Mycetoma diagnosis, Computer-aided diagnosis

## 1 Introduction

Mycetoma is a neglected tropical disease (NTD) that exerts a profound and far-reaching impact on individuals, families, communities, and health systems, particularly in endemic regions [1, 2]. It is a chronic and progressive disease, characterised by slow-growing lesions that can lead to severe deformities, disability, and social stigma, and compound the suffering of those affected [3, 4]. Despite its classification as an NTD, the burden of mycetoma is disproportionately high in many low-resource settings, especially in countries like Sudan [1, 5]. Although mycetoma was officially declared a health priority by the WHO in 2016 due to its devastating consequences on public health, yet, there is no genuine progress in the management of the affected patients and communities [6, 7].

Mycetoma is caused by a wide variety of microorganisms, including more than 70 causative organisms, which can be broadly classified into two categories: Actinomycetoma (AM), caused by bacterial species and Eumycetoma (EM), caused by fungi [8, 9]. The infection primarily occurs when microorganisms from the environment penetrate the skin, often through minor injuries. The chronic granulomatous inflammation caused by these organisms results in localised swelling, sinus tract formation, and the presence of grains in the discharge, which are a diagnostic hallmark of mycetoma [10, 11].

The disease's insidious onset and slow progression often lead to a delayed diagnosis and subsequent treatment, exacerbating the physical and psychological toll on patients [12]. This delay, combined with the limited access to adequate healthcare facilities in many endemic regions, creates significant barriers to timely and effective intervention [13]. Mycetoma patients frequently experience severe disability due to the destruction of soft tissues, bones, and joints, which can lead to amputation in advanced cases [14, 15]. The affected individuals often face social exclusion, economic hardship, and limited access to education and employment due to the debilitating nature of the disease [16].

The limited availability and accuracy of diagnostic tools in many endemic regions hamper effective diagnosis and management [17, 18]. In many endemic regions, the histopathological technique remains the only available diagnostic tool, as more advanced diagnostic methods, such as molecular testing or imaging techniques, may be inaccessible due to high costs or limited infrastructure [19, 20]. However, its effectiveness is highly dependent on the skill of a well-trained histopathologist, which is often unavailable in rural or resource-limited areas. Furthermore, the complexity of histopathological interpretation poses a challenge for less experienced practitioners, underscoring the need for more robust diagnostic solutions [21]. One potential advancement is the development of automated artificial intelligence (AI)-based diagnostic tools, which could assist in the interpretation of histopathological findings. AI-driven systems hold promise for improving diagnostic accuracy and accessibility, particularly in regions where trained specialists are scarce. Such tools could facilitate early diagnosis and ensure that patients receive the appropriate treatment in a timely manner, thereby reducing the burden of mycetoma on affected individuals and health systems alike [22].

In 2023, the first study that used a machine learning approach to semi-automatically analyse histopathological microscopic images of grains and provide a classification of the disease as eumycetoma or actinomycetoma was reported. This computational method achieved an accuracy rate of 91.89% in identifying the causative agents. This approach could greatly benefit rural areas with limited access to specialized clinical centres [23]. Building on this semi-automated model for the classification of mycetoma, we organized the Mycetoma MicroImage: Detect and Classify Challenge (mAICetoma) to make significant strides towards automating and enhancing the diagnostic process. The challenge shows successful outcomes and five finalist teams achieved promising results in the development of AI-driven models.

## 2 Challenge Design

mAICetoma challenge was organised with the conjunction of the Medical Image Computing and Computer-Assisted Interventions Conference (MICCAI) 2024 to promote the development of automated methods for detecting and classifying Mycetoma grains from histopathological images [24]. The ultimate goal of this challenge was to improve mycetoma diagnostic efficiency and aid in treatment decision-making.

The challenge consisted of two core tasks designed to simulate the diagnostic workflow in clinical pathology:

## Task 1: Mycetoma Grain Segmentation

Participants were required to develop algorithms to detect and segment Mycetoma grains within histopathological images. Accurate segmentation is crucial, as the presence of grains is a definitive diagnostic feature of Mycetoma. The challenge dataset provided expert-annotated grain masks, and participants' results were evaluated based on segmentation performance. Developed algorithms should output the boundary definitions of each detected Mycetoma grain within the images.

## Task 2: Mycetoma Type Classification

The second task involved classifying the detected grains into Actinomycetoma (bacterial) or Eumycetoma (fungal). Correct classification is critical for determining treatment, as bacterial infections are treated with antibiotics, while fungal infections often require antifungal therapy or surgical intervention. For each detected Mycetoma grain, the algorithm was supposed to give a classification label indicating whether the detected grain is Actinomycetoma or Eumycetoma. Fig.1 illustrates typical histopathological images of both classes and their segmentation masks.

### 2.1 Challenge statistics

The challenge successfully engaged multiple international teams from diverse backgrounds, including computer vision researchers, AI developers, and medical imaging experts. A total of 34 teams, each comprising one to five members, registered for the challenge from 16 different countries Fig.2. The teams had various degrees of experience in medical image analysis; 20.59% were beginners, 64.71% were intermediate, and 14.71% were advanced.

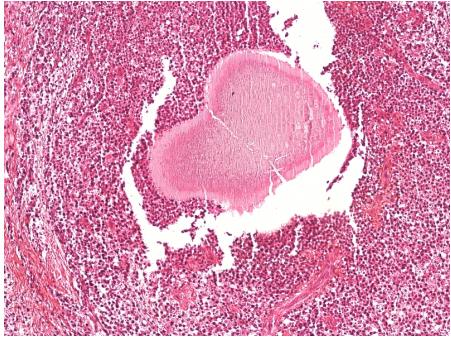
Five teams successfully completed the challenge (Table 1). Three of the teams presented their results in person on the day of the challenge, while the other two teams sent recordings of their presentations.

