JOURNAL OF SOFT COMPUTING AND DATA MINING
e-ISSN: 2716-621X

JSCDM

Vol. 6 No. 3 (2025) 200-218
https://publisher.uthm.edu.my/ojs/index.php/jscdm

Multi-Classes of Microorganisms Classification Using Hybrid Deep Stacked Autoencoder

Marwa T. Albayati[1]*, Mohd Ezanee Rusli[2,3], Moamin A. Mahmoud[2,3], Muhammed Ibrahim[1], Salam Omar Alo[4]

[1] *College of Graduate Studies, Universiti Tenaga Nasional, Kajang 43000, MALAYSIA*
[2] *College of Computing and Informatics, Universiti Tenaga Nasional, Kajang 43000, MALAYSIA*
[3] *Institute of Informatics and Computing in Energy, Universiti Tenaga Nasional, Kajang 43000, MALAYSIA*
[4] *Department of Artificial Intelligence Engineering Techniques College of Technical Engineering, Alnoor University, Mosul 41012, Nineveh, IRAQ*

*Corresponding Author: compmmtt@gmail.com
DOI: https://doi.org/10.30880/jscdm.2025.06.03.014

Article Info

Received: 13 August 2025
Accepted: 7 November 2025
Available online: 30 December 2025

Keywords

Convolutional neural network, artificial neural network, recurrent neural network, auto-encoder-decoder, long short-term memory, bidirectional long short-term memory

Abstract

Classification of microorganisms is critical for clinical and medical research; however, it remains challenging due to two key limitations: the limited availability of datasets of microscope images of microorganisms and the high morphological similarity between different classes. These limitations often result in misclassification and poor accuracy. To address these issues, the study proposes a hybrid model that combines a convolutional network and recurrent neural networks to extract more efficient features. This study evaluated the model on benchmark datasets. The model achieved 86% accuracy on the large-scale dataset and 97% accuracy on the small-scale dataset. This raises the probability of overfitting. For this, we utilized regularization and data augmentation techniques to reduce this risk. The main contribution is the development of a hybrid architecture that combines CNN, Bi-LSTM, and LSTM to capture bidirectional temporal dependencies and spatial sequential patterns, and compares the results between the proposed model and baseline model, which used CNN, BILSTM, LSTM, CNN-BILSTM, and CNN-LSTM. This integration of the proposed model, which enhances feature extraction and provides a more robust learning compared to conventional CNN–LSTM or CNN–BiLSTM approaches. These results emphasize the model's potential to enhance diagnostic processes, reduce time and labor, and support global health efforts by alleviating the workload on biologists and specialists.

1. Introduction

Microorganisms play a crucial role in the ecosystem, contributing to environmental monitoring, waste decomposition, wastewater management, and food processing. However, some are pathogenic, causing severe diseases in humans, plants, and animals. These include tuberculosis, anthrax, plague, toxoplasmosis, HIV, and COVID-19, the latter caused by SARS-CoV-2 and responsible for over 623,000 deaths globally. Consequently, microorganism classification is essential in fields like clinical microbiology, agriculture, medicine, and food

science. Traditional culture-based identification methods are laborious, expensive, and time-consuming [1], underscoring the need for automated classification using computer vision. In recent years, deep learning has made significant advancements in fields such as data mining, artificial intelligence, and pattern recognition [2]. Deep learning models, particularly neural networks with multiple layers, extract and represent complex data features at various abstraction levels [3]. Several challenges in the classification of microorganisms, such as the overfitting caused by the limited dataset and the importance of using the augmentation methods, shaped the design of the proposed model by integrating CNN with BILSTM and LSTM. This integration strengthens the feature extraction to address these limitations. This study introduced data augmentation and regularization methods, such as dropout, which improved the ability of the model to generalize. That prior research has not integrated autoencoders in multi-class microorganism classification.

However, a common challenge is overfitting, where a model memorizes the training data and fails to generalize to new inputs. One solution is to increase the dataset size, though this is not always feasible, especially with medical images. Data augmentation addresses this by generating synthetic samples through transformations like rotation, translation, shear, flipping, and adjusting brightness or contrast. However, excessive reduction can cause underfitting. Another strategy is to reduce the number of model parameters to limit overfitting; however, excessive reduction can cause underfitting, where the model performs poorly on both training and unseen data. Dropout regularization is also used, where neurons are randomly deactivated during training to prevent over-reliance on specific features. While dropout requires longer training and is more effective in fully connected layers than in convolutional layers, it enhances model robustness. The appropriate optimization techniques lead to training speed improvement and decrease overfitting. There is the existence of medical images in high-dimensional forms, making it necessary to use methods for dimensionality reduction like Principal Component Analysis (PCA) as employed in multiple studies in the literature [4].

Rasool and Abdel Amir [5] proposed a hybrid Convolutional Neural Network–Long Short-Term Memory (CNN-LSTM) model for classifying items in the COCO dataset, including animals, objects, and vehicles. Their approach for dimensionality reduction and feature extraction used histogram equalization and PCA with Histogram of Oriented Gradients (HOG) and Gray-Level Co-occurrence Matrix (GLCM). CNNs extracted spatial features, while LSTM networks captured sequential data patterns. As shown in the study by Sudha and Kumar [6], a hybrid CNN-LSTM was employed to classify heart disease data into normal and abnormal categories. To select relevant features, a Support Vector Machine (SVM) was used to address missing values with their preprocessing, combined with normalization. For loss minimization, they used backpropagation, and for model validation, k-fold cross-validation was utilized. El-Sayed [7] presents a hybrid model based on CNN-LSTM for the detection of Parkinson's disease. Significant features were extracted using HOG descriptors, and data augmentation techniques such as color adjustments, cropping, and flipping were employed to boost the training. Across the two datasets, the model indicated high classification accuracy.

Muttaqin et al. [8] developed a CNN-Recurrent Neural Network (RNN) model to identify bacteria that are harmful in drinking water. The CNN component extracted spatial features from microscopic images, while the RNN module analyzed temporal changes in bacterial growth patterns. Detection accuracy was enhanced by combining these techniques, indicating the effectiveness of integrating LSTM with CNNs or RNNs across multiple medical and environmental classification tasks. Ranjan Sahoo and Chakraborty [9] employed a neural network architecture that combines Bidirectional Long Short-Term Memory (Bi-LSTM), 2D CNN, and 3D CNN. The 3D CNN captures joint spatial-spectral characteristics, while the 2D CNN focuses on spatial characteristics. Bi-LSTM performs classification using outputs from the 2D CNN. Based on this combination, the hyperparameters and model complexity were reduced. Building on this, an autoencoder-based hybrid CNN-LSTM model was proposed, integrating recurrent neural networks with deep convolutional layers. A component of CNN operates within the structure of a modified DenseNet-201 autoencoder. Results indicated a 7.12% enhancement in classification accuracy over the standalone DenseNet, with a 17% growth compared to traditional configurations. The model stacks LSTM, CNN, and Bi-LSTM around an autoencoder framework. While minimizing noise, CNN-autoencoder effectively extracts and compresses spatial characteristics, providing clean, high-quality inputs to the LSTM and Bi-LSTM layers for sequential analysis [10].

