International Journal of Robotics and Control Systems
Vol. 5, No. 6, 2025, pp. 2729-2749
ISSN 2775-2658
http://pubs2.ascee.org/index.php/ijrcs

Modified CBAM-VGG16 with Sequential SAM-CAM and Image Filtering for Automated Microscopic Fungi Classification

Muhammad Zulfikar Fauzi [a,1], Nanik Suciati [a,2,*]

[a] Department of Informatics, Institut Teknologi Sepuluh Nopember, Sukolilo, Surabaya 60111, Indonesia
[1] mzulfikarfauzi08@gmail.com; [2] nanik@its.ac.id
* Corresponding Author

ARTICLE INFO

Article history
Received July 26, 2025
Revised September 23, 2025
Accepted November 26, 2025

Keywords
Microscopic Fungi;
CBAM;
Attention Mechanism;
Image Filtering;
Data Augmentation

ABSTRACT

Fungal infections can lead to severe tissue invasion and multi-organ dysfunction, making the rapid and accurate identification of these infections critical. Traditional methods rely on culture-based and morphological examination, requiring extensive expertise. This study develops a modified attention-enhanced deep learning approach for automated microscopic fungi classification, utilizing VGG16 integrated with a decoupled Convolutional Block Attention Module (CBAM) that separates spatial and channel attention (SAM-CAM) to preserve spatial information. Preprocessing included data augmentation (rotation and flipping) to address class imbalance, as well as multiple standard image filtering techniques to enhance structural features. Three architectures—baseline VGG16, original CBAM-VGGNet, and modified CBAM-VGGNet—were systematically evaluated using transfer learning across five filtered datasets. The proposed model achieved a maximum accuracy of 94.33% with the HFE filter, outperforming the original CBAM-VGGNet (86.86%) and baseline VGG16 (78.88%). Attention map visualization and quantitative metrics (entropy and variance) confirmed that the decoupled SAM preserves spatial information, enabling more effective feature extraction. Class-wise recall improved substantially, particularly for challenging classes (from 0.418 to 0.878). These results suggest that carefully designed attention architectures, combined with appropriate preprocessing, enhance classification performance for complex microscopic specimens.

1. Introduction

Fungal infections caused by pathogenic fungi represent a significant global health concern affecting people worldwide [1]. These infections can progress rapidly, with invasive species causing tissue damage, vascular invasion, and even multi-organ failure if left untreated [2]-[4]. The identification of pathogenic fungi, such as mucormycosis, aspergillosis, and systemic candidiasis, remains dependent on swift and accurate diagnosis, as delays contribute to increased mortality and suboptimal therapeutic interventions [5]-[7]. Additionally, dermatophyte infections present diagnostic challenges due to morphological similarities between species and the need for prolonged culture

periods that can delay appropriate antifungal treatment [8], [9]. These phenomena highlight the opportunity for automated classification systems to mitigate the limitations of conventional morphological identification techniques. Traditionally, fungal identification has relied on culture-based methods and morphological examination under light microscopy, often supported by staining techniques and biochemical assays [10]. These approaches are inherently slow, resource-intensive, and dependent on specialized expertise. While these methods have long served as the standard in clinical and research laboratories, recent evaluations have revealed considerable variability and inconsistency in their implementation. Evaluations show that essential procedures such as direct microscopic examinations are frequently omitted, despite being crucial for diagnosis and treatment guidance. Similarly, inappropriate specimen handling, suboptimal stain selection, and inconsistent incubation durations have been observed, leading to reduced diagnostic accuracy [11]-[13]. The requirement for mycological expertise that is increasingly scarce in many clinical settings further compounds these challenges. These systematic limitations have highlighted the need for efficient and automated identification approaches.

The development of machine learning and deep learning has opened new possibilities for automated medical image analysis. Convolutional neural networks (CNNs) have demonstrated considerable performance in computer vision tasks and have been increasingly applied to medical imaging [14], [15]. These technologies have shown potential in identifying and classifying fungal species and the infections they cause [16]-[18]. Moreover, deep learning models have shown effectiveness in identifying specific fungal structures, such as hyphae in microscopic images, allowing clinicians to concentrate their attention on relevant findings and substantially reducing the time required for manual microscopic observation [19]. These successful implementations show the versatility and scalability of deep learning approaches in mycological diagnostics, establishing a strong foundation for more sophisticated automated classification systems. One of the early studies in fungal diagnostics utilized a combination of bag-of-words algorithms with support vector machine classifiers for yeast classification [17]. Another study conducted using ResNet-50 model and the transfer learning method to detect filamentous fungi and yeasts primarily detected in human nail, hair, and skin [20]. A study on onychomycosis detection utilizing deep learning models such as VGG16 and InceptionV3 was also implemented using grayscale images [21]. Another application of deep learning in the study of microscopic fungi utilized VGG16 to classify Cryptococcus neoformans yeast, indicating whether an infection is positive or negative in humans [22]. Most of these existing studies have only focused on single fungal types rather than comprehensive multi-species classification systems. This limitation restricts their applicability in clinical settings where accurate differentiation between multiple fungal species is crucial for selecting appropriate treatment and managing patients, particularly when various species of fungi appear morphologically identical under microscopic examination. Though scarce, several studies have been leaning toward a multi-species classification system.

Previous study [23] trained seven well-known CNN architectures to classify 89 distinct fungal genera using microscopic images. Their dataset comprised 1,079 high-resolution images captured at standard 400X magnification. The research methodology involved data splitting with a 7:1:2 distribution for training, validation, and testing, complemented by 10 image augmentation transformation techniques. The lack of data used in this study results in very few images available for validation and testing for some of the classes. DenseNet achieved the best accuracy overall, yielding primary prediction accuracy of approximately 65.35% and tertiary prediction accuracy reaching 75.19%. The study revealed that the implementation of data augmentation contributed to an approximate 10% gain in accuracy, while leveraging pre-trained ImageNet parameters provided an additional 15% gain across the board. When focusing on genera with adequate sample representation and excluding under-represented categories, classification performance surpassed 80%. Cross-validation was applied to the best-performing model and confirmed the model's consistency, with performance metrics of 69.08%, 70.57%, and 69.32% across the three folds.

Another study introduced MeFunX [24], an innovative meta-learning framework designed for early detection of fungal infections using microscopic images. This architecture employed a novel

stacked ensemble approach, combining two distinct convolutional base learners (CBL1 and CBL2) with XGBoost serving as the meta-learner for final classification decisions. The first base learner (CBL1) utilized a custom-built architecture incorporating convolutional blocks, max pooling, batch normalization, and fully connected layers, while the second base learner (CBL2) leveraged ConvNeXt architecture for feature extraction with global average pooling. The researchers evaluated their approach using the DeFungi dataset [25] which contains 9,114 images representing five different superficial fungal infections, partitioned into a 8:2 ratio for the training and testing sets. MeFunX demonstrated exceptional performance, achieving an overall accuracy of 92.49%. The results highlight the effectiveness of meta-learning strategies in fungal classification tasks.

Previous studies have demonstrated the promise of deep learning for fungal classification when supported by appropriate architectures. Attention mechanisms, which allow neural networks to focus on the most informative regions of the input, have shown potential in microscopy-based analysis. For example, soft attention mechanisms integrated into CNNs improved accuracy in classifying fungal genera and *Aspergillus* species from microscopic slides in a clinical laboratory context [26]. Two studies added attention mechanism to the U-Net architecture for microscopic cell segmentation and indicated improvement in model performance [27], [28]. In another microscopic domain, integrating the Convolutional Block Attention Module (CBAM) with U-Net and ResNet significantly improved thin-section rock image segmentation [29]. These findings suggest that incorporating attention mechanisms into established architectures can enhance classification and segmentation performance in microscopy, providing a strong motivation for exploring their role in fungal image analysis.

This study contributes by attempting an attention-mechanism CNN in the fungal classification domain. Experimentation with multiple standard image filtering techniques was also conducted to test whether the model's performance could be further improved. The architecture used in this study adopts the CBAM-VGGNet [30] with slight modification. This architecture was chosen because of its relatively small parameter count compared to the previously mentioned study, making it less computationally demanding and more suitable for settings with limited resources. Nonetheless, attention mechanisms may carry risks, such as overfitting to dataset-specific patterns or reduced interpretability, which warrant careful evaluation. Furthermore, issues of computational cost, ethical considerations, and deployment feasibility in real clinical workflows remain significant open challenges. These aspects, while beyond the scope of the current work, are acknowledged as crucial directions for future investigation into the safe and effective translation of such models into biomedical applications.

2. Method

To provide a clear overview of the research methodology, a flowchart summarizing the main steps of this study is presented in Fig. 1. The methodology consists of three sequential stages. The first stage, data preprocessing involves partitioning the dataset, applying data augmentation, and performing image filtering to enhance relevant features. The second stage, model training, includes training three models: VGG16, CBAM-VGGNet, and the modified CBAM-VGGNet. The final stage, evaluation and analysis, comprises calculating classification metrics, generating confusion matrices, and conducting statistical validation to assess model performance. Detailed descriptions is provided in each section.