### 2.2 Challenge Workflow and Timeline

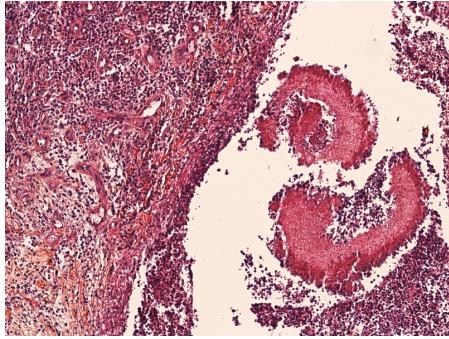
The challenge followed a well-defined structure, including multiple phases to ensure fair evaluation and participant engagement (Table 2). Participants were required to submit their results through an online submission portal, ensuring a blind evaluation where test set ground truth labels were not available to teams.

**Table 1:** Demographics of the five teams that completed the Challenge

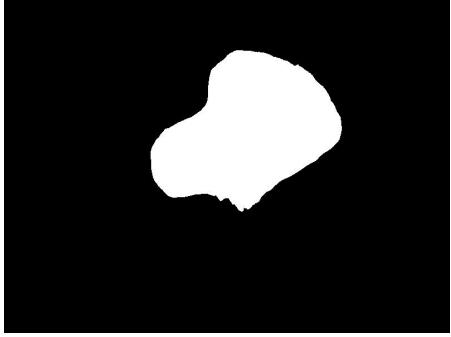
Team Name	Number of Team Members	Country	Level of Experience
VSI	2	United States	Intermediate
Minions	5	United Kingdom	Intermediate
Macaron	5	China	Intermediate
TeamTiger	3	United States	Intermediate
Adrian Galdran	1	Spain	Advanced



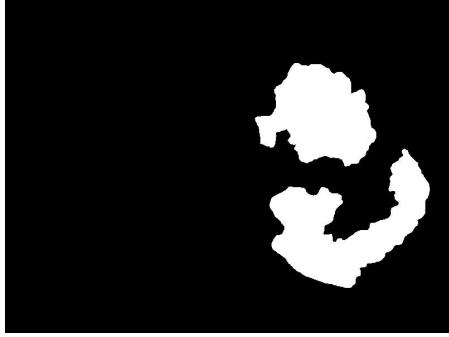
(a) Actinomycetoma (AM) image



(b) Eumycetoma (EM) image



(c) AM Segmentation Mask

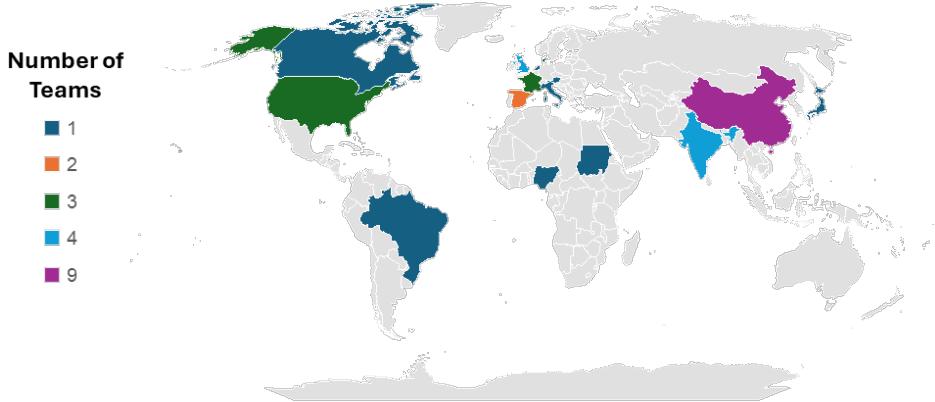


(d) EM Segmentation Mask

**Fig. 1:** Representative histopathological images of mycetoma. The first row shows examples of the two classes: Actinomycetoma (AM) and Eumycetoma (EM). The second row presents the corresponding segmentation masks highlighting the mycetoma grains within the tissue samples.

### 3 Dataset Description

A novel mycetoma histopathological images dataset designed for automated segmentation and classification was used in this challenge [25]. The clinical data (surgical biopsies) were collected from patients seen at the Mycetoma Research Centre (MRC) in Sudan during the last five years with various mycetoma classes, clinical presentations and durations. The data usage was approved by the Soba University Hospital Ethical Committee, Khartoum, Sudan. Tissue blocks were prepared from the surgical biopsies. From each tissue block, we acquire 2-3 tissue sections. These sections were stained with Haematoxylin and Eosin stain (H&E) and imaged using a Nikon Eclipse 80i digital optical microscope. Images were captured in RGB colour space with 10X magnification. On average, six images were taken for each patient. The dataset includes 864 images from 142 patients, each of which is annotated with a binary mask indicating the presence of grains. This facilitates both the detection and segmentation



**Fig. 2:** The distribution of teams that registered for the challenge around the world

**Table 2:** The different phases of the challenge

Phase	Description	Date
Challenge announcement and registration opening	Participants register and gain access to labelled training and validation data.	May 2, 2024
Submission portal opening	Participants submit predictions on the test dataset.	August 19, 2024
Submission Deadline	All final results must be submitted for evaluation.	September 6, 2024
Evaluation & Results Announcement	The challenge team evaluates the submissions and announces the results to participants.	September 23, 2024
MICCAI 2024 Presentation	The teams' rankings were revealed, and the top teams presented their approaches at the challenge session.	October 6, 2024

tasks. The manual segmentation is used as a ground-truth annotation for the detection of mycetoma grain, while the differentiation of the types defined by the expert pathologists is used to assign mycetoma class. In this dataset, the image data consists of 471 EM and 393 AM from 80 and 62 patients, respectively. This dataset was used for the training, validation, and testing with respective percentages of 65%, 70% and 20%. A careful splitting strategy was employed to prevent potential statistical bias that might arise from allocating images from the same patient into different sets. Consequently, each patient's images were exclusively assigned to either the training, validation, or test sets, ensuring that they are not shared across multiple sets. This approach aimed to maintain the integrity of patient-specific data within each split and ensure a well-balanced distribution of data for effective model training and evaluation.