Without sacrificing key information, an autoencoder also enhances computational efficiency by reducing input dimensionality. Even on small datasets, the hybrid model performs well, showing a strong generalization across multiple microorganism classes [11]. The challenges in microorganism classification were addressed using a hybrid architecture by combining LSTMs' sequential learning with CNNs' spatial feature extraction. Both efficiency and noise suppression were enhanced by the autoencoder, resulting in a model that ensures efficient, accurate, and robust classification. Its strong performance across microbial datasets positions it as a valuable tool in medical diagnostics and global health. Across limited data scenarios, the approach was integrated to enable effective feature extraction, spatiotemporal analysis, and generalization.

This study aims to categorize microorganisms into four distinct groups using a robust dataset of microscopic images. A hybrid architecture is designed by integrating CNNs, LSTMs, Bi-LSTMs, and an autoencoder using the unique advantages of each component. This proposed CNN-Bi-LSTM-LSTM model is evaluated for its classification

performance and generalization on an unseen publicly available dataset. Figure 1 presents a multilayer neural network showing that input data is passed into the hidden layer. Outputs are conveyed as input to the second hidden layer, which allows the network to learn more complicated patterns. Finally, an activation function was used to pass the results, producing nonlinearity, allowing learning more features. Finally, the network predicts the outputs for classification. This strategy effectively minimizes overfitting and achieves a 97% accuracy rate, outperforming existing classification techniques. The contribution of this study is:

- The research developed a hybrid model which combines the Convolutional Neural Networks (CNN), Bidirectional Long Short-Term Memory (Bi-LSTM), and Long Short-Term Memory (LSTM)
- An autoencoder is applied to obtain more representation learning and efficient accuracy
- Data augmentation and regularization strategies were added to reduce overfitting and address the limitations of the dataset.

The related work section outlines the architecture design of various deep learning models and their corresponding performance results. Section three presents the method needed to complete this study. Section four views the results and discusses the analysis of the results. The conclusion is addressed in section five, the final section.

2. Related Work

Previous studies applied CNN with RNN network for bacterial classification to achieve accurate classification, but still suffer from a limited dataset and a poor ability to capture long dependencies effectively. Similarly, CNN–LSTM studies the authors applied in neurodegenerative disease research to develop the learning dependencies, but they did not use the mechanisms for robust feature extraction, making them less effective when datasets are small or highly imbalanced. These limitations highlighted the need for more comprehensive work. The proposed hybrid model handled these gaps by using an integrating CNN with both Bi-LSTM and LSTM, along with an autoencoder, to reduce overfitting and enhance the classification accuracy of multiple microorganism classes

The related work in this area is extensive, featuring numerous advanced solution models that aim to enhance the classification performance of deep learning algorithms. Zhang et al. [12] employed image augmentation to expand their dataset and utilized various deep learning CNN models, including ResNet18, ResNet50, DenseNet121, DenseNet169, EfficientNet-b0, b5, b7, and ResNeXt101_32x8d and 32x16d, to classify slit-lamp images from patients with infectious keratitis. The conditions included Bacterial Keratitis (BK), Fungal Keratitis (FK), Herpes Simplex Keratitis (HSK), and Acanthamoeba Keratitis (AK). The study reported misclassifications between BK and FK, with classification accuracies of 70.27% for BK, 77.71% for FK, 83.81% for AK, and 79.31% for HSK.

The study of Shen et al. [13] indicated that high accuracy and reported misclassification were not achieved in two data categories. In the domain of radar image recognition, synthetic aperture radar (SAR) imagery challenges often emerge due to the scarcity of training samples and the inherent complications of imaging conditions. To address these weaknesses, they introduced a target recognition framework that integrates constrained naive generative adversarial networks (GANs) with convolutional neural networks (CNNs). By adopting a shallow CNN architecture, the model reduces computational complexity and mitigates overfitting, a common issue in deep learning when training data is limited. This hybrid approach improved recognition performance while effectively managing data scarcity and mitigating the effects of speckle noise in SAR images.

Kholerdi et al. [14] developed a self-adaptive multi-classifier system designed for environments with limited training samples. Their model combines SVM, neural networks, and Naive Bayes, automatically selecting the most reliable classifier for each instance based on confidence measures. This adaptive mechanism successfully minimized the decline in predictive accuracy often associated with insufficient datasets, providing comprehensive results across varied experimental conditions. In the field of facial recognition and medical image analysis CNNs are broadly identified for their ability to learn complicated patterns. They rely on substantial labeled data, which frequently results in overfitting and reduced classification accuracy when applied to small-scale datasets. To address these shortcomings, a hybrid methodology that integrates conventional machine learning techniques with deep learning was proposed. Based on this approach, in the early layers, CNNs were used for feature extraction, and subsequently, the extracted features are categorized using traditional supervised algorithms. Based on small-sample scenarios, this combination not only reduces overfitting but also consistently outperforms stand-alone machine learning models.

The application of CNNs in Raman spectroscopy was examined by Tewes et al. [15] for the purpose of classifying fungal spores and carotenoid-pigmented microorganisms. Depending strongly on classification accuracy, both the quality and volume of training data were emphasized in their study, strengthening the wider observation that dataset adequacy remains a critical determinant of CNN performance across various fields. When datasets lack diversity or fail to represent the full range of microorganism types, model performance can decline. To prevent overfitting, the authors employed cross-validation. However, the results, based on specific

microorganism types, may not generalize well to other categories or contexts, limiting the broader applicability and robustness of the predictive models.

Nneji et al. [16] employed CNN, CECED Inception-V3, and MobileNet3 architectures to classify COVID-19 cases. While the study reported promising results, its performance metrics were dataset-specific, suggesting that the model's effectiveness may vary when applied to different data sources. Wang et al. [17] utilized CNNs combined with Raman spectroscopy to classify Arcobacter bacteria species, though the study was limited to a single microbial class due to data constraints. Misclassifications were most frequent in A. faecis (89.6%) and A. skirrowii (90.4%), with 9.6% of A. skirrowii misclassified as A. butzleri. Most incorrect A. faecis predictions were assigned to A. butzleri, A. cibarius, and A. molluscorum. Nonetheless, 15 out of 18 species achieved classification accuracies above 95%. Other studies explored hybrid approaches.