2.1. Dataset

This study utilized the publicly accessible DeFungi dataset, a comprehensive collection of microscopic fungi images specifically designed for direct mycological examination, as introduced in previous research [25]. The images are from superficial fungal infections caused by moulds, yeasts, or dermatophyte fungi and have been manually labelled into 5 classes. The classes represented in this dataset are tortuous septate hyaline hyphae, beaded arthroconidial septate hyaline hyphae, groups or mosaics of arthroconidia, septate hyaline hyphae with chlamydioconidia, and broad brown hyphae. Sample image for each species can be seen in Fig. 2. Data distribution of each class is listed in Table

1. The dataset underwent preprocessing, resulting in 9,114 RGB images resized to 500×500 pixels. For naming convenience, the labelling of each species in the dataset will be as follows:

- Hypha 1: tortuous septate hyaline hyphae.
- Hypha 2: beaded arthroconidial septate hyaline hyphae.
- Hypha 3: groups or mosaics of arthroconidia.
- Hypha 5: septate hyaline hyphae with chlamydioconidia.
- Hypha 6: broad brown hyphae.

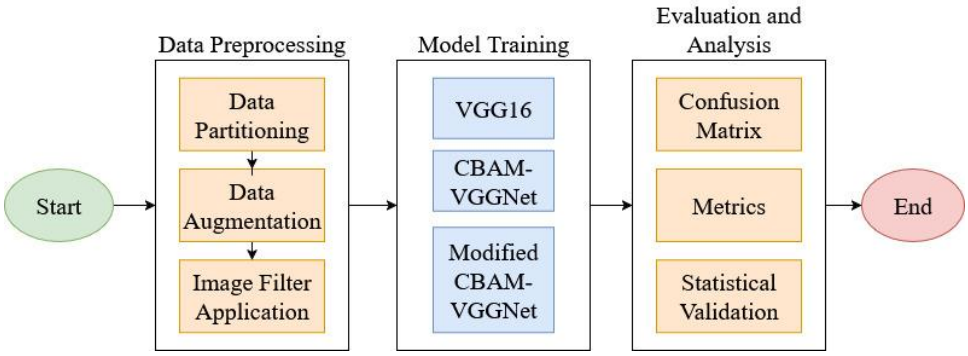

Fig. 1. Research methodology

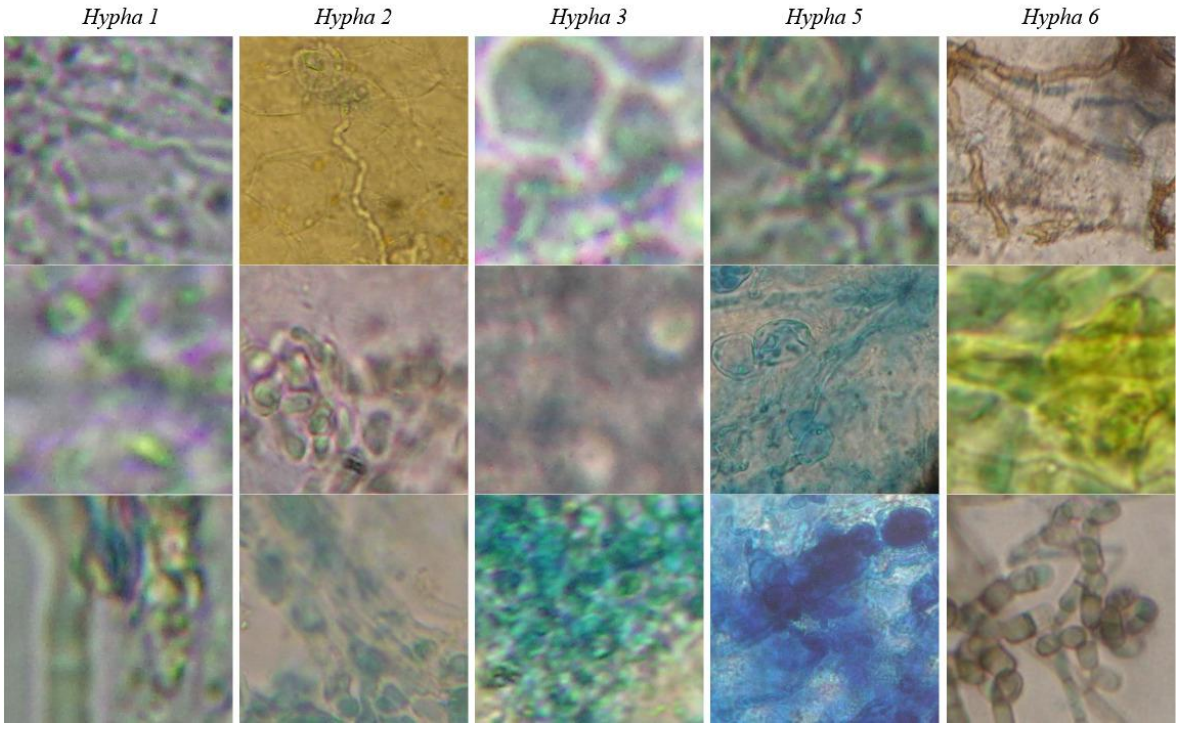

Fig. 2. DeFungi dataset sample images

Table 1. DeFungi dataset class distribution

Class	Image Count
Hypha 1	4404
Hypha 2	2334
Hypha 3	819
Hypha 5	818
Hypha 6	739

2.2. Data Augmentation

In deep learning, data augmentation serves as a prevalent approach to artificially increase the size of datasets by applying various transformations to the original data [31], [32]. This technique can address the problem of data imbalance and lack of data for cases where the dataset is small [33]-[35]. However, improper implementation of data augmentation can lead to issues such as data leakage [36]. Data augmentation was necessary to address the inherent class imbalance present in the DeFungi dataset. The dataset was first partitioned into training, validation, and testing subsets using stratified splitting, with augmentation applied only to the training data to preserve evaluation integrity. Given that the dataset images had already undergone preprocessing to capture all regions of interest containing microscopic fungal structures, a conservative augmentation strategy was adopted to minimize the risk of distorting diagnostically relevant features. Only two types of geometric transformations were applied: rotation (90°, 180°, 270°) and horizontal/vertical flipping, as these preserve spatial relationships and morphological characteristics. Additional combinations of rotation with flipping were also included, yielding seven unique transformations per image. This choice reflects the natural variability of microscopic sample orientation while avoiding more aggressive operations (e.g., cropping, scaling, color jitter) that could obscure fine-grained structures.

To prevent oversampling artifacts, the augmentation algorithm incorporated duplicate-prevention mechanisms, ensuring that no identical images were generated. The target number of samples per class (2,820) corresponds to the size of the majority class (Hypha 1) and was within the feasible range of unique augmentations even for the smallest class (Hypha 6, with 480 originals, which can yield up to 3,360 unique augmented images). As a result, all classes were balanced at 2,820 images without reusing the same augmentation multiple times per original, thereby improving dataset diversity while maintaining biological fidelity. The final distribution of augmented training data is provided in Table 2. To evaluate the impact of the data augmentation method, a preliminary ablation study was conducted by comparing classification metrics before and after augmentation. The results show consistent improvements across all classes, with macro-averaged precision, recall, and F1-score increasing from 0.771, 0.722, and 0.738 to 0.851, 0.822, and 0.834, respectively. Classes that were previously underrepresented, such as H2 and H3, showed the most notable gains, indicating that these augmentations enhance the model's ability to generalize while preserving the morphological features critical for accurate classification. The graph showing this improvement is shown in Fig. 3.

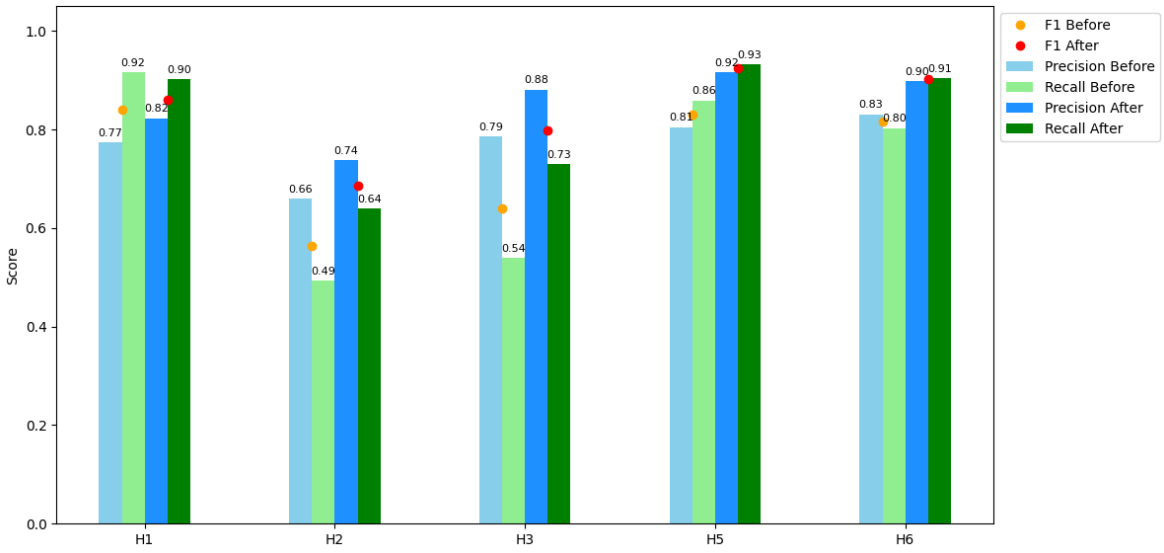

Fig. 3. Data augmentation ablation study results

2.3. Image Filter

Image filtering methods are commonly employed in computer vision applications to enhance model accuracy by improving feature discrimination and reducing imaging artifacts that may interfere

with classification [37], [38]. In this study, image preprocessing through filtering was explored to leverage these benefits and potentially improve classification performance on the microscopic fungi images. Various filtering approaches were systematically evaluated to determine their impact on model accuracy, including spatial domain filters and frequency domain enhancement methods. The following filtering techniques were implemented and assessed:

- Box filtering: A simple averaging filter that reduces noise by replacing each pixel with the average value of pixels within a defined kernel. This technique could potentially improve model accuracy by smoothing out imaging artifacts and noise that might confuse the classifier [39]-[41], allowing it to focus on genuine morphological patterns. A 3x3 box filter was implemented in this study using the `cv2.blur()` function from the OpenCV library.
- Gaussian filtering: A weighted averaging filter that reduces noise while better preserving edges compared to box filtering, using a Gaussian distribution to determine pixel weights. This approach may enhance classification performance by maintaining important morphological boundaries while reducing background noise, thus improving the signal-to-noise ratio of critical diagnostic features [42], [43]. A Gaussian filter is implemented in this study using the `cv2.GaussianBlur()` function from the OpenCV library.
- High-pass filtering: A technique that emphasizes fine structural details and edges by attenuating low-frequency components while preserving high-frequency information [44]. This filter could enhance model accuracy by increasing the visibility of crucial diagnostic features such as septation patterns and hyphal boundaries that distinguish between different fungal morphologies. High-pass filtering is implemented in this study using a combination of `cv2.GaussianBlur()` and `cv2.subtract()` functions from the OpenCV library.
- High Frequency Emphasis (HFE) filtering: An adaptive filtering method that selectively enhance components in high-frequency while preserving information in low-frequency through controlled enhancement [45]. This technique may improve classification performance by making subtle morphological differences between fungal classes more prominent while maintaining overall structural context. In this study, the implementation of HFE used a combination of `cv2.GaussianBlur()`, `cv2.subtract()` and `cv2.addWeighted()` function from OpenCV library.

The filtering operations were implemented using the default parameter settings of the OpenCV library. Each technique was evaluated independently on the preprocessed dataset to review its contribution. The comparative analysis aimed to identify optimal image filtering strategies that could complement the existing dataset preprocessing while maintaining the diagnostic integrity of the fungal specimens.

Table 2. Augmented data distribution of training set

Class	Image Count
Hypha 1	2820
Hypha 2	2820
Hypha 3	2820
Hypha 5	2820
Hypha 6	2820

2.4. Visual Geometry Group-16 Architecture

Visual Geometry Group-16 or VGG16 is a neural network deep learning model that employs a uniform architectural design consisting of sixteen weight layers, specifically thirteen convolutional layers paired with three fully connected layers [46]. The VGG16 model architecture is depicted in Fig. 4. The network's design utilizes 3×3 convolutional filters throughout all layers, which enables the model to capture fine-grained morphological details [47]. Max-pooling operations with 2×2 windows and a stride value of 2 are used to downsample the features across layers. VGG16 uses 3 fully connected layers as its classifier network.

VGG16 has been extensively applied to various tasks in medical image analysis [48]-[50]. Its application has also been explored in previous studies for classifying microscopic fungi images, yielding promising results [23], [24]. This proven track record in medical imaging suggests that VGG16's feature extraction capabilities are well-suited for the morphological analysis [51] required in microscopic fungi classification. Therefore, this architecture was chosen for this study.

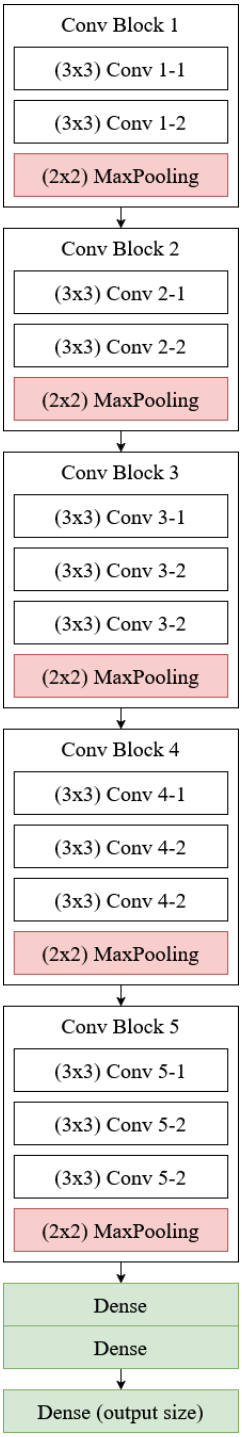

Fig. 4. VGG16 architecture

2.5. Convolutional Block Attention Module

CBAM is an attention module that is applicable to feed-forward convolutional neural networks to enhance their representational capabilities [52], [53]. This module incorporates two key

components: the Channel Attention Module (CAM) and the Spatial Attention Module (SAM). The module architecture is illustrated in Fig. 5. CBAM employs a sequential attention refinement strategy, which operates through its dual-component design. The module first applies channel attention, followed by spatial attention.

The CAM, as shown in Fig. 6, evaluates the importance of each feature channel by quantifying information from the spatial dimension of feature maps using average-pooling and max-pooling operations. These methods create two unique spatial context profiles that encompass varying dimensions of feature distributions. The profiles are subsequently fed through a unified network structure comprising a multi-layer perceptron (MLP) containing one hidden layer to construct the channel attention map. Once the shared network has completed processing each profile, the output vectors are integrated via element-wise summation.

The SAM, as shown in Fig. 7, complements the channel attention by focusing on the most relevant spatial locations within the feature maps. This component consolidates channel information through two pooling techniques, creating two 2D representations that correspond to average-pooled and max-pooled characteristics across all channels. These representations are subsequently merged and fed through a conventional convolutional layer to generate a 2D spatial attention map.

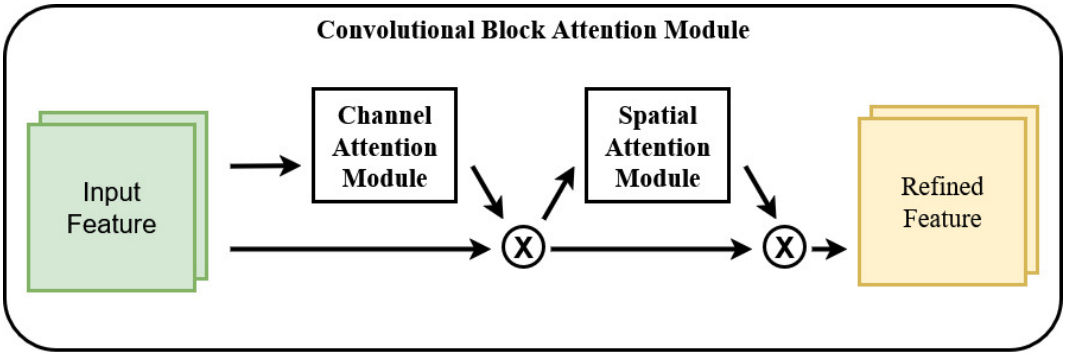

Fig. 5. Convolutional block attention module

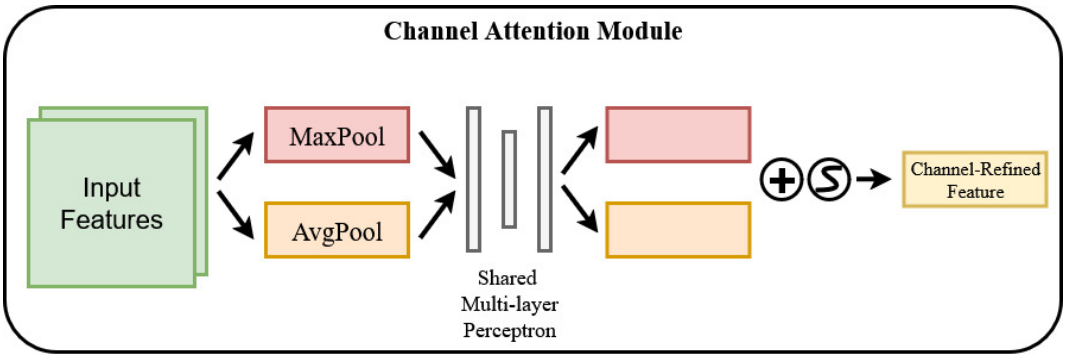

Fig. 6. CBAM: Channel attention module

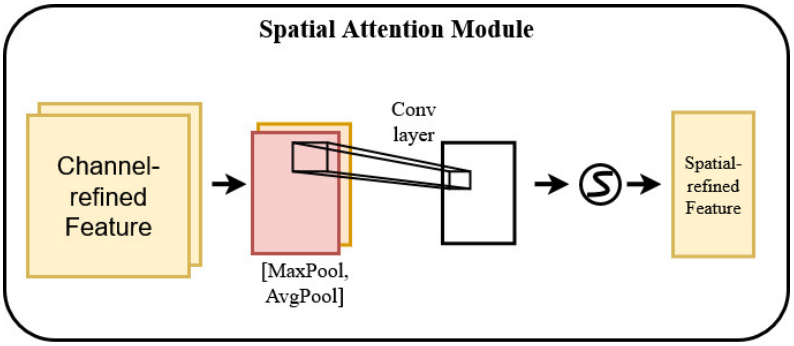

Fig. 7. CBAM: Spatial attention module

The integration of CBAM into convolutional networks provides several advantages. The attention mechanism enhances the network's discriminative capability by directing focus toward taxonomically significant features and regions, improving classification accuracy between morphologically similar species [47].