## 4 Participating Teams and Methods

### 4.1 Methods

This section presents the proposed approaches of the finalist teams.

#### 1. Adrian

The first step was of data preprocessing. Having noticed the presence of several semi-duplicates in the ground-truth segmentation annotations, the team removed them by merging images containing the same visual content. Two different models were then trained for each of the tasks. For the segmentation task, the team experimented with various architectures, before settling on an ensemble of six randomly initialized encoder-decoder nets. The encoder was a Feature-Pyramid Network [26], with the decoder being a Mix Vision Transformer [27]. Following the same setup as in [28], the networks were trained by optimizing a binary cross-entropy loss function, with a batch size of 4, and a learning rate of  $10^{-4}$ , which was annealed towards zero during training following a cosine law. For the classification task, the team also experimented with several architectures but quickly realized that smaller models could already achieve a high level of accuracy. To avoid overfitting, the team opted for a modest number of trainable parameters and ultimately submitted an ensemble of three EfficientNet-B0 networks [29]. The training process was similar to that for segmentation networks but with a higher batch size of 8 samples. In addition, to optimize of the models, Adrian compared the conventional Adam algorithm with a newer optimization algorithm called AdEMAMix [30]. The latter achieved to reach marginally but consistently better performance with less overfitting, so it was adopted it for our final submission.

#### 2. VSI

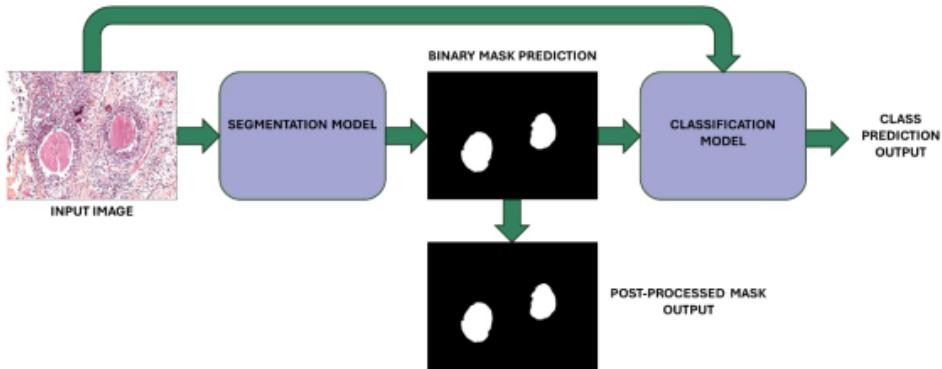
For the segmentation task, the VSI team utilized the nnUNet [31] architecture, specifically a PlainConvUNet with 8 stages and 32 to 512 feature maps per stage. This architecture was chosen due to its adaptability for biomedical image segmentation tasks, providing efficient feature extraction across multiple scales. Key strategies for enhancing performance included resizing the images to 640x896 pixels, using a batch size of 6 for 2D segmentation and employing the Dice loss function for optimization, which is particularly effective for segmentation tasks. As for post-processing, Conditional Random Fields (CRF) [32] were applied to the segmentation output to refine boundaries, significantly enhancing the results by focusing on clear delineation between BM and FM images. Specifically, the final softmax masks were converted to binary format, and CRF was applied to further refine them, improving overall segmentation consistency.

The pretrained ResNet50 model [33] was adapted to distinguish between BM and FM images. Adjustments included resizing images to 224x224 pixels and fine-tuning the model with a fully connected output layer optimized for binary classification. Sigmoid activation was applied to produce binary output probabilities, and binary cross-entropy was used as the loss function, a common choice for binary classification tasks to manage the differences in image distributions.

### 3. Minions

Minions's approach consisted of multiple stages including preprocessing, segmentation, classification, hyperparameter tuning, and postprocessing, as shown in Fig.3. The team made two submissions which used two different classifiers. The first submission's classifier included clinical variables and binary multi-task segmentation masks. The second submission's classifier was a baseline model using just the images as input.

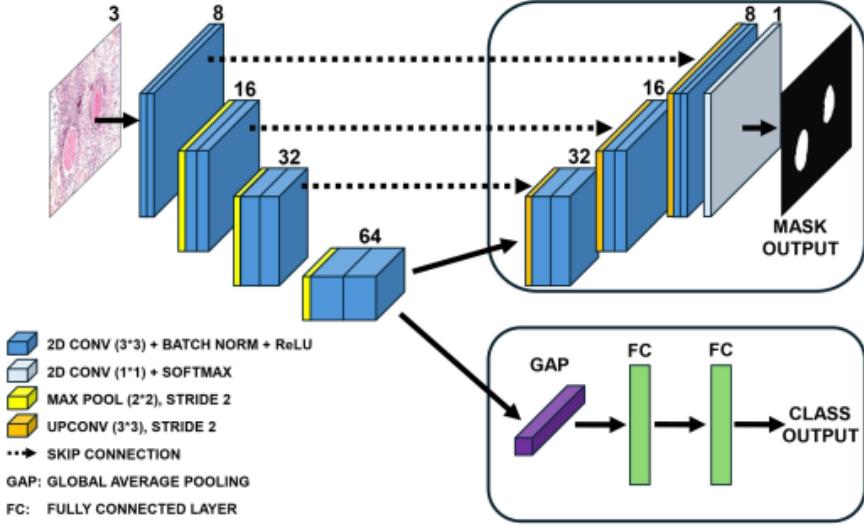
Pre-processing involved data visualization and correcting issues like mask shapes, and duplicate images. Duplicate images with different masks were merged into single ground truth masks while overlapping regions across non-identical images were identified and merged using OpenCV's ORB features and BFMatcher algorithm. Masks were updated to ensure consistent grain labelling and a small number of image/mask pairs with large unlabelled grains were excluded. This process resulted in cleaned training and validation datasets with corrected masks, extensions, and non-overlapping data.