Mansour et al. [18] proposed a model using Inception v4 with AdaGrad for feature extraction and a Variational Autoencoder (VAE) for classification. The autoencoder effectively reduced feature redundancy and misclassifications. However, the model's limitation lies in its evaluation solely based on chest X-ray data, with no assessment of other imaging modalities, such as CT scans, which affects its generalizability. Dev et al. [19] and Madhu [20] also explored hybrid deep learning models for disease detection. [20] proposed a CNN for feature extraction combined with cascaded RNN classifiers, including GRU, LSTM, and Bi-LSTM units. Their CNN-LSTM-Bi-LSTM model achieved 96.20% accuracy, with a 2.23% type I error rate and a combined 3.80% type I and II error rate, outperforming other tested models in both accuracy and computational efficiency. Table 1 shows the comparison of related work.

Table 1 *A comparison of related work*

Authors	Dataset	Method	Accuracy	Limitations
[12]	Slit-lamp images	CNN(ResNeXt101_32x16d+ DenseNet169)	70.27% for BK, 77.71% for FK, 83.81% for AK, and 79.31% for HSK.	An imbalanced data set is poor in generalization
[14]	Iris flower dataset	NN-SVM-NB	94.25%	Small data set
[15]	Raman-micro spectroscopy images	CNN	98%	The study focuses on one class (fungus) and lacks generalization
[18]	Human skin images	CNN	95%	Small data set
[19]	Public data set	CNN-RNN	96.2%	Overfitting and the study focus on binary classification (infected, uninfected)

3. Methodology

3.1 Model Structure

The proposed model integrates four CNN layers, an autoencoder, two LSTM layers, and two Bi-LSTM layers to classify microorganisms from image datasets. This hybrid architecture combines multiple deep learning components to extract spatial features and learn temporal dependencies effectively. The research workflow consists of three phases: data preprocessing, model architecture design, and performance evaluation. The proposed model integrates CNNs and LSTM networks to classify bacteria, viruses, fungi, and parasites. CNN layers capture spatial features from high-dimensional images, while LSTM layers learn sequential dependencies. The inclusion of Bi-LSTM layers further enhances contextual learning by considering both past and future information. Autoencoder–decoder modules and normalization methods are employed to improve performance and reduce overfitting. To strengthen generalization, data augmentation with dropout layers was used to expand variability within the data. These approaches improve accuracy and robustness in microbial image classification.

3.1.1 Feature Extraction

To extract spatial features from multidimensional image data within its encoder, the model used CNNs. Hierarchical visual patterns were used for identifying convolutional layers like textures, edges, and shapes, which are significant for distinguishing microorganisms with slight morphological differences. By reducing spatial dimensions, pooling layers improve this process by enhancing computational efficiency and reducing overfitting

risks. The ability of CNNs to filter and convey spatial information facilitates the generation of a compact, detailed representation in the latent space. By processing complicated image data, CNNs have become an essential component.

3.1.2 Dimensionality Reduction and Reconstruction

The autoencoder serves a critical role in data reduction by producing low-dimensional latent representations that protect vital information. The dimensionality of high-resolution medical images was reduced by the encoder part, which addresses the challenges related to processing large-scale data. The decoder restores the input from the reduced representation by minimizing reconstruction loss, and the entire model is optimized. This compression process effectively removes noise and redundant data while preserving key diagnostic features. As a result, the autoencoder serves as an effective preprocessing tool for classification tasks in downstream models. It reduces input dimensionality while maintaining important spatial characteristics. Formally, let the encoder EE and decoder DD be parameterized as described by [21]:

$$z = E(x) = \sigma(W_e * x + b_e)\ (Encoder) \tag{1}$$

$$\hat{x} = D(x) = \sigma(W_d * x + b_d)\ (Encoder) \tag{2}$$

where x is the input, z is the latent representation, $\hat{x}$ is the reconstructed output, $*$ denotes convolution, σ is a non-linear activation (e.g., ReLU), and W_e, b_e, W_d, b_d, are learnable parameters. The reconstruction loss $\mathcal{L}_{recon}$, typically, Mean Squared Error (MSE) ensures fidelity:

$$\mathcal{L}_{recon} = \frac{1}{N}\sum_{i=1}^{N} \left|\left|x_i - D\big(E(xi)\big)\right|\right|^2 \tag{3}$$

3.1.3 Sequential Pattern Learning

LSTM networks are integrated to process the sequential information contained in the flattened latent features output from the encoder. LSTMs are particularly effective at learning temporal dependencies, which is essential for understanding structural variations and progression patterns in different types of microorganisms. By maintaining long-term dependencies and addressing the vanishing gradient issue, LSTMs improve the model's ability to recognize relationships and transitions across image features, ultimately boosting classification accuracy. The latent vectors {z_1, z_2, ..., z_T}, representing a sequence of T timesteps, are flattened and input into the LSTM layers. At each time step t, the LSTM cell updates its hidden state h_t and cell state c_t according to the formulation provided by [21].

$$f_t, i_t, o_t = \sigma\big(W_{f,i,o}[h_{t-1}, z_t] + b_{f,i,o}\big), \tag{4}$$

$$\tilde{c}_t = tanh(W_e[h_{t-1}, z_t] + b_e), \tag{5}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \tag{6}$$

$$h_t = o_t \odot \tanh(c_t) \tag{7}$$

where $f_t, i_t, and\ o_t$ denote the forget, input, and output gates, $\odot$denotes element-wise multiplication, and W, b are learnable weights

3.1.4 Enhanced Contextual Understanding

Bi-LSTM layers extend the functionality of standard LSTMs by processing data in both forward and backward directions, enabling the model to learn from past and future sequences simultaneously. This bidirectional structure enables a deeper understanding of temporal relationships, which is particularly beneficial in microorganism classification, where some visual features may depend on both prior and subsequent context. By capturing these complex dependencies, Bi-LSTMs enhance the model's ability to recognize subtle patterns, resulting in improved generalization and robustness across various and diverse datasets. A Bi-LSTM processes the sequence in both forward and backward directions, concatenating the outputs [22].

$$h_t^{Bi} = [\overrightarrow{h_t};\ \overleftarrow{h_t}], \tag{8}$$

Bi-LSTM can enhance context awareness for microorganisms with variable temporal dynamics.