2.6. Proposed Model Architecture

The integration of attention mechanisms with an established CNN architecture has shown promising results across various computer vision tasks [54]-[58]. Previous study [54] has implemented a novel architecture combining both VGG16 and CBAM, called CBAM-VGGNet, to classify breast histopathology images. This architecture incorporates Global Average Pooling (GAP) layers after the first layer of each convolutional block. GAP layers from all convolutional blocks are interconnected, with CBAM modules positioned after each GAP layer to refine the extracted feature maps. The outputs from all CBAM blocks are subsequently concatenated to form a comprehensive feature representation before being forwarded to the fully connected classifier block inherited from the original VGG16 architecture for final classification. This architecture is shown in Fig. 8.

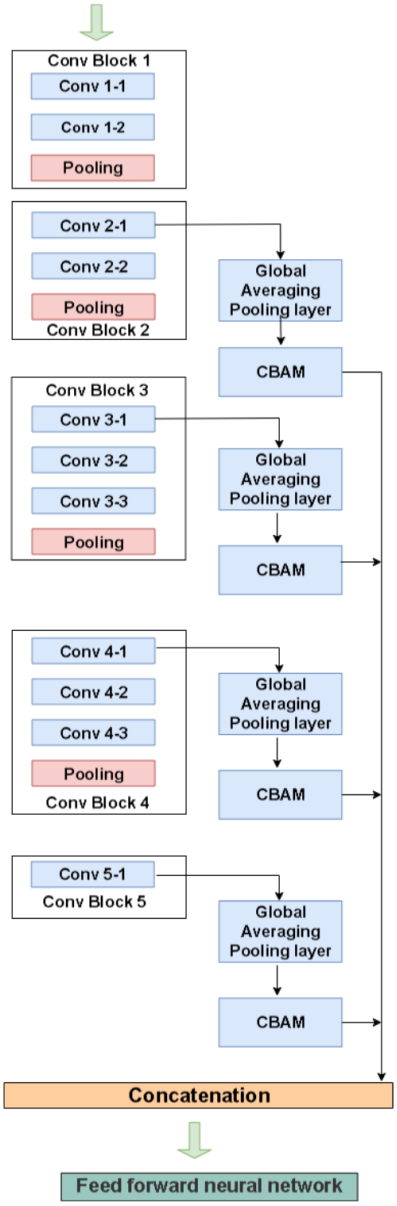

Fig. 8. CBAM-VGGNet architecture [54]

While CBAM comprises two modules that provide attention mechanisms in both spatial and channel dimensions, the original CBAM-VGGNet implementation presents a potential limitation in spatial information processing. The architecture applies CBAM after features pass through the Global Average Pooling layer, which reduces the feature dimensions from H × W × C (height × width × channel) to 1 × 1 × C. This dimensional reduction eliminates spatial information before CBAM processing, potentially compromising the effectiveness of the spatial attention module within CBAM.

To address this limitation, this study proposes a modified configuration where the two CBAM modules are decoupled and applied sequentially at different stages of the feature processing pipeline. Specifically, the SAM is positioned before the Global Average Pooling layer, allowing the attention mechanism to operate on the full spatial dimensions of the feature maps. Fig. 9 shows the attention maps produced by SAM (overlaid on the input image) for both models after passing the input image through the GAP layer in one of the blocks, highlighting the effect of GAP on spatial information. To quantify this effect, the variance and entropy of SAM activations are calculated for each block, with the results presented in Table 3. and Fig. 10. These metrics demonstrate that, in the original CBAM-VGGNet, SAM activations carry virtually no spatial information due to the collapsing effect of GAP, whereas in the modified architecture, spatial information is preserved and can contribute meaningfully to feature refinement. Meanwhile, CAM is applied after GAP, where it effectively captures channel-wise dependencies without requiring spatial dimensions. The full modified model architecture is shown in Fig. 11.

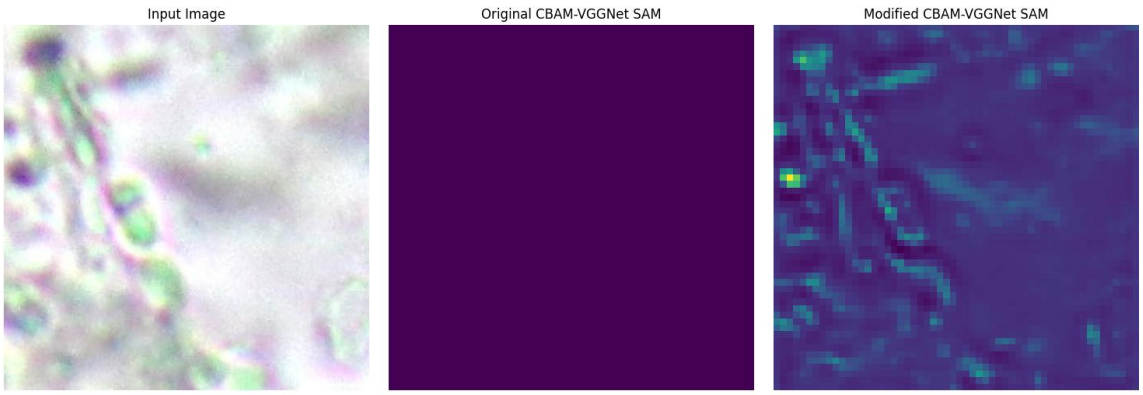

Fig. 9. SAM attention map in original and proposed architecture

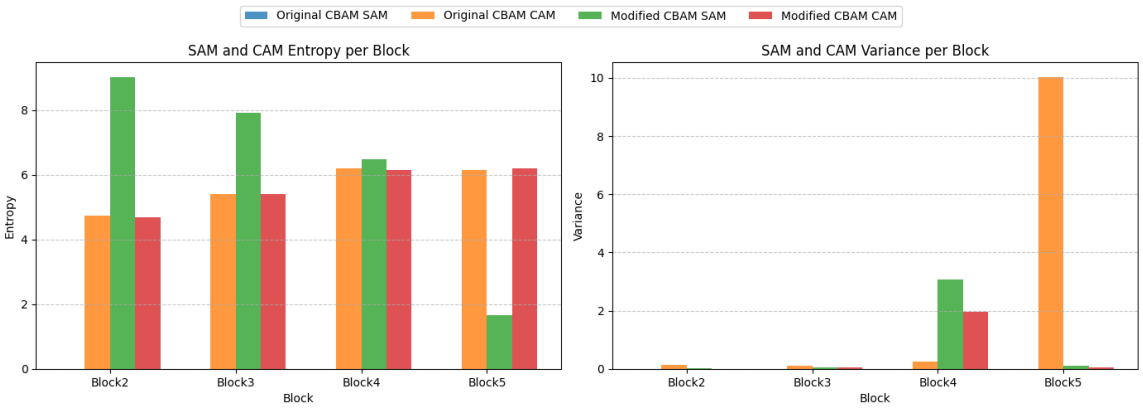

Fig. 10. Entropy and variance plotting for SAM and CAM for original and modified CBAM-VGGNet

2.7. Evaluation Metrics

Precision, recall, f1-score, accuracy, and confusion matrix were used to evaluate the model performance on classifying microscopic fungi images, which are the generally used metrics [59], [60]. Accuracy calculates the proportion of correctly classified instances relative to the total number of classification instances. The equation for this metric is:

$$Accuracy = \frac{Correctly\ classified\ samples}{Total\ samples} \times 100\% \quad (1)$$

Precision represents the proportion of true positives (TP) among the total positive predictions, consisting of both true positives and false positives (FP) classifications, for each class. The equation for this metric is:

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Recall represents the proportion of true positive predictions among all actual positive instances, which are true positives and false negatives (FN). The equation for this metric is:

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

F1-score represents the harmonic mean of recall and precision, a metric that considers both FP and FN. The equation for this metric is:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

The confusion matrix is a table used to evaluate classification performance on a labelled dataset. This metric provides detailed breakdowns of model predictions, showing the number of correct and wrong classification instances for each class.

Table 3. Variance (var) and entropy (ent) of SAM and CAM from original and modified CBAM-VGGNet

Model	Block	SAM var	SAM ent	CAM var	CAM ent
Original CBAM-VGGNet	2	0.0000	0.0000	0.1242	4.7439
	3	0.0000	0.0000	0.9934	5.3998
	4	0.0000	0.0000	0.2423	6.2027
	5	0.0000	0.0000	10.0238	6.1513
Modified CBAM-VGGNet	2	0.0066	9.0257	0.0043	4.6944
	3	0.0436	7.9201	0.0374	5.4157
	4	3.0773	6.4909	1.9472	6.1537
	5	0.1146	1.6594	0.0564	6.1898

2.8. Experimentation

Data augmentation and image filtering techniques were applied to the DeFungi dataset, as illustrated in Fig. 12. The microscopic fungi images were resized to 224×224 pixels to meet the input requirements of the VGG16-based architectures. Dataset partitioning was performed using a 65:15:20 split ratio for the training, validation, and test sets, respectively. Data augmentation was subsequently applied solely to the training set to prevent data leakage and maintain evaluation integrity. The partitioning process is stratified to ensure a balanced distribution between classes, as the dataset is imbalanced. The preprocessing pipeline generated a total of five distinct datasets: four datasets with individual image filter applications and one baseline dataset without filtering. This systematic approach enables a comprehensive evaluation of how different image preprocessing techniques affect model performance in microscopic fungi classification tasks.

Three deep learning architectures were evaluated in this study: the baseline VGG16 model, the original CBAM-VGGNet implementation, and the proposed modified CBAM-VGGNet architecture. The transfer learning technique was employed across all models to leverage established feature representations for improved convergence and performance. This technique has been proven to be effective on medical images [61]-[63].