**Fig. 3:** Minions's schematic illustrating the data processing strategy

A variant of the U-Net architecture was used for segmenting mycetoma grains, incorporating both a segmentation and a classification head-to-output segmentation masks and class predictions (Fig.4). While the class prediction was not the final classification output, the model was trained in a multi-task way using both segmentation and classification losses. This approach aimed to produce more accurate, class-specific segmentation masks. The loss was computed as  $\text{TotalLoss} = \text{SegmentationLoss} + a * \text{ClassificationLoss}$ , where the segmentation loss combined binary cross-entropy (BCE) loss and dice loss, while classification loss was based on the BCE. Hyperparameter tuning determined optimal values of batch size:8, learning rate  $2 * 10^{-4}$ , and  $a = 0.2$ . Using a threshold of 0.5, the trained model generated binary masks for training and validation data. To improve mask clarity, automated post-processing steps were implemented including morphological operations, such as opening and closing. This reduced noise and filled holes, while connected component analysis filled out irrelevant regions by retaining components

above a certain size. Lastly, the remaining holes within components were filled to enhance the interpretability of the masks for visualisation and further analysis.

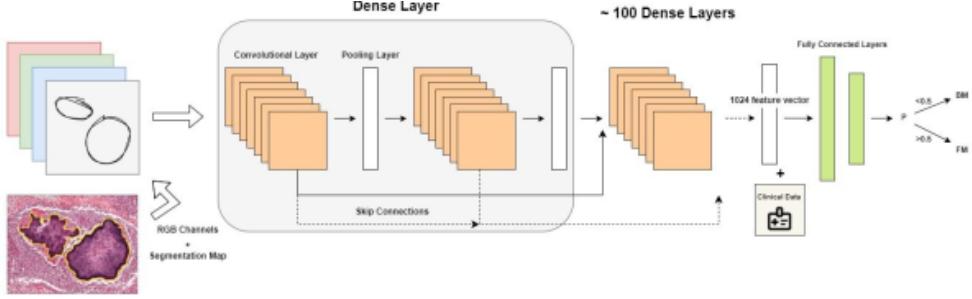


**Fig. 4:** Minions's schematic illustrating the segmentation model architecture

The classification model was built using a DenseNet-169 backbone combined with a two-layer multi-layered perceptron classifier, Fig.5. DenseNet weights were pre-trained on ImageNet and implemented via MONAI [34]. Experiment 1 utilized four input channels, including the predicted mask from the segmentation model and 27 handcrafted features that describe grain morphology. In contrast, Experiment 2 used only three channels as a baseline. Handcrafted features included colour (mean, variance, skewness), texture (GLCM and LBP), and shape descriptors (grain border irregularity). These features were concatenated with the features representation of DenseNet and processed through the classifier. The model was trained using BCE loss for 30 epochs with an Adam optimizer. Data augmentation, including flips and colour jitter, was applied to improve robustness. Bayesian optimization was used for hyperparameter tuning, focusing on learning rate, batch size, weight decay, and mask channels. Optimal hyperparameters were found to have a learning rate of  $4 * 10^{-4}$ , batch size of 12, and weight decay of  $10^{-3}$ . Hyperparameter tuning improved performance by 1–2%. Both experiments achieved similar performance, with 96.4% agreement on the testing data. The baseline model was retrained for the final experiments due to its superior validation metrics. For the testing set, the model correctly predicted most cases, with only 8 out of 167 predictions falling into the uncertain range (0.25–0.75). In clinical applications, these uncertain cases could be flagged for expert review.

#### 4. Macaroon

The Macaroon team proposed a framework comprising a segmentation model and a



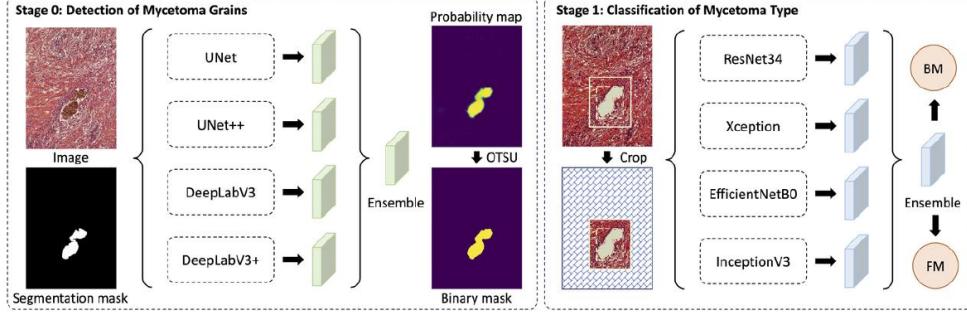
**Fig. 5:** Minions's schematic illustrating the classification model architecture

classification model designed for the tasks of detecting Mycetoma grains and classifying Mycetoma types, respectively. The segmentation model generated probability map which was then transformed into a binary image using the Otsu algorithm. Based on the segmentation results, the classification model used cropped regions of interest for both training and prediction purposes, thereby focusing more closely on the relevant areas. Both tasks used an ensemble of four models to combine the results. All models were initialized pre-trained parameters from ImageNet. Both models adhered to the official data split and utilized the Adam optimizer with a batch size of 16. An early stopping strategy was implemented to terminate the training if no improvement was observed on the validation set within 5 epochs. Data augmentation for the classification model includes: RandomCrop, HorizontalFlip, VerticalFlip, GaussianBlur, RandomRotate90, RandomBrightnessContrast, and ShiftScaleRotate. The data augmentation strategies for the segmentation model include: Resize, HorizontalFlip, VerticalFlip, RandomRotate90, and RandomResizedCrop. It is important to note that during classification, images are resized directly to  $512 \times 512$  for training and prediction. However, during segmentation, a sliding window approach with a window size of  $512 \times 512$  was used for prediction.