1. **Regularized Classification Layers:** The Bi-LSTM outputs are passed through dense layers with L_2-regularization to mitigate overfitting:

$$\mathcal{L}_{reg} = \lambda \sum_{j=1}^{N} \omega_j^2 \tag{9}$$

where λ is the regularization strength and ω_j are layer weights. The final SoftMax layer computes class probabilities:

$$P(y = j|X) = \frac{\exp(z_j)}{\sum_{k=1}^{K} \exp(z_k)} \tag{10}$$

where z_j is the logit for class j, and K is the total number of classes.

2. **Regularized Classification Layers:** The total loss combines reconstruction, classification (cross-entropy), and regularization terms:

$$\mathcal{L}_{total} = \alpha \mathcal{L}_{recon} + \beta \mathcal{L}_{class} + \gamma \mathcal{L}_{reg} \tag{11}$$

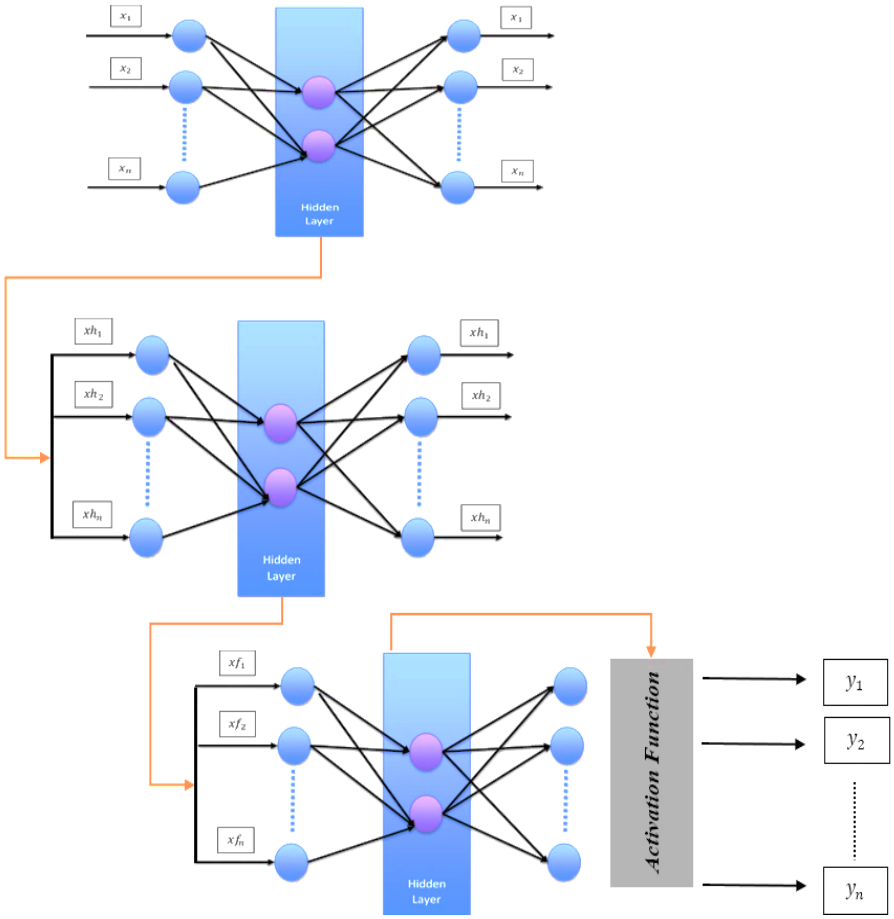

Fig. 1 *Structure of the proposed stacked autoencoder for a microorganism classifier*

3.2 Dataset

3.2.1 Dataset Preparation

The dataset used in this study was compiled from four primary sources: The Virus Texture Dataset [23], the Digital Images of Bacteria Species (DIBaS) dataset [23], the Digital Images of Fungus Species (DIFaS) database [25], and the Parasite Species Dataset [26]. A detailed overview of each dataset is provided below:

1. Virus texture dataset: The dataset includes transmission electron microscopy (also called TEM images for 15 different virus types. The texture patterns have been automatically derived from objects. The following is a concise summary of dataset characteristics [23].
    - There are 15 different texture classes, as seen in Fig.2.
    - There are 100 distinct texture patches for each class.
    - The dimensions of the texture patch are 41x41 pixels.
    - The file format is a lossless compressed 16-bit PNG.
    - The naming convention for files is as follows: The file name is class-003-sample-036.tif, with "class-003" representing the class number and "sample-036" representing the sample number.
2. Bacteria: The DIBaS consists of 32 bacterial species, with 20 images for each species. The Chair of Microbiology collected this dataset at Jagiellonian University in Krakow, Poland. All samples were stained using Gram's method. The images were captured with an Olympus CX31 Upright Biological Microscope with a SC30 camera (Olympus Corporation, Japan). The evaluation was conducted using a 100x objective lens under oil immersion (Nikon, Japan).
3. Fungus: The DIFaS database is a collection of microscopic images depicting various species of fungi. The system helps identify and sort fungi through deep learning methods and was specifically designed for this purpose. The database comprises 180 fungal strain images organized by multiple preparations of each strain and multiple images. The images within this database have a resolution of 3600 x 5760 pixels, utilizing a 16-bit intensity scale. The DIFaS database offers valuable support to mycological diagnostics by reducing both the time requirements and expenses associated with conventional biochemical identification systems [25].
4. Parasites: Mendeley Data hosts the Microscopic Images of Parasite Species dataset containing a total of 34,298 microscopic images that display different parasites alongside host cells. The dataset contains parasite images that include Plasmodium, together with Toxoplasma gondii, Babesia, Leishmania, and Trypanosoma, as well as Trichomonas, along with host cell images of Red Blood Cells (RBCs) and leukocytes. The imaging process occurred at magnification levels of 400x and 1000x [26]. Fig. 2 illustrates the samples of microorganism types.