All three model architectures were trained and evaluated using all five preprocessed datasets, providing a thorough assessment of both architectural modifications and preprocessing effects. Early

stopping method was the chosen regularization technique to prevent overfitting during the training process [64], [65]. The specific training parameters and hyperparameters utilized in the experiment are detailed in Table 4. The selected values (optimizer, learning rate, activation function, dense layer units, dropout rate) were determined through manual experimentation on preliminary runs to achieve stable training and reasonable performance.

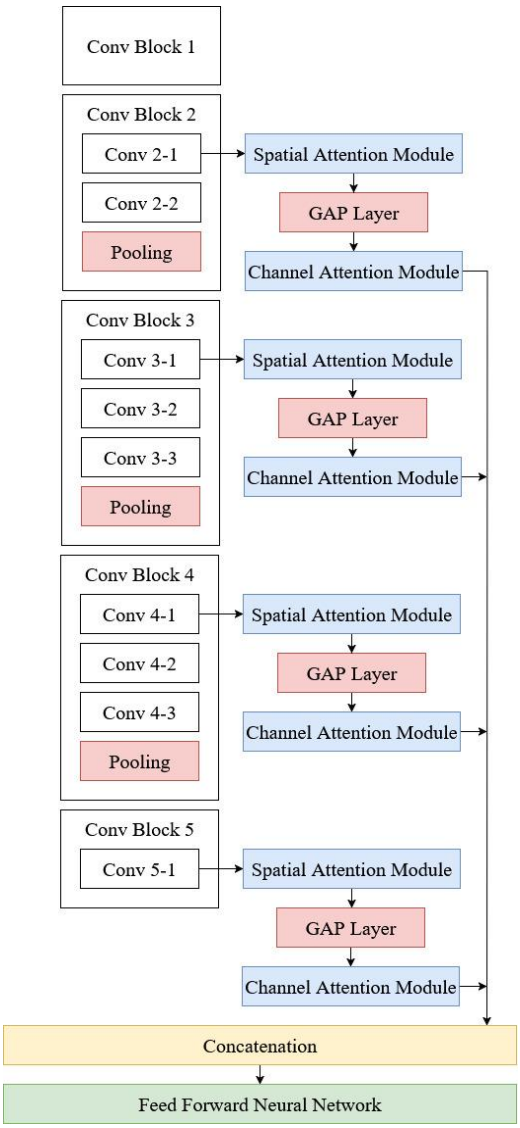

Fig. 11. Proposed CBAM-VGGNet architecture

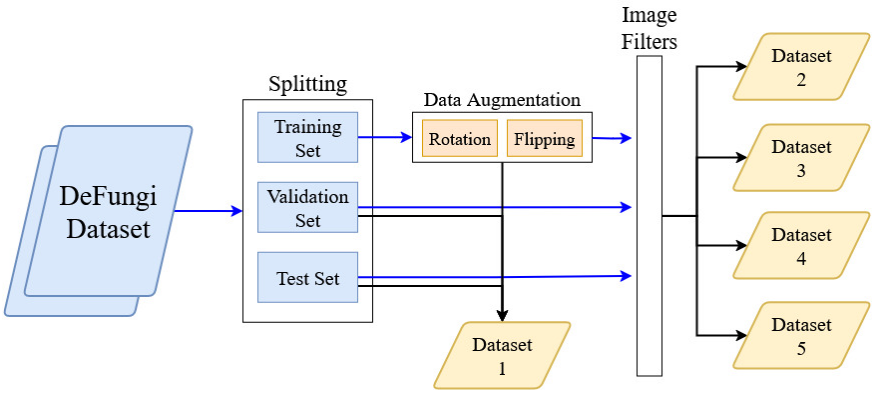

Fig. 12. Data preprocessing diagram

Table 4. Training parameters

Parameter	Value
Optimizer	adam
Learning rate	0.0001
Loss function	categorical_crossentropy
Dropout value	0.4
Activation function	LeakyRelu
dense Layer neuron	256

3. Results and Discussion

3.1. Baseline VGG16 Model Results

The baseline VGG16 model demonstrated varying performance across different preprocessing techniques, as shown in Table 5. The box filter achieved the highest accuracy (78.88%), followed by the Gaussian filter (78.66%) and the HFE filter (77.18%). The unfiltered dataset yielded 75.53%, while the high-pass filter performed worst at 48.37%. The confusion matrix presented in Fig. 13 reveals the model's classification patterns, with H1 showing exceptional performance (832 correct predictions with minimal misclassification), while H2 exhibited the most confusion with significant misclassifications across multiple classes, particularly being confused with H1 (225 instances).

Per-class metrics listed in Table 6 highlight that H6 achieved the highest precision (0.942) and F1-score (0.908), whereas H2 had the poorest recall (0.418), indicating class-specific challenges. Overall, smoothing filters (box and Gaussian) improved model stability, whereas edge-enhancing filters, such as high-pass, reduced performance—suggesting that baseline VGG16 is sensitive to high-frequency noise. The baseline VGG16 exhibits moderate overall accuracy, with performance sensitive to preprocessing type and persistent difficulties distinguishing H2 from H1.

Table 5. Baseline VGG16 model accuracy

Dataset	Accuracy
No filter	0.7553
Box filter	0.7888
Gaussian filter	0.7866
High-pass filter	0.4837
HFE filter	0.7718

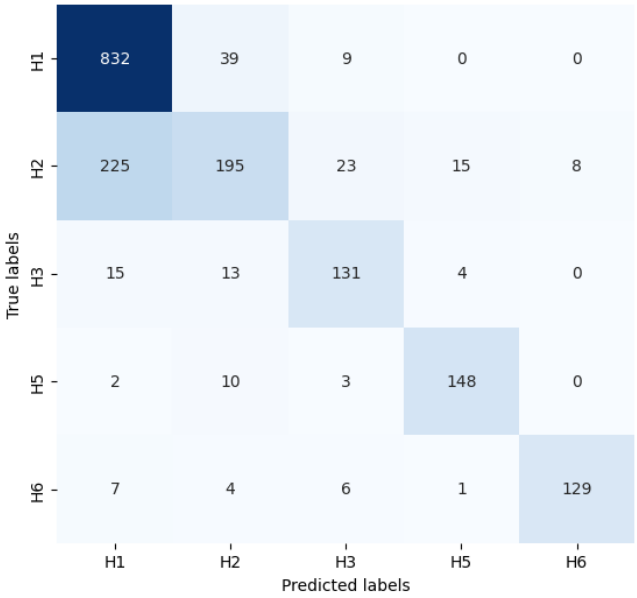

Fig. 13. Baseline VGG16 with box image filter dataset confusion matrix

Table 6. Baseline VGG16 with box image filter dataset performance metrics

Class	Metrics		
	Precision	Recall	F1-Score
Hypha 1	0.770	0.945	0.849
Hypha 2	0.747	0.418	0.536
Hypha 3	0.762	0.804	0.782
Hypha 5	0.881	0.908	0.894
Hypha 6	0.942	0.878	0.908

3.2. Original CBAM-VGGNet Model Result

The original CBAM-VGGNet outperformed the baseline VGG16 across all preprocessing conditions, as illustrated in Table 7. Peak accuracy was 86.86% on the unfiltered dataset, followed closely by Gaussian (86.58%) and box filters (84.33%). Even on the high-pass filter, performance improved to 60.47%, demonstrating the attention mechanism's ability to emphasize relevant features. The confusion matrix depicted in Fig. 14 for the unfiltered dataset reveals significantly improved classification performance, with H1 achieving 797 correct predictions and substantially reduced misclassifications compared to the baseline model. The attention mechanism particularly benefited H2 classification, which showed marked improvement with 348 correct predictions, though some confusion with H1 persisted (96 instances). As detailed in Table 8. metrics show high precision and recall across all classes, with H6 achieving a precision of 0.959 and a recall of 0.952. Integrating CBAM enhanced feature discrimination and improved H2 classification, but certain inter-class confusions remain. The model consistently benefits from attention mechanisms, particularly on unfiltered and smoothing-filtered datasets, although performance on high-pass filtering indicates sensitivity to high-frequency components.

Table 7. Original CBAM-VGGNet model accuracy

Dataset	Accuracy
No filter	0.8686
Box filter	0.8433
Gaussian filter	0.8658
High-pass filter	0.6047
HFE filter	0.8526

Table 8. Original CBAM-VGGNet with no image filter dataset performance metrics

Class	Metrics		
	Precision	Recall	F1-Score
Hypha 1	0.879	0.906	0.892
Hypha 2	0.780	0.747	0.763
Hypha 3	0.860	0.865	0.862
Hypha 5	0.987	0.945	0.966
Hypha 6	0.959	0.952	0.956

3.3. Proposed CBAM-VGGNet Model Results

The proposed modified CBAM-VGGNet demonstrated the strongest overall performance, as shown in Table 9. The proposed model reached its peak accuracy of 94.33% with the HFE filter, substantially outperforming both the original CBAM-VGGNet (86.86%) and baseline VGG16 (78.88%). Across all preprocessing methods, the model maintained high performance, with the box filter at 91.97%, the Gaussian at 87.24%, and the high-pass at 50.75%. The confusion matrix shown in Fig. 15 indicates that H1 is correctly classified in 856 instances, and H2 is substantially improved with 409 correct predictions. The per-class performance metrics detailed in Table 10 indicate that H1 achieved a precision of 0.933 and a recall of 0.973, H6 achieved a precision of 0.979 and a recall of 0.952, and H2 achieved a precision of 0.947 and a recall of 0.878, highlighting that the modification improves the discrimination of previously challenging classes. The HFE filter's strong performance suggests the model effectively leverages edge-enhanced features; however, persistent misclassifications (e.g., occasional H2-H1 confusion) indicate morphological similarity between

classes remains challenging. Moreover, while the model achieves high accuracy, the performance drop on high-pass filtered images suggests sensitivity to high-frequency noise and potential limitations in generalizing to datasets with stronger edge-dominant artifacts. The modified architecture consistently outperforms both baseline models and demonstrates improved class-wise discrimination, particularly for H2, while maintaining robustness across multiple preprocessing techniques. Remaining confusions and sensitivity to high-pass filtering highlight areas for future refinement. Computationally, the modifications introduced minimal additional overhead relative to the original CBAM-VGGNet.