The model incorporated two keys training strategies: a multi-model ensemble approach and a cascaded classification and segmentation strategy. Firstly, numerous studies have shown that multi-model ensemble strategy improves accuracy, reduces overfitting, and enhances model robustness [35]. Secondly, the team believes that the discriminative regions in an image should incorporate two types of information: the region of the image that is diseased, as indicated by the provided annotations, and the surrounding environment of this area. However, it was observed that the diseased region occupies only a small portion of the image. This means that a large amount of environmental information is present in the image, which could negatively impact the model's diagnostic performance. Therefore, having obtained the segmentation results from the previous step, we performed rectangular cropping based on the diseased region in order to minimize the impact of background noise on the diagnostic process. This approach is similar to the cascade training strategy used in nnUNet [31]. The overall framework is illustrated in Fig.6.

##### 5. Tiger

The models developed by the Tiger team were designed to accurately detect



**Fig. 6:** The framework proposed by macaroon’s team

and classify mycetoma grains, which are crucial for distinguishing between Actinomycetoma and Eumycetoma. For the segmentation task, the team used a ResNet50 backbone with DeeplabV3+, a state-of-the-art architecture for semantic segmentation tasks. The model layers were structured to effectively capture spatial hierarchies and contextual information and consisted of three key components: ResNet50 Encoder, Atrous Spatial Pyramid Pooling (ASPP) for capturing multi-scale information, and Decoder with Upsampling Layers for refining spatial resolution. This model incorporates a Cross Knowledge Distillation module to enhance the segmentation performance by enabling cross-task knowledge sharing and applying a distillation technique across multiple heads in the model, leveraging the advantages of attention mechanisms to focus on specific image regions. This approach enhances the models focus on mycetoma grain features within histopathological images. The training pipeline for segmentation included the Adam optimizer, Dice Loss as the loss function, and Dice Score as the evaluation metric. The classification model is based on DenseNet121, which is an advanced convolutional neural network architecture that is renowned for its dense connectivity pattern. The dense connectivity of DenseNet121 enables the network to make efficient use of features, leading to high classification accuracy. By sharing feature maps across layers, the network captures both local and global features, making it particularly effective for distinguishing subtle differences between the two types of mycetoma. The model was used with Adam as the optimizer and Cross-Entropy as the loss function. metrics of Accuracy and AUC were used to assess performance. Key insights from these architectures suggest that the integration of multi-head attention enhances the ability to capture fine-grained differences between Actinomycetoma and Eumycetoma.

## 5 Evaluation method

The evaluation framework was designed to ensure a comprehensive assessment of participants’ algorithms. The primary evaluation criteria included sensitivity, specificity,

accuracy, and the sensitivity, specificity, accuracy, and Matthews Correlation Coefficient (MCC) [36] for the classification task, and sensitivity, specificity, accuracy, and Dice coefficient [37] for the detection task.

Sensitivity and Specificity were chosen to minimize false negatives and false positives, ensuring a reliable diagnosis and classification of Mycetoma types. Accuracy provided an overall measure of correctness, while MCC accounts for imbalanced class distributions, offering a robust measure of classification performance. For detection, the Dice Coefficient was used to assess the spatial overlap in the detection of Mycetoma grains within histopathological images.

To maintain fairness in the evaluation process, submissions with missing results were assigned default scores for each metric. Missing Sensitivity, Specificity, and Accuracy values were set to 0.5, representing an average performance, while a missing MCC value was assigned 0, indicating no correlation. Handling missing results ensures that the incomplete submissions are not given an unfair advantage or disadvantage in the ranking, while still allowing them to be considered.

## 5.1 Ranking Scheme

A weighted sum approach was employed to rank the algorithms for both classification and detection tasks, emphasizing the relative importance of each metric. Combining multiple evaluation metrics ensured a comprehensive assessment of algorithm performance. Assigning different weights to each metric enables for flexibility in reflecting their relative importance in relation to the objectives of the challenge.

Using a weighted sum enables for ranking based on the importance of each metric. While different participants may prioritize metrics differently, this method ensured a balanced evaluation that aligned with the challenge's goals while allowing flexibility in addressing the specific priorities of each task. For instance, MCC was given the highest weight in classification due to its robustness in handling class imbalance. Similarly, the Dice coefficient was given a high weight in the detection process, as it best captures the overlap in segmentation tasks.

The weights assigned for classification were as follows: 15% for Sensitivity and Specificity, 30% for Accuracy and the highest weight of 40% for MCC. For detection, the assigned weights were 20% for Sensitivity and Specificity and 25% for Accuracy and the Dice coefficient. The classification weighted sum formula is:

$$\text{WeightedSum} = w_{\text{sen}} \text{sensitivity} + w_{\text{spec}} \text{specificity} + w_{\text{accu}} \text{accuracy} + w_{\text{mcc}} \text{MCC}$$

For detection, the formula is:

$$\text{WeightedSum} = w_{\text{sen}} \text{sensitivity} + w_{\text{spec}} \text{specificity} + w_{\text{accu}} \text{accuracy} + w_{\text{dice}} \text{Dice}$$

where  $w_{\text{sen}}$ ,  $w_{\text{spec}}$ ,  $w_{\text{accu}}$  and  $w_{\text{dice}}$  are the weights for sensitivity, specificity, accuracy, MCC and Dice, respectively. Finally, the ranking was determined by ordering the weighted sum in descending order, with the highest-scoring submission ranked as the top performer. This ranking method provides a straightforward and interpretable approach to evaluating performance. Furthermore, the weighted sum approach allowed

flexibility in addressing specific task priorities while reflecting the overall quality of each submission.

## 6 Results

The evaluation was conducted separately for the segmentation and classification tasks. The performance of each team was assessed using the metrics above, and the ranking was determined through a weighted sum approach. The segmentation task focused on detecting mycetoma grains within histopathological images. Table 3 presents the results achieved by the finalist teams. Adrian achieved the highest weighted sum score (93.56%), demonstrating superior performance for accuracy and Dice coefficient. Macaroon and Tiger followed closely, with scores of 92.52% and 91.98%, respectively. Despite variations in the teams model performance, all the teams demonstrated strong performance in segmenting mycetoma grains.