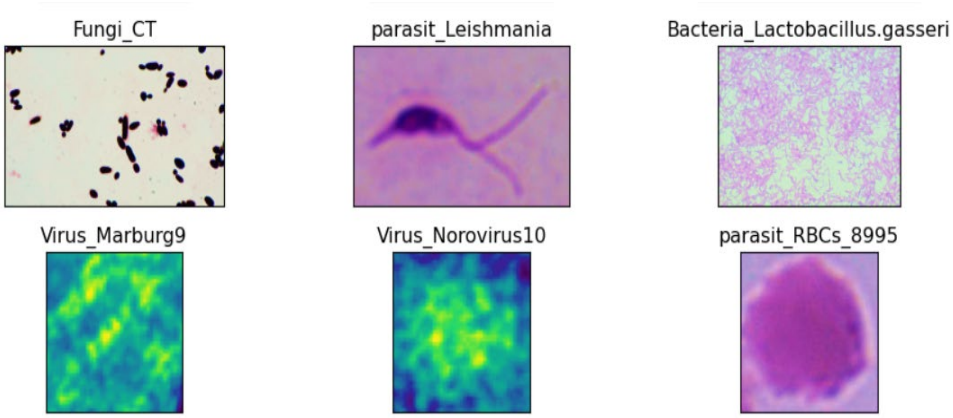

Fig. 1 *Samples of microorganism types*

3.2.2 Dataset Analysis

The analysis process facilitates efficient computation and offers clear interpretations of the relationships among various classes within the dataset. This approach supports numerical assessment of dataset similarity, which is valuable for tasks such as clustering, anomaly detection, and validating pre-trained models. The procedure is executed through the following implementation steps:

1. Feature Extraction: The dataset's inter-class relationships are quantified through a feature extraction process that begins with the pre-trained VGG16 CNN [27], followed by feature aggregation at the class level and pairwise similarity computation. The method allows researchers to analyze high-dimensional semantic patterns and their class-based distributions effectively.
    - Pre-processing: Input images are resized to 224 x 224 pixels and normalized to match the VGG16 training distribution (ImageNet).

- Model Adaptation: The classification layer of VGG16 is removed, and a Global Average Pooling (GAP) layer is appended to the final convolutional block. This produces a 512-dimensional feature vector $f_i \in R^{512}$ for each image i, capturing high-level spatial semantics.
- Feature Freezing: The CNN weights remain frozen during extraction, ensuring computational efficiency and leveraging pre-trained representations without fine-tuning.

2. Class-Level Features: Using the average feature vector for each class, we reduce noise and represent class-specific characteristics effectively. For each class c, a representative feature vector F_c is computed by averaging the feature vectors of all N_c images within the class:

$$F_c = \frac{1}{N_c}\sum_{i=1}^{N_c} f_i \tag{12}$$

where f_i denotes the feature vector of the ii-th image in class cc. This approach mitigates the impact of outliers and noise while preserving the central tendency of the class distribution. The mean is chosen over alternatives (e.g., median) due to its computational efficiency and smooth differentiability, which aligns with downstream applications like gradient-based optimization.

3. Pairwise Class Similarity Computation: The similarity between two classes c_1 and c_2 is quantified using cosine similarity, which measures the angular alignment of their aggregated feature vectors F_{c_1} and F_{c_2}:

$$Similarity(c_1, c_2) = \frac{F_{c_1}.F_{c_2}}{||F_{c_1}||\,||F_{c_2}||} \tag{13}$$

Since feature vectors are non-negative due to rectified linear unit ReLU activations in VGG16, similarity scores range between 0 (orthogonal, maximally dissimilar) and 1 (identical direction, maximally similar). A symmetric (K×K) similarity matrix S is constructed for all K classes, where diagonal entries $S_{ii} = 1$.

3.2.3 Data Preprocessing (Augmentation and Normalization)

This phase involves splitting the data into training, validation, and testing sets, followed by normalization. All images are resized to 128 × 128 pixels, and pixel values are normalized. Normalization and preprocessing are carried out using the MinMaxScaler and fit functions from the scikit-learn library. To address class imbalances commonly found in the datasets, data augmentation techniques are applied. These include random rotations, height and width shifts, zooming, and flipping. Such augmentation ensures the model is better trained in underrepresented classes. The original images are loaded, augmented, and the processed data is saved into a new directory to form the final augmented dataset. The number of original data includes 2893 images of all classes, in the proposed method, the data augmentation was applied utilizing on-the-fly techniques, which means the augmented images were generated during training, and the number of images on the disk did not change, the augmented data which are generated during training was 289.300 over 100 epochs and batch size 32, the advantages of this approach is Unlimited variety across epochs which led to better generalization, Zero extra disk space, and easy to bias augmentation toward minority classes without saving files. Table 2 shows the amount of data.

Table 2 *Sample of data*

Classes	Training	Validation	Testing
Bacteria_Acinetobacter	16	2	2
Bacteria_Actinomyces	18	2	2
Bacteria_Bacteroides	18	2	2
Bacteria_Bifidobacterium	18	2	2
Bacteria_Candida	16	2	2
Bacteria_Clostridium	17	2	3
Bacteria_Enterococcus	32	4	4
Bacteria_Escherichia	16	2	2
Bacteria_Fusobacterium	18	2	2
Bacteria_Lactobacillus.casei	16	2	2
Bacteria_Lactobacillus.crispatus	16	2	2
Bacteria_Lactobacillus.delbrueckii	16	2	2
Bacteria_Lactobacillus.gasseri	16	2	2

J. of Soft Computing and Data Mining Vol. 6 No. 3 (2025) p. 200-218

Classes	Training	Validation	Testing
Bacteria_Veionella	17	3	2
Fungi CA	32	4	4
Fungi_SB	32	4	4
Fungi_SC	32	4	4
Virus_Adenovirus1	80	10	10
Virus_Astrovirus2	80	10	10
Virus_CCHF3	80	10	10
Virus_Cowpox4	80	10	10
Virus_Dengue5	80	10	10
Virus_Ebola6	80	10	10
Virus_Influenza7	80	10	10
parasit_Toxoplasma_400X	32	4	4
parasit_Trichomonad	32	4	4

3.2.4 Experimental Setting

For training and testing the model, a workstation has an AMD Threadripper CPU 3990x of 64 cores and 128 threads and 64GB DDR4 RAM. The operating system is Windows 11, and the programming language is Python. The Python environment is generated using Anaconda, which is a distribution of the Python and R programming languages for scientific computing. Table 3 shows the parameters for training.

Table 3 *Parameters for training*

Name	Parameters
Operating system	windows 11
Processor	AMD Threadripper CPU 3990x of 64 cores and 128 threads
Ram	64GB DDR4
Programming language	Python
Development environment	Jupyter 7.0.8
Loss	sparse categorical cross-entropy loss
Input dimensions	224x224
Batch size	64
Activation function	Relu
Epochs	100
Optimizer	Adam
Filter of CNN	(128,64)
Units of BILSTM	256 units
Units of LSTM	128 units