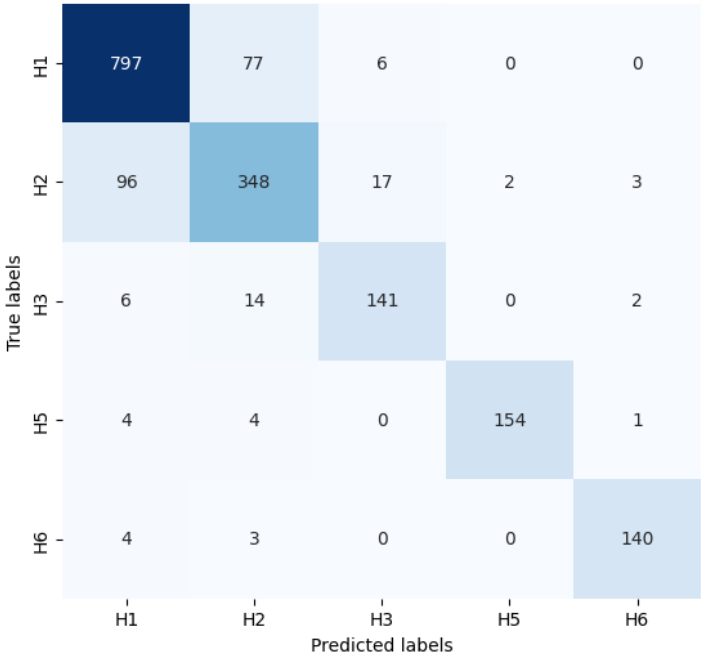

Fig. 14. Original CBAM-VGGNet with no image filter dataset confusion matrix

Table 9. Proposed model accuracy

Dataset	Accuracy
No filter	0.8829
Box filter	0.9197
Gaussian filter	0.8724
Highpass filter	0.5075
HFE filter	**0.9433**

3.4. Statistical Validation Test and Additionals

To validate the performance differences between the original CBAM-VGGNet and the proposed modified CBAM-VGGNet, both models were trained and evaluated over three independent runs using the same dataset (dataset without any image filter). The repeated evaluations account for stochastic variation in training. The modified CBAM-VGGNet achieved a mean accuracy of 0.893, with a 95% confidence interval (CI) of 0.874–0.910, compared to 0.856 (95% CI: 0.848–0.871) for the original CBAM-VGGNet, indicating a consistent improvement. McNemar's test was conducted on the predictions from the first run yielding a p-value of 0.0027 which demonstrates that the improvement in classification performance is statistically significant. The average per-class metrics, summarized in Table 11 further confirm that the gains are not limited to a single class; improvements in precision, recall, and F1-score were observed across underrepresented classes such as H2 and H3. These results indicate that the proposed modifications enhance model generalization and reliability while accounting for variability due to limited dataset size and class imbalance.

Overall, the experimental results demonstrate a clear performance progression from the baseline VGG16 to the original CBAM-VGGNet and finally the proposed modified CBAM-VGGNet, with the latter achieving the highest accuracy across all preprocessing conditions, particularly with the HFE

filter (94.33%). The modified model consistently improved per-class precision and recall, notably for previously challenging classes such as H2, while maintaining strong performance for H1 and H6. Despite these improvements, all models showed reduced accuracy on high-pass filtered images, suggesting that edge-focused preprocessing could be further refined to better preserve biologically relevant features. Additionally, due to the lack of expert mycological evaluation, it remains unclear whether certain preprocessing techniques, particularly high-pass and HFE filtering, introduce artifacts or obscure subtle fungal structures. These limitations highlight opportunities for further optimization of preprocessing strategies and the potential benefit of interdisciplinary collaboration to fully assess the biological validity of attention-enhanced feature representations.

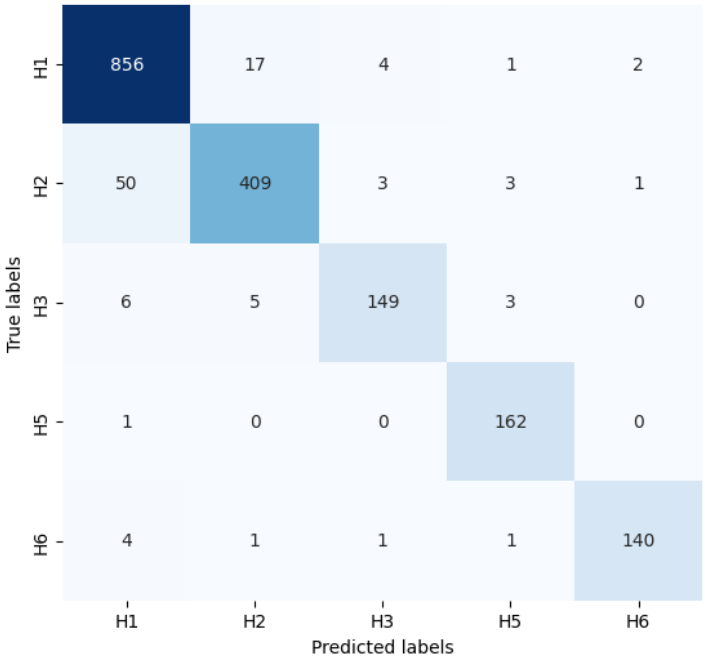

Fig. 15. Proposed model with HFE image filter dataset confusion matrix

Table 10. Proposed model with HFE image filter dataset performance metrics

Class	Metrics		
	Precision	Recall	F1-Score
Hypha 1	0.933	0.973	0.953
Hypha 2	0.947	0.878	0.911
Hypha 3	0.949	0.914	0.931
Hypha 5	0.953	0.994	0.973
Hypha 6	0.979	0.952	0.966

Table 11. Average per-class metrics from 3 runs using dataset with no filter

Class	CBAM-VGGNet			Modified CBAM-VGGNet		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Hypha 1	0.851	0.926	0.887	0.909	0.914	0.912
Hypha 2	0.812	0.687	0.738	0.834	0.820	0.826
Hypha 3	0.885	0.814	0.844	0.027	0.853	0.888
Hypha 5	0.953	0.941	0.946	0.934	0.959	0.946
Hypha 6	0.911	0.925	0.918	0.910	0.971	0.939

4. Conclusion

In conclusion, this study demonstrates that attention-enhanced deep learning architectures, combined with systematic preprocessing strategies, can substantially improve microscopic fungi classification performance despite dataset limitations. The proposed modified CBAM-VGGNet

architecture achieved a classification accuracy of 94.33% using the HFE filter, showing notable improvements over both the original CBAM-VGGNet (86.86%) and the baseline VGG16 (78.88%). These results highlight the effectiveness of integrating CBAM attention mechanisms into VGG16, enabling the model to focus on distinctive morphological features within fungal images and improving recall in previously problematic classes (e.g., H2 recall increased from 0.418 to 0.878).

While the reported performance exceeds the 92.49% accuracy of the MeFunX meta-learning framework on the DeFungi dataset, this comparison should be interpreted with caution, as no direct head-to-head evaluation under identical experimental conditions or statistical significance testing was performed. Moreover, the findings are currently limited to a single dataset, and external validation on independent samples is needed to assess generalizability across diverse imaging conditions. Additional limitations include reliance on legacy architecture (VGG16), lack of interpretability analysis, and absence of computational benchmarking, all of which are important considerations for clinical deployment.

Future work should explore comparisons with more modern and parameter-efficient architectures (e.g., ConvNeXt, Vision Transformers), incorporate model interpretability and error analysis to better understand misclassifications, and evaluate lightweight optimization strategies such as pruning or quantization to improve feasibility in resource-limited settings. Despite these limitations, the study provides evidence that attention mechanisms can meaningfully enhance fungal classification performance and underscores the potential of attention-based architectures in advancing automated mycological diagnostics.

Author Contribution: All authors contributed equally to the main contributor to this paper. All authors read and approved the final paper.

Funding: This research received no external funding

Acknowledgment: The authors would like to thank the Department of Informatics, Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia, for supporting this research.

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] D. W. Denning, "Global incidence and mortality of severe fungal disease," *Lancet Infectious Diseases*, vol. 24, no. 7, pp. e428-e438, 2024, https://doi.org/10.1016/S1473-3099(23)00692-8.

[2] M. K. Khoury *et al*., "Antifungal Therapy in Fungal Necrotizing Soft Tissue Infections," *Journal of Surgical Research*, vol. 256, pp. 187-192, 2020, https://doi.org/10.1016/j.jss.2020.06.013.

[3] J. Acharya *et al*., "Review of neuroimaging findings of intracranial angioinvasive fungal infections," *Clinical Imaging*, vol. 115, p. 110306, 2024, https://doi.org/10.1016/j.clinimag.2024.110306.