**Table 3:** Segmentation

Team	Sensitivity	Specificity	Accuracy	Dice	Weighted Sum
Adrian	0.9914	0.8885	0.9806	0.8820	93.56%
Tiger	0.9956	0.8406	0.9786	0.8646	91.98%
Macaroon	0.9740	0.9124	0.9706	0.8439	92.52%
VSI	0.9834	0.8696	0.9723	0.8295	91.37%
Minions	0.9863	0.9863	0.9785	0.8362	91.42%

For the classification task, models were evaluated based on their ability to differentiate between Eumycetoma and Actinomycetoma. Table 4 summarizes the classification results. Adrian achieved the highest classification performance, with a weighted sum of 96.14%, followed by Macaroon (93.11%) and Minions (92.06%). Notably, MCC played a crucial role in ranking as it effectively accounted for imbalanced class distributions.

**Table 4:** Classification

Team	Sensitivity	Specificity	Accuracy	MCC	Weighted Sum
Adrian	0.9659	0.9828	0.9726	0.9435	96.14%
Tiger	0.8068	0.9655	0.8699	0.7559	82.91%
Macaroon	0.9545	0.9483	0.9521	0.9003	93.11%
VSI	0.9091	0.9310	0.9178	0.8317	88.40%
Minions	0.9545	0.9310	0.9452	0.8856	92.06%

Since all models achieved high performance in the segmentation task, this strengthens the overall results and aligns with the challenge design. Accurate detection and segmentation of mycetoma grains is crucial for diagnosis, and the consistent high performance of the segmentation models developed by all teams reinforces their ability.

This further supports the challenge approach of separating segmentation and classification, ensuring that even if classification errors occur, the presence of correctly segmented grains allows for expert intervention and validation.

Considering the combined results of both tasks, the top three performing models overall were: Adrian, Macaroon, and Minions/Tiger. These models adopted different approaches and consistently achieved high scores across both tasks. Adrian excelled with the use of an ensemble of transformer-based architecture for segmentation and EfficientNet-B0 for classification. On the other hand, Macaroon demonstrated the power of cascaded segmentation-classification strategies. Minions and Tiger, who shared third place, used different segmentation and classification strategies. Minions employed a U-Net-based segmentation model with a dual-head architecture, while Tiger used a ResNet50 backbone with DeeplabV3+ and a Cross Knowledge Distillation module. For classification, Minions combined a DenseNet-169 model with handcrafted features, whereas Tiger utilized DenseNet121 with multi-head attention. The diversity of the model and its performance emphasized the importance of design choices for models and preprocessing strategies in optimizing AI-driven histopathological analysis for diagnosing mycetoma.

## 7 Discussion

### 7.1 Key Lessons Learned

Macaroon reported that their proposed framework is divided into two stages rather than an end-to-end approach. This leads to a problem where classification and segmentation labels cannot be considered simultaneously. Generally, jointly considering two related labels may yield better performance. Future work could explore incorporating both labels into a single model, enabling an end-to-end model that can perform both classification and segmentation. In addition, the lack of medical background and being unfamiliar with the differences among the categories involved, as well as the distinctions between diseased areas and normal tissue. This information is crucial for effective model design. For example, when diagnosing flow cytometry data, doctors can make an accurate diagnosis based on a few key markers. Without this background knowledge, we can only input all markers into the model for evaluation, which would likely introduce excessive redundant information. Finally, the models used are all based on convolutional neural networks, and their performance on natural images may be lower than the most advanced classification models, such as the Vision Transformer [38].

Several challenges encountered during the development of the models were reported by Tiger. Managing the variability in histopathological image quality and ensuring model robustness across different patient samples. They found that incorporating Cross cross-knowledge distillation and attention mechanisms was essential for improving model precision, particularly for nuanced distinctions between mycetoma types. Additionally, the importance of a well-balanced training pipeline was revealed, enabling them to achieve consistent performance in both segmentation and classification tasks.

Team Minions emphasized the importance of thoroughly examining data and understanding its characteristics before model development. During the challenge

workshop, they were surprised to learn that only one other team, Team Adrian (the winner) had visualized the images and masks in order to identify and correct issues with grain labelling. By addressing these labelling inconsistencies, their approach achieved better performance compared to other teams that relied on larger models, more complex architectures, and ensemble methods. They noted that even the best architecture can only perform well if the data is properly pre-processed. The challenge also highlighted the success of deep learning in image classification. In one of their submissions, Team Minions experimented with creating clinical features and incorporating segmentation masks to assist with classification. However, they observed no significant performance improvement compared to using a DenseNet on the RGB image alone without additional features. Furthermore, they underscored the importance of extensive hyperparameter tuning to enhance model performance and identify suitable parameter ranges. They also observed that peaks in midrange RGB signal intensities were associated with images that were commonly misclassified, pointing to inconsistencies in tissue staining. To mitigate the impact of such inconsistencies, they recommended applying preprocessing and augmentation techniques, such as colour jitter and histogram scaling. Future work might include evaluating performance on an external validation cohort to provide a better indication of model generalizability.

Team Adrian highlighted the importance of visually inspecting the data, emphasizing that visualizing the ground truth for segmentation and cleaning the training data were crucial aspects of their approach. They stressed that for the classification task, it is essential to experiment with smaller models and consider using them if they achieve accuracy comparable to larger models, as smaller models are less prone to overfitting.