3.2.5 Performance Evaluation

Precision, recall, F1-score, and accuracy are widely used metrics for evaluating classification performance and are adopted in this study to assess the proposed model. Precision is defined as the ratio of true positive (TP) predictions to all predicted positive cases, indicating the proportion of predicted positives that are actually correct. Recall, also known as sensitivity, measures the proportion of actual positive cases correctly identified by the model, calculated as the ratio of TP to all actual positives. Accuracy reflects the overall effectiveness of the model by calculating the ratio of correctly classified samples, both true positives (TP) and true negatives (TN), to the total number of samples. In this context, TP refers to correctly predicted positive samples, while TN denotes correctly identified negative samples. The evaluation metrics are computed using the following formulas, as referenced in [28]:

$$\text{Precision} = \frac{TP}{TP+FP} \tag{14}$$

$$\text{Recall (Sensitivity)} = \frac{TP}{TP+FN} \quad (15)$$

$$\text{F1-Score} = 2*\frac{Precision*Recall}{Precision+Recall} \quad (16)$$

$$\text{Accuracy} = \frac{\text{TP+TN}}{TP+TN+FP+FN} \quad (17)$$

4. Results and Discussion

4.1 Dataset Analysis Result

The pairwise data analysis results are visually represented in the similarity heat map shown in Fig. 3. This heat map is generated from a full cosine similarity matrix, computed using features extracted via the VGG16 model. It offers valuable insights into the relationships among the microbial classes, including bacteria, fungi, parasites, and viruses, within the dataset. Key findings and their implications are summarized as follows:

1. Intra-Domain Similarities
    - Bacterial Classes: High intra-bacterial similarities are observed (e.g., Bacteria_Acinetobacter vs. Bacteria_Escherichia: 0.924, Bacteria_Neisseria.gonorrhoeae vs. Bacteria_Pseudomonas.aeruginosa: 0.928). This suggests shared spatial features (e.g., rod-like morphology, Gram-negative staining). Lactobacillus subspecies (e.g., Lactobacillus casei vs. Lactobacillus johnsonii: 0.939) indicate tight clustering, likely due to conserved structural traits.
    - Fungal Classes: Fungi exhibit strong intra-group similarities (e.g., Fungi_CA vs. Fungi_SC: 0.924, Fungi_CL vs. Fungi_MF: 0.783), reflecting shared hyphal or yeast-like morphologies.
    - Viral Classes: High similarity among enteric viruses (e.g., Virus_Rotavirus14 vs. Virus_Norovirus10: 0.977) suggests overlapping imaging characteristics, potentially due to similar capsid structures.
    - Parasites: Bloodborne parasites (e.g., parasit_Plasmodium_843 vs. parasit_Babesia: 0.916) cluster closely, likely due to erythrocyte invasion mechanisms visible in imaging.
2. Inter-Domain Relationships
    - Bacteria vs. Fungi: Moderate overlaps (e.g., Bacteria_Staphylococcus.aureus vs. Fungi_CA: 0.696) may stem from shared staining properties or biofilm structures.
    - Bacteria vs. Parasites: Bacteria_Veionella vs. parasit_Trichomonad: 0.675 hints at potential morphological overlaps (e.g., flagellar structures).
    - Viruses vs. Other Domains: Viruses generally exhibit low similarity with bacteria/fungi (e.g., Virus_Adenovirus1 vs. Bacteria_Escherichia: 0.361), aligning with their distinct structural biology.
3. Outliers and Anomalies
    - Unexpected Bacterial Similarities: Bacteria_Staphylococcus.aureus vs. Bacteria_Micrococcus.spp: 0.958—unusually high for distinct genera, possibly due to cocci morphology or imaging artifacts.
    - Parasite-Bacteria Overlaps: Parasit_Leukocyte_1000X vs. Bacteria_Listeria.monocytogenes: 0.637—may reflect leukocyte invasion by Listeria in images.
4. Taxonomic Validation
    - Consistent Genus/Phylum Groupings: Lactobacillus subspecies (casei, reuteri, rhamnosus) form a coherent cluster (similarities > 0.800), validating taxonomic relationships. Streptococcus vs. Staphylococcus: Lower similarity (0.759) aligns with their distinct Gram-positive morphologies.
    - Viral Family Clustering: Virus_Lassa8 vs. Virus_Marburg9: 0.903—reflects shared filamentous structures in Arenaviridae and Filoviridae.
5. Implications for Model Performance
    - High Similarity Classes: Pairs like Bacteria_Clostridium vs. Bacteria_Enterococcus (0.900) may lead to misclassification; targeted data augmentation or attention mechanisms could improve distinction.
    - Low Similarity Classes: Viruses (e.g., Virus_CCHF3 vs. Virus_West Nil15: 0.988) indicate near-perfect self-similarity but low cross-domain scores, suggesting robust feature extraction for viruses. The high similarity between these classes makes it difficult for the model to distinguish between the classes because of the overlap between them, which leads to misclassification among classes. Figure 3 shows the similarity heatmap distribution provides some pros because it represents the visual analysis of similarity between classes, such as the similarity of bacteria groups, which are the same classes. Conversely, the proposed model can differentiate between classes based on their distinct shapes, such as bacteria versus viruses, indicating that the model can generalize across various classes. For instance, bacterial classes such as Lactobacillus and Streptococcus exhibit high inter-class similarity because of their morphological differences.

J. of Soft Computing and Data Mining Vol. 6 No. 3 (2025) p. 200-218

Fig. 2 *Similarity of dataset classes results using a heat-map distribution*

6. Visualization of Key Patterns

- Heatmap Clusters:
  Block 1: Bacterial families (e.g., Enterobacteriaceae, Lactobacillaceae).
  Block 2: Fungal species (e.g., Candida, Aspergillus).
  Block 3: Parasites and viruses form distinct off-diagonal clusters.
- Dendrogram Analysis: Hierarchical clustering would likely group Staphylococcus spp., Lactobacillus spp., and blood parasites into distinct clades.

Thus, the analysis of the dataset can demonstrate the following aspects:

1. Feature Discrimination: The model robustly separates domains (bacteria/fungi vs. viruses) yet struggles with morphologically similar subclasses (e.g., Clostridium vs. Enterococcus).
2. Taxonomic Alignment: High similarities within genera/phyla validate the biological relevance of extracted features.
3. Model Refinement: Address high-similarity pairs via contrastive learning or hybrid architectures (e.g., CNNs + Transformers).

4.2 Dataset Preprocessing Result

The initial dataset consisted of 2,893 input images, following data augmentation. The augmented data was evenly distributed across all classes, as illustrated in Fig. 4.

Fig. 3 *The distribution of data in the proposed dataset before and after augmentation. (a) before data augmentation; (b) after data augmentation*

J. of Soft Computing and Data Mining Vol. 6 No. 3 (2025) p. 200-218

4.3 Model Evaluation Result

The experiments detailed in this section were performed using the proposed dataset, which was split into training and testing sets. The model's performance was assessed using several evaluation metrics. The training progress is illustrated through the loss and accuracy curves shown in Figs. 5 and 6, respectively.