[4] C. Rodríguez-Vargas, A. Alastruey-Izquierdo, D. W. Denning, and A. Belén Araúz, "Estimated burden of fungal infections in Panama," *Journal of Medical Mycology*, vol. 34, no. 1, p. 101466, 2024, https://doi.org/10.1016/j.mycmed.2024.101466.

[5] Z. U. Mustafa *et al*., "Predictors and outcomes of patients with COVID-19 admitted to intensive care units in Pakistan and the development of nosocomial fungal infections: Findings and implications," *IJID Regions*, vol. 13, p. 100445, 2024, https://doi.org/10.1016/j.ijregi.2024.100445.

[6] W. Fang *et al*., "Diagnosis of invasive fungal infections: challenges and recent developments," *Journal of Biomedical Science*, vol. 30, no. 1, p. 42, 2023, https://doi.org/10.1186/s12929-023-00926-2.

[7] M. Zhang *et al*., "Appropriate empirical antifungal therapy is associated with a reduced mortality rate in intensive care unit patients with invasive fungal infection: A real-world retrospective study based on the MIMIC-IV database," *Frontiers in Medicine*, vol. 9, 2022, https://doi.org/10.3389/fmed.2022.952611.

[8] T. Pablo, A. Mzabi, M. Meo, F. Decruyenaere, and M. Perrin, "Routine laboratory test enabling the detection of dermatophytes and the identification of Trichophyton rubrum by means of in-house duplex real-time PCR," *Journal of Microbiological Methods*, vol. 185, p. 106229, 2021, https://doi.org/10.1016/j.mimet.2021.106229.

[9] C. Leeyaphan, C. Chai-Adisaksopha, N. Tovanabutra, P. Phinyo, and S. Bunyaratavej, "Developing diagnostic criteria to differentiate fungal foot infections caused by Neoscytalidium dimidiatum and dermatophytes," *Heliyon*, vol. 9, no. 8, p. e18963, 2023, https://doi.org/10.1016/j.heliyon.2023.e18963.

[10] A. K. Gautam *et al*., "Current Insight into Traditional and Modern Methods in Fungal Diversity Estimates," *Journal of Fungi*, vol. 8, no. 3, p. 226, 2022, https://doi.org/10.3390/jof8030226.

[11] A. J. Morris *et al*., "Update on methods used for mycological testing: wide diversity and opportunities for improvement persist," *Pathology*, vol. 56, no. 7, pp. 1021-1027, 2024, https://doi.org/10.1016/j.pathol.2024.06.007.

[12] J. Salmanton-García *et al*., "The current state of laboratory mycology in Asia/Pacific: A survey from the European Confederation of Medical Mycology (ECMM) and International Society for Human and Animal Mycology (ISHAM)," *International Journal of Antimicrobial Agents*, vol. 61, no. 3, p. 106718, 2023, https://doi.org/10.1016/j.ijantimicag.2023.106718.

[13] C. Driemeyer *et al*., "The current state of clinical mycology in Africa: a European Confederation of Medical Mycology and International Society for Human and Animal Mycology survey," *The Lancet Microbe*, vol. 3, no. 6, pp. e464-e470, 2022, https://doi.org/10.1016/S2666-5247(21)00190-7.

[14] C. Chakraborty, M. Bhattacharya, S. Pal, and S.-S. Lee, "From machine learning to deep learning: Advances of the recent data-driven paradigm shift in medicine and healthcare," *Current Research in Biotechnology*, vol. 7, p. 100164, 2024, https://doi.org/10.1016/j.crbiot.2023.100164.

[15] A. Ettalibi, A. Elouadi, and A. Mansour, "AI and Computer Vision-based Real-time Quality Control: A Review of Industrial Applications," *Procedia Computer Science*, vol. 231, pp. 212-220, 2024, https://doi.org/10.1016/j.procs.2023.12.195.

[16] J. Lv *et al*., "Deep learning-based automated diagnosis of fungal keratitis with in vivo confocal microscopy images," *Annals of Translational Medicine*, vol. 8, no. 11, pp. 706-706, 2020, https://doi.org/10.21037/atm.2020.03.134.

[17] B. Zieliński, A. Sroka-Oleksiak, D. Rymarczyk, A. Piekarczyk, and M. Brzychczy-Włoch, "Deep learning approach to describe and classify fungi microscopic images," *PLoS One*, vol. 15, no. 6, p. e0234806, 2020, https://doi.org/10.1371/journal.pone.0234806.

[18] S. Karthikeyan, G. Ramkumar, S. Aravindkumar, M. Tamilselvi, S. Ramesh, and A. Ranjith, "A Novel Deep Learning-Based Black Fungus Disease Identification Using Modified Hybrid Learning Methodology," *Contrast Media & Molecular Imaging*, vol. 2022, no. 1, pp. 1-11, 2022, https://doi.org/10.1155/2022/4352730.

[19] T. Koo, M. H. Kim, and M.-S. Jue, "Automated detection of superficial fungal infections from microscopic images through a regional convolutional neural network," *PLoS One*, vol. 16, no. 8, p. e0256290, 2021, https://doi.org/10.1371/journal.pone.0256290.

[20] W. Gao *et al*., "The design and application of an automated microscope developed based on deep learning for fungal detection in dermatology," *Mycoses*, vol. 64, no. 3, pp. 245-251, 2021, https://doi.org/10.1111/myc.13209.

[21] A. Yilmaz, F. Göktay, R. Varol, G. Gencoglan, and H. Uvet, "Deep convolutional neural networks for onychomycosis detection using microscopic images with KOH examination," *Mycoses*, vol. 65, no. 12, pp. 1119-1126, 2022, https://doi.org/10.1111/myc.13498.

[22] A. S. Cagatan, M. T. Mustapha, C. Bagkur, T. Sanlidag, and D. U. Ozsahin, "An Alternative Diagnostic Method for C. neoformans: Preliminary Results of Deep-Learning Based Detection Model," *Diagnostics*, vol. 13, no. 1, p. 81, 2022, https://doi.org/10.3390/diagnostics13010081.

[23] M. A. Rahman *et al*., "Classification of fungal genera from microscopic images using artificial intelligence," *Journal of Pathology Informatics*, vol. 14, p. 100314, 2023, https://doi.org/10.1016/j.jpi.2023.100314.

[24] S. Rawat, B. Bisht, V. Bisht, N. Rawat, and A. Rawat, "MeFunX: A novel meta-learning-based deep learning architecture to detect fungal infection directly from microscopic images," *Franklin Open*, vol. 6, p. 100069, 2024, https://doi.org/10.1016/j.fraope.2023.100069.

[25] M. A. V. Álvarez, L. Sopó, C. J. P. Sopo, F. Hajati, and S. Gheisari, "P456 Defungi: direct mycological examination of microscopic fungi images," *Medical Mycology*, vol. 60, no. Supplement_1, p. myac072P456, 2022, https://doi.org/10.1093/mmy/myac072.P456.

[26] T.-S. Huang, K. Wang, X.-Y. Ye, C.-S. Chen, and F.-C. Chang, "Attention-Guided Transfer Learning for Identification of Filamentous Fungi Encountered in the Clinical Laboratory," *Microbiology Spectrum*, vol. 11, no. 3, p. e04611-22, 2023, https://doi.org/10.1128/spectrum.04611-22.

[27] Y. Zhou, K. Ma, Q. Sun, Z. Wang, and M. Liu, "Edge-Guided Cell Segmentation on Small Datasets Using an Attention-Enhanced U-Net Architecture," *Information*, vol. 15, no. 4, p. 198, 2024, https://doi.org/10.3390/info15040198.

[28] J. Niyogisubizo, K. Zhao, J. Meng, Y. Pan, R. Didi, and Y. Wei, "Attention-Guided Residual U-Net with SE Connection and ASPP for Watershed-Based Cell Segmentation in Microscopy Images," *Journal of Computational Biology*, vol. 32, no. 2, pp. 225-237, 2025, https://doi.org/10.1089/cmb.2023.0446.

[29] L. Zhao, H. Zhang, X. Sun, Z. Ouyang, C. Xu, and X. Qin, "Application of ResUNet-CBAM in Thin-Section Image Segmentation of Rocks," *Information*, vol. 15, no. 12, p. 788, 2024, https://doi.org/10.3390/info15120788.

[30] A. Ijaz *et al*., "Modality Specific CBAM-VGGNet Model for the Classification of Breast Histopathology Images via Transfer Learning," *IEEE Access*, vol. 11, pp. 15750-15762, 2023, https://doi.org/10.1109/ACCESS.2023.3245023.

[31] K. Maharana, S. Mondal, and B. Nemade, "A review: Data pre-processing and data augmentation techniques," *Global Transitions Proceedings*, vol. 3, no. 1, pp. 91-99, 2022, https://doi.org/10.1016/j.gltp.2022.04.020.

[32] T. Islam, Md. S. Hafiz, J. R. Jim, Md. M. Kabir, and M. F. Mridha, "A systematic review of deep learning data augmentation in medical imaging: Recent advances and future research directions," *Healthcare Analytics*, vol. 5, p. 100340, 2024, https://doi.org/10.1016/j.health.2024.100340.

[33] A. A. Khan, O. Chaudhari, and R. Chandra, "A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation," *Expert Systems with Applications*, vol. 244, p. 122778, 2024, https://doi.org/10.1016/j.eswa.2023.122778.

[34] R. Poojary, R. Raina, and A. Kumar Mondal, "Effect of data-augmentation on fine-tuned CNN model performance," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 10, no. 1, pp. 84-92, 2021, http://doi.org/10.11591/ijai.v10.i1.pp84-92.