Throughout the challenge, Team VSI gained insights while facing several challenges. They emphasized the significant impact of data preprocessing, noting that standardizing image sizes and normalizing intensity values were crucial for model performance. Resizing images to different dimensions for segmentation and classification (640x896 and 224x224, respectively) ensured that each model had input images tailored to its specific architecture. The adaptability of the models was another key consideration. While utilizing a self-configuring model such as nnUNet simplified the segmentation process, it also posed challenges in terms of balancing model complexity and computational cost. Fine-tuning the segmentation model while maintaining efficiency required iterative experimentation. Additionally, boundary refinement played a vital role in improving segmentation accuracy. Notably, the application of the CRF improved boundary precision. However, they acknowledged that implementing CRF in real-time diagnostic settings might be computationally intensive, highlighting the need for further research into lighter post-processing techniques or adaptive CRF implementations. Lastly, handling imbalanced data was addressed through the use of the MCC, which provided a robust metric by balancing performance across specificity, sensitivity, and accuracy. MCC was particularly valuable as it considers all four confusion matrix components (true positive, true negative, false positive, and false negative) offering a balanced evaluation of the model's performance across both classes.

## 7.2 General Discussion

The Mycetoma MicroImage Challenge aimed to promote the development of AI and computer vision solutions that can automatically detect mycetoma infection in histopathological images, as well as classify the disease. It explored the feasibility of using AI to effectively diagnose mycetoma from histopathological images of mycetoma grains. Therefore, the challenge focused on developing fully automated models for the detection and classification of mycetoma, which could significantly improve diagnostic accuracy and patient management, particularly in low-resource settings.

Despite the success of the challenge, some challenges and limitations of the dataset were identified and highlighted by the participants. The dataset of the challenge does not include additional clinical and demographic data that could enhance model performance and provide deeper clinical insights. The variability in staining techniques was reported as another challenge, which if not addressed properly, can affect the models' generalizations. Moreover, the use of single-expert annotations can limit the robustness of data by lacking inter-observer variability which could impact annotation quality and potentially diagnostic accuracy. Future efforts should focus on enhancing the dataset and assessing inter- and intra-rater variability to further strengthen the annotations. Additionally, expanding the dataset to include more diverse samples from different centres around the world, as well as cases of other diseases with similar differential diagnoses, will enable the models to effectively diagnose mycetoma alongside these diseases.

The evaluation of submitted models utilized standard performance metrics, including sensitivity, specificity, accuracy, and Matthews Correlation Coefficient for classification tasks, while segmentation performance was assessed using the Dice coefficient, specificity, sensitivity and accuracy. Although these metrics provided a comprehensive assessment of model performance, some improvements could be considered in future iterations of this challenge. The inclusion of other evaluation metrics that could offer deeper insights into model effectiveness, for example investigating the F1-score for the classification task, with the imbalanced classes, can maintain balanced sensitivity and specificity. Another consideration is the analysis of inter- and intra-rater variability in ground truth annotations to further refine the evaluation process. Additionally, pre-processing techniques such as colour normalization and data augmentation were found to play a crucial role in improving model robustness and should be considered in future model development.

As the first challenge for mycetoma diagnosis, mAIcetoma provides a solid foundation for future challenges that address the aforementioned limitations to further enhance its utility and comprehensiveness. Validating the developed models with multi-centre data and independent testing could offer valuable insights and recommendations for improvement. Future steps may involve the clinical integration of these models as prototypes or testing modules. Developing offline applications, such as desktop and mobile versions, represents a crucial step toward integrating the AI models into clinical workflows. Another promising avenue worth exploring is the application of advanced AI techniques, such as self-supervised learning and explainable AI, which can enhance model interpretability and facilitate adoption by healthcare professionals.

To ensure the successful implementation and sustained impact of the challenge outcomes, collaboration among researchers, healthcare providers, and policymakers must be fostered, alongside efforts to encourage the adoption of these AI solutions.

## 8 Conclusion

This paper overviewed the findings of the Mycetoma MicroImage Challenge held in conjunction with the 27th International Conference on Medical Image Computing and Computer-Assisted Interventions (MICCAI 2024). The challenge aimed to demonstrate the potential of AI-driven models in advancing mycetoma diagnosis. Five finalist teams proposed efficient models with segmentation Dice scores exceeding and classification accuracies ranging from 82.91% to over 96.14%. The findings of the challenge were highly encouraging, confirming that such models can enhance diagnostic capabilities, particularly in settings with limited resources. Future efforts will focus on integrating the challenge models into clinical workflows for real-world use. In order to enable more researchers to develop new models further, the challenge dataset is publicly available.

## Declarations

### Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

### Competing Interests

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.

### Data Availability

MyData histopathology dataset used in this study is publicly hosted in the AFRICAI repository and Zenodo. Access to the dataset requires submitting a formal request through the hosting platforms. Full details on licensing, access procedures, and FAIR compliance are described in [25].

### Code Availability

Not applicable.

### Ethics Approval

Not applicable.

### Consent to Participate

Not applicable.

## References

- [1] Fahal AH, Ahmed KO, Saeed AA, Elkhawad AO, Bakhet SM.: Why the mycetoma patients are still neglected. Public Library of Science San Francisco, CA USA.
- [2] Zijlstra EE, Van De Sande WW, Welsh O, Mahgoub ES, Goodfellow M, Fahal AH. Mycetoma: a unique neglected tropical disease. *The Lancet Infectious Diseases*. 2016;16(1):100–112.
- [3] Fahal AH, Hassan MA. Mycetoma. *British Journal of Surgery*. 1992 11;79(11):1138–1141. <https://doi.org/10.1002/bjs.1800791107>.
- [4] Fahal A. Mycetoma: a thorn in the flesh. *Transactions of the Royal Society of Tropical Medicine and Hygiene*. 2004;98(1):3–11.
- [5] Alhaj AAM, Ahmed ES, Hassan A, Fahal AH. Epidemiological observations and management challenges in extrapedal mycetoma: A three-decade review of 420 cases. *PLoS Neglected Tropical Diseases*. 2024;18(5):e0011841.
- [6] World Health Organization.: Addressing the burden of mycetoma. Accessed: Feb. 06, 2025. <https://iris.who.int/handle/10665/252801>.
- [7] Hay RJ, Fahal AH.: Mycetoma: an old and still neglected tropical disease. Oxford University Press.
- [8] Guerrant R, Walker D, Weller P, Fahal A. Mycetoma. In: *Tropical Infectious Diseases: Principles, Pathogens and Practice*. 3rd ed. Elsevier Health Science; 2011.
- [9] van de Sande WW, Fahal AH. An updated list of eumycetoma causative agents and their differences in grain formation and treatment response. *Clinical Microbiology Reviews*. 2024;37(2):e00034–23.
- [10] Fahal AH, Suliman SH, Hay R. Mycetoma: the spectrum of clinical presentation. *Tropical medicine and infectious disease*. 2018;3(3):97.
- [11] Williams N, O'Connell PR. *Bailey & Love's short practice of surgery* 26E. CRC press; 2013.
- [12] Ahmed AO, van Leeuwen W, Fahal A, van de Sande W, Verbrugh H, van Belkum A. Mycetoma caused by *Madurella mycetomatis*: a neglected infectious burden. *The Lancet infectious diseases*. 2004;4(9):566–574.
- [13] Hay R, Asiedu KB, Fahal A.: Mycetoma—a long journey out of the shadows. Oxford University Press.