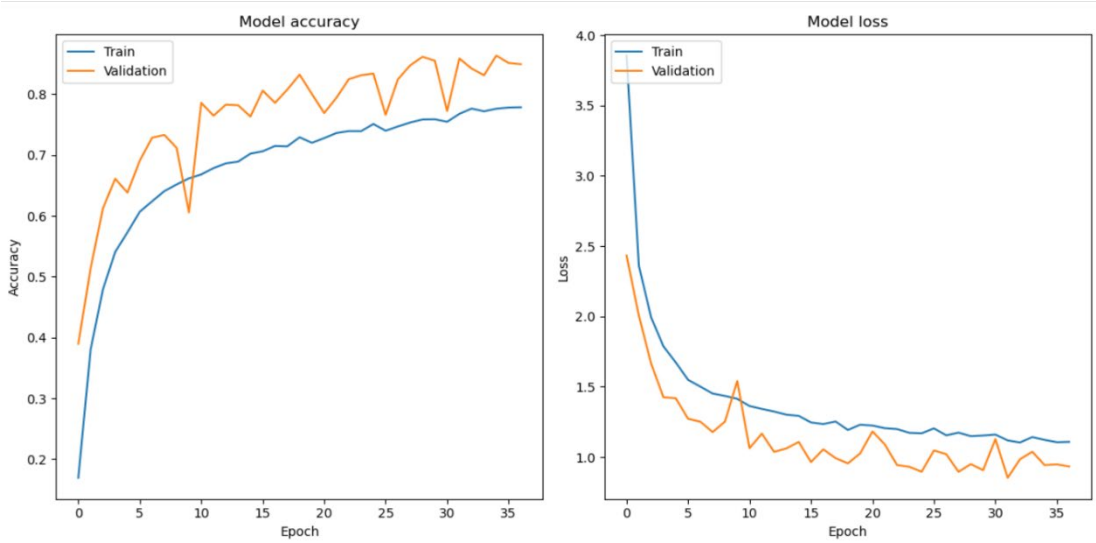

Fig. 4 *Training and validation accuracy*

Fig. 5 presents the training and validation accuracy, along with the corresponding loss curves, for the proposed classification model under the defined optimizer settings. The model achieves an overall accuracy of 86%, indicating strong performance across bacterial, fungal, viral, and parasitic categories. The model achieves an F1-score of 0.84, indicating effective classification across most classes, although some variations persist. The model performs exceptionally well in identifying parasitic species such as Babesia, Leishmania, and Plasmodium, as well as fungal species like Candida glabrata (CG), Candida albicans (CL), Candida paralytica (CP), and Candida tropicalis (CT), achieving perfect precision, recall, and F1-scores. This highlights both strong dataset representation and the presence of distinct, recognizable features in these classes. However, challenges persist with certain bacterial and viral species. Neisseria gonorrhea was not classified, likely due to limited representation or high feature overlap with other bacteria. Similarly, Pseudomonas aeruginosa demonstrated poor performance, with a precision of 0.23 and a recall of 0.50, indicating difficulty in distinguishing it from similar species. Among viral classes, Dengue, Lassa, Marburg, and West Nile viruses reported low recall scores, indicating frequent misclassification due to overlapping features in the latent space. Fig. 6 displays the classification results for 15 test samples across the four microbial classes (bacteria, fungi, parasites, and viruses). The model correctly classified 11 of these samples, while four were misclassified, reflecting both its strengths and remaining limitations in handling diverse microorganism images. The results showed a decrease in accuracy due to the larger dataset, high variety, and similarity among certain classes, all of which led to misclassifications.