[35] K. Alomar, H. I. Aysel, and X. Cai, "Data Augmentation in Classification and Segmentation: A Survey and New Strategies," *Journal of Imaging*, vol. 9, no. 2, p. 46, 2023, https://doi.org/10.3390/jimaging9020046.

[36] I. E. Tampu, A. Eklund, and N. Haj-Hosseini, "Inflation of test accuracy due to data leakage in deep learning-based classification of OCT images," *Scientific Data*, vol. 9, no. 1, p. 580, 2022, https://doi.org/10.1038/s41597-022-01618-6.

[37] J. Rasheed, "Analyzing the Effect of Filtering and Feature-Extraction Techniques in a Machine Learning Model for Identification of Infectious Disease Using Radiography Imaging," *Symmetry*, vol. 14, no. 7, p. 1398, 2022, https://doi.org/10.3390/sym14071398.

[38] L. Huang, C. Yao, L. Zhang, S. Luo, F. Ying, and W. Ying, "Enhancing computer image recognition with improved image algorithms," *Scientific Reports*, vol. 14, no. 1, p. 13709, 2024, https://doi.org/10.1038/s41598-024-64193-3.

[39] M. Fukatsu, S. Yoshizawa, H. Takemura, and H. Yokota, "Fast Scale-Aware Image Filtering via Half-Box Regions," *Journal of the Japan Society for Precision Engineering*, vol. 90, no. 3, pp. 313-320, 2024, https://doi.org/10.2493/jjspe.90.313.

[40] M. Fukatsu, S. Yoshizawa, H. Takemura, and H. Yokota, "Bilateral Half-Box Image Filtering," *International Journal of Automation Technology*, vol. 18, no. 3, pp. 427-432, 2024, https://doi.org/10.20965/ijat.2024.p0427.

[41] D. Gautam, K. Khare, and B. P. Shrivastava, "A Novel Guided Box Filter Based on Hybrid Optimization for Medical Image Denoising," *Applied Sciences*, vol. 13, no. 12, p. 7032, 2023, https://doi.org/10.3390/app13127032.

[42] H. Lin *et al*., "Gaussian filter facilitated deep learning-based architecture for accurate and efficient liver tumor segmentation for radiation therapy," *Frontiers in Oncology*, vol. 14, 2024, https://doi.org/10.3389/fonc.2024.1423774.

[43] O. Turk, E. Acar, E. Irmak, M. Yilmaz, and E. Bakis, "A Hybrid 2D Gaussian Filter and Deep Learning Approach with Visualization of Class Activation for Automatic Lung and Colon Cancer Diagnosis," *Technology in Cancer Research & Treatment*, vol. 23, 2024, https://doi.org/10.1177/15330338241301297.

[44] Z. Jin, J. Cao, M. Zhang, and Q.-S. Xiang, "Using High-Pass Filter to Enhance Scan Specific Learning for MRI Reconstruction without Any Extra Training Data," *Neuroimage*, vol. 303, p. 120926, 2024, https://doi.org/10.1016/j.neuroimage.2024.120926.

[45] W. Pan, W. Ma, X. Wu, and W. Liu, "High Frequency Component Enhancement Network for Image Manipulation Detection," *Electronics*, vol. 13, no. 2, p. 447, 2024, https://doi.org/10.3390/electronics13020447.

[46] X. Lu, H. Wang, J. Zhang, Y. Zhang, J. Zhong, and G. Zhuang, "Research on J wave detection based on transfer learning and VGG16," *Biomedical Signal Processing and Control*, vol. 95, p. 106420, 2024, https://doi.org/10.1016/j.bspc.2024.106420.

[47] Z. Cao *et al*., "Fine-grained image classification on bats using VGG16-CBAM: a practical example with 7 horseshoe bats taxa (CHIROPTERA: Rhinolophidae: Rhinolophus) from Southern China," *Frontiers in Zoology*, vol. 21, no. 1, p. 10, 2024, https://doi.org/10.1186/s12983-024-00531-5.

[48] Y. Chen, Y. Chen, S. Fu, W. Yin, K. Liu, and S. Qian, "VGG16-based intelligent image analysis in the pathological diagnosis of IgA nephropathy," *Journal of Radiation Research and Applied Sciences*, vol. 16, no. 3, p. 100626, 2023, https://doi.org/10.1016/j.jrras.2023.100626.

[49] S. U. Krishna, A. N. B. Lakshman, T. Archana, K. Raja, and M. Ayyadurai, "Lung Cancer Prediction and Classification Using Decision Tree and VGG16 Convolutional Neural Networks," *The Open Biomedical Engineering Journal*, vol. 18, no. 1, 2024, https://doi.org/10.2174/0118741207290271240322061032.

[50] Z. Liu, J. Peng, X. Guo, S. Chen, and L. Liu, "Breast cancer classification method based on improved VGG16 using mammography images," *Journal of Radiation Research and Applied Sciences*, vol. 17, no. 2, p. 100885, 2024, https://doi.org/10.1016/j.jrras.2024.100885.

[51] M. Lachmann *et al*., "Harnessing feature extraction capacities from a pre-trained convolutional neural network (VGG-16) for the unsupervised distinction of aortic outflow velocity profiles in patients with severe aortic stenosis," *European Heart Journal - Digital Health*, vol. 3, no. 2, pp. 153-168, 2022, https://doi.org/10.1093/ehjdh/ztac004.

[52] S. Agac and O. Durmaz Incel, "On the Use of a Convolutional Block Attention Module in Deep Learning-Based Human Activity Recognition with Motion Sensors," *Diagnostics*, vol. 13, no. 11, p. 1861, 2023, https://doi.org/10.3390/diagnostics13111861.

[53] Y. Wang, X. Chen, J. Li, and Z. Lu, "Convolutional Block Attention Module–Multimodal Feature-Fusion Action Recognition: Enabling Miner Unsafe Action Recognition," *Sensors*, vol. 24, no. 14, p. 4557, 2024, https://doi.org/10.3390/s24144557.

[54] V. Winnarasi A, B. Vaishnavi and A. Veluppal, "Enhanced Breast Cancer Diagnosis: Leveraging Customized Transfer Learning with Machine Learning and Attention Mechanisms for Histopathology Image Classification," *2024 7th International Conference on Circuit Power and Computing Technologies (ICCPCT)*, pp. 1540-1545, 2024, https://doi.org/10.1109/ICCPCT61902.2024.10673240.

[55] C. H. Praharsha and A. Poulose, "CBAM VGG16: An efficient driver distraction classification using CBAM embedded VGG16 architecture," *Computers in Biology and Medicine*, vol. 180, p. 108945, 2024, https://doi.org/10.1016/j.compbiomed.2024.108945.

[56] M. M. Ali, F. Maqsood, S. Liu, W. Hou, L. Zhang, and Z. Wang, "Enhancing Breast Cancer Diagnosis with Channel-Wise Attention Mechanisms in Deep Learning," *Computers, Materials & Continua*, vol. 77, no. 3, pp. 2699-2714, 2023, https://doi.org/10.32604/cmc.2023.045310.

[57] Z. A. Khan *et al*., "EA-CNN: Enhanced attention-CNN with explainable AI for fruit and vegetable classification," *Heliyon*, vol. 10, no. 23, p. e40820, 2024, https://doi.org/10.1016/j.heliyon.2024.e40820.

[58] C. Sitaula and M. B. Hossain, "Attention-based VGG-16 model for COVID-19 chest X-ray image classification," *Applied Intelligence*, vol. 51, no. 5, pp. 2850-2863, 2021, https://doi.org/10.1007/s10489-020-02055-x.

[59] N. Drissi, H. El-Kassabi, and M. A. Serhani, "A multi-criteria decision analysis framework for evaluating deep learning models in healthcare research," *Decision Analytics Journal*, vol. 13, p. 100523, 2024, https://doi.org/10.1016/j.dajour.2024.100523.

[60] D. Valero-Carreras, J. Alcaraz, and M. Landete, "Comparing two SVM models through different metrics based on the confusion matrix," *Computers & Operations Research*, vol. 152, p. 106131, 2023, https://doi.org/10.1016/j.cor.2022.106131.

[61] A. Inés, C. Domínguez, J. Heras, E. Mata, and V. Pascual, "Biomedical image classification made easier thanks to transfer and semi-supervised learning," *Computer Methods and Programs in Biomedicine*, vol. 198, p. 105782, 2021, https://doi.org/10.1016/j.cmpb.2020.105782.

[62] C. Gu and M. Lee, "Deep Transfer Learning Using Real-World Image Features for Medical Image Classification, with a Case Study on Pneumonia X-ray Images," *Bioengineering*, vol. 11, no. 4, p. 406, 2024, https://doi.org/10.3390/bioengineering11040406.

[63] A. A. Mukhlif, B. Al-Khateeb, and M. A. Mohammed, "An extensive review of state-of-the-art transfer learning techniques used in medical imaging: Open issues and challenges," *Journal of Intelligent Systems*, vol. 31, no. 1, pp. 1085-1111, 2022, https://doi.org/10.1515/jisys-2022-0198.

[64] B. M. Hussein and S. M. Shareef, "An Empirical Study on the Correlation between Early Stopping Patience and Epochs in Deep Learning," *ITM Web of Conferences*, vol. 64, p. 01003, 2024, https://doi.org/10.1051/itmconf/20246401003.

[65] A. Rezaeezade and L. Batina, "Regularizers to the rescue: fighting overfitting in deep learning-based side-channel analysis," *Journal of Cryptographic Engineering*, vol. 14, no. 4, pp. 609-629, 2024, https://doi.org/10.1007/s13389-024-00361-5.