- [14] Abbas M, Scolding PS, Yosif AA, El Rahman RF, El-Amin MO, Elbashir MK, et al. The disabling consequences of Mycetoma. PLoS neglected tropical diseases. 2018;12(12):e0007019.
- [15] Omer RF, Seif EL Din N, Abdel Rahim FA, Fahal AH. Hand mycetoma: The Mycetoma Research Centre experience and literature review. PLoS Neglected Tropical Diseases. 2016;10(8):e0004886.
- [16] Fahal AH, Ahmed ES, Mahmoud AH, Saaed AA. The Mycetoma Research Center, University of Khartoum, Sudan's experience in community engagement initiatives spans 3 decades. PLOS Neglected Tropical Diseases. 2024;18(8):e0012304.
- [17] Ahmed AA, van de Sande W, Fahal AH. Mycetoma laboratory diagnosis. PLoS neglected tropical diseases. 2017;11(8):e0005638.
- [18] van de Sande WW, Fahal AH, Goodfellow M, Mahgoub ES, Welsh O, Zijlstra EE. Merits and pitfalls of currently used diagnostic tools in mycetoma. PLoS neglected tropical diseases. 2014;8(7):e2918.
- [19] Ahmed AO, Mukhtar MM, Kools-Sijmons M, Fahal AH, de Hoog S, van den Ende BG, et al. Development of a species-specific PCR-restriction fragment length polymorphism analysis procedure for identification of *Madurella* mycetomatis. Journal of clinical microbiology. 1999;37(10):3175–3178.
- [20] Bahar ME, Bakheet OEH, Fahal AH. Mycetoma imaging: the best practice. Transactions of The Royal Society of Tropical Medicine and Hygiene. 2021;115(4):387–396.
- [21] Yousif B, Fahal A, Shakir M. A new technique for the diagnosis of mycetoma using fixed blocks of aspirated material. Transactions of the Royal Society of Tropical Medicine and Hygiene. 2010;104(1):6–9.
- [22] Ali HO, Elkheir LYM, Fahal AH. The use of artificial intelligence to improve mycetoma management. PLOS Neglected Tropical Diseases. 2024;18(2):e0011914.
- [23] Omar Ali H, Abraham R, Desoubeaux G, Fahal AH, Tauber C. Evaluation of a computational model for mycetoma-causative agents identification. Transactions of The Royal Society of Tropical Medicine and Hygiene. 2024;118(4):253–263.
- [24] Mycetoma Research Centre.: Mycetoma MicroImage: Detect and Classify Challenge (mAICetoma). Accessed: Mar. 16, 2025. [https://mycetoma.edu.sd/?page\\_id=5110](https://mycetoma.edu.sd/?page_id=5110).
- [25] Ali HO, Abraham R, Desoubeaux G, Fahal A, Tauber C. MyData: A Comprehensive Database of Mycetoma Tissue Microscopic Images for Histopathological Analysis. arXiv preprint arXiv:241012833. 2024;.

- [26] Lin TY, Dollár P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 2117–2125.
- [27] Chen JN, Sun S, He J, Torr PH, Yuille A, Bai S. Transmix: Attend to mix for vision transformers. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2022. p. 12135–12144.
- [28] Galdran A, Carneiro G, Ballester MAG. On the optimal combination of cross-entropy and soft dice losses for lesion segmentation with out-of-distribution robustness. In: Diabetic Foot Ulcers Grand Challenge. Springer; 2022. p. 40–51.
- [29] Tan M, Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In: International conference on machine learning. PMLR; 2019. p. 6105–6114.
- [30] Pagliardini M, Ablin P, Grangier D. The ademamix optimizer: Better, faster, older. arXiv preprint arXiv:240903137. 2024;.
- [31] Isensee F, Jaeger PF, Kohl SA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods. 2021;18(2):203–211.
- [32] Sutton C, McCallum A, et al. An introduction to conditional random fields. Foundations and Trends® in Machine Learning. 2012;4(4):267–373.
- [33] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 770–778.
- [34] Cardoso MJ, Li W, Brown R, Ma N, Kerfoot E, Wang Y, et al. Monai: An open-source framework for deep learning in healthcare. arXiv preprint arXiv:221102701. 2022;.
- [35] Ganaie MA, Hu M, Malik AK, Tanveer M, Suganthan PN. Ensemble deep learning: A review. Engineering Applications of Artificial Intelligence. 2022;115:105151.
- [36] Boughorbel S, Jarray F, El-Anbari M. Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric. PloS one. 2017;12(6):e0177678.
- [37] Setiawan AW. Image segmentation metrics in skin lesion: accuracy, sensitivity, specificity, dice coefficient, Jaccard index, and Matthews correlation coefficient. In: 2020 International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM). IEEE; 2020. p. 97–102.

- [38] Dosovitskiy A. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:201011929. 2020;.