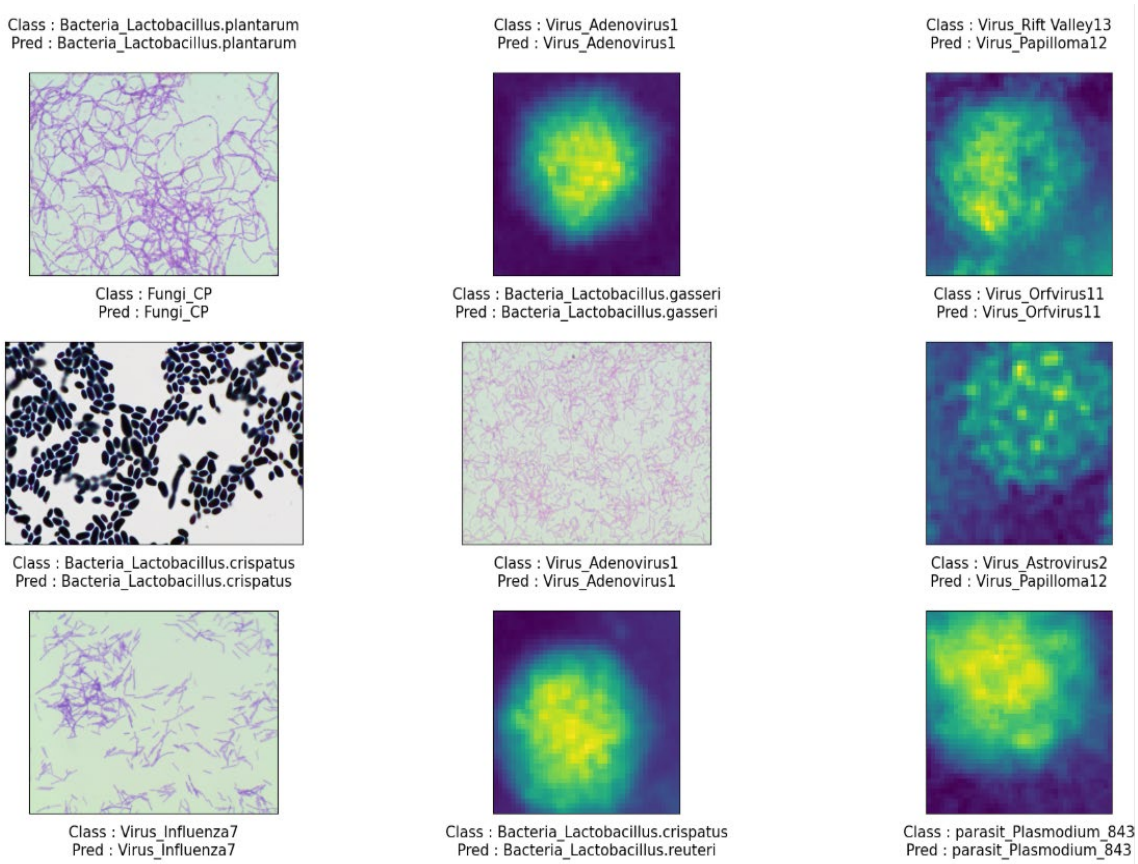

Fig. 6 *Examples of misclassifications across microbial classes*

4.4 Performance Comparison with Other Models

The proposed model's effectiveness was validated by benchmarking its performance against several state-of-the-art machine learning models. All models were evaluated using consistent performance metrics, precision, recall, F1-score, and accuracy on the test dataset. The classification outcomes for these models are summarized in Table 4, which presents the average metric values across all categories. A detailed comparative analysis was also performed on an external benchmark dataset, EMDS-6 [29], to assess the proposed model's generalizability. The results demonstrated that the proposed hybrid model significantly outperforms existing approaches. Specifically, it achieved 97% accuracy on EMDS-6, compared to just 67.62% reported by [30] on the same dataset. The earlier study utilized only CNN-based architectures to determine the best-performing model. In contrast, this study integrates CNNs with LSTM and Bi-LSTM layers, enhancing its ability to capture both spatial and temporal features. For instance, MobileNetV2 achieved an accuracy of 83.9%, while a study [7] using HOG features with CNN and LSTM reported 81% accuracy in a Parkinson's disease classification task. [31] proposed a CNN-BiLSTM model for wildfire prediction, achieving a relatively low precision of 0.62%. Another comparative method, the CViT model with a stacked technique for surface electromyography (sEMG) classification [32], reached an accuracy of 80.02%. [33] used a CNN-LSTM framework for COVID-19 classification, achieving 67.7% accuracy. The comparative results, as highlighted in Tables 4 and 5, demonstrate the superiority of the proposed model in terms of both accuracy and robustness across various domains and datasets. Notably, it maintained strong generalization across both large-scale datasets and smaller ones, such as EMDS-6, which contains only 2,240 images. This adaptability confirms the model's potential for real-world applications where data diversity and volume may vary significantly.

J. of Soft Computing and Data Mining Vol. 6 No. 3 (2025) p. 200-218

Table 4 *Comparison using the EMSD-6 dataset*

Class	Recall (%)	Precision (%)	F1 score (%)
Amoeba	100	97	99
Euglena	98	98	98
Hydra	97	100	99
Paramecium	98	100	99
Rod Bacteria	100	91	96
Spherical Bacteria	100	92	96
Spiral Bacteria	97	100	99
Yeast	81	96	88
Accuracy	97%		

As shown in Table 4, when trained on a small dataset, a high accuracy of 97% was achieved by the model as indicated in this study [30]. However, lower accuracy (86%) was achieved by the proposed model when trained on a larger dataset, as shown in Table 5. The reduction in accuracy can be connected to the higher variety and complication of the large data, making it more difficult to obtain the features in the images. Despite the accuracy reduction, this shows the capability of the model to handle complicated and multiple data by enhancing its ability to generalize across various data. Compared with the large data, the high accuracy identified with the small data is due to the complexity being lower, which provides critical challenges when working with large data.

Table 5 *Existing models versus proposed model*

Ref.	Model	Precision	Recall	FI-score	Accuracy
[28]	EMViTmodel 0.6953	0.6762		0.6689	0.6762
[29]	MobilNetv2 model 0.833	0.857		0.845	0.8390
[31]	CviT model			___	0.8002
[30]	CNN+LSTM model 0.62% 0.66%			0.64%	___
[6]	Histogram of oriented gradients CNN+LSTM	___	___	____	0.8100
[32]	Hybrid CNN+LSTM model	___	___	____	0.6770
	Proposed model 0.85	0.86		0.84	0.8600

Table 5 presents studies [7], [29], and [33] that used small data, such as the EMD-6 data. Study [29] employed the VIT model, which showed lower accuracy than other models due to the limited dataset size, leading to overfitting. Research [28] addressed the limitation of the study made by [30] by introducing EMVIT, a model that combines CNN with MLP, to overcome the weaknesses of the VIT model. Research [33] applied CNN-LSTM with an Autoencoder to reduce image dimensions and remove noise. CNN was used to extract fine spatial features, while LSTM processed the temporal sequence of images. One of the drawbacks of the study [33] was overfitting, which was again due to the use of small datasets. Study [7] utilized a CNN with the ResNet-50 model to extract important spatial features and an LSTM to capture temporal features. A key challenge in this research was the reliance on small datasets. Another study [32] utilized the CViT model to extract spatial features from sEMG, employing a stacked technique for classifying surface electromyography (sEMG). The proposed model used a larger dataset and achieved an accuracy of 86%, outperforming previous studies. Table 6 compares the proposed model with the machine learning algorithm, showing a superiority in accuracy compared to the machine learning algorithm. This indicates the ability of the proposed model to adapt to complex data.

Table 6 *Comparison between machine learning algorithms and the proposed model*

Machine learning classifier	Precision	Recall	F1 score	Accuracy
K-Nearest Neighbor	0.44	0.27	0.19	0.25
Support Vector Machine (SVM)	0.58	0,58	0.57	0.59
Decision Tree (DT)	0.14	0.20	0.14	0.21
Stochastic Gradient Descent (SGD)	0.23	0.16	0.13	0.16
Logistic Regression (LR)	0.44	0.44	0.44	0.45
eXtreme Gradient Boosting (XGBoost)	0.46	0.43	0.43	0.44
Proposed model	0.86	0.85	0.84	0.86

4.5 Discussion

The classification of microorganisms, including viruses, bacteria, fungi, and parasites, is crucial due to their role in causing various infectious diseases. However, this process is often time-intensive and resource-demanding. Researchers used multiple approaches to mitigate these challenges, including CNNs and traditional machine learning models such as SVMs and Decision Trees (DTs) [34]. Regardless of these initiatives, limitations such as insufficient datasets and poor application within various microbial classes continue to restrict performance. The current study provides a hybrid deep learning model that was designed for microorganisms' classification.

The compiled dataset consists of images of 32 bacterial species, 15 viral species, nine fungal species, and nine parasitic species, all sourced from various sources. Within the techniques for augmentation of the data, as a result, reducing class imbalance and improving the robustness of the model. The architecture of the model comprises a four-layer CNN used to retrieve salient image features. To identify temporal trends and relevant relationships a Bi-LSTM network is executed twice in progression. This is succeeded by two standard LSTM layers that use past sequential relationships to refine the process of classification.

To manage the medical image data's high dimensionality, an encryption-decryption-based reduction mechanism was used. This method reduces data volume while preserving significant features, thus reducing differences between the input and the images reconstructed. Table4 presents the model's performance, indicating that it achieved a precision of 86%, a recall of 85%, an F1-score of 84%, and an overall accuracy of 86%. Additionally, its generalization ability was tested on a small-scale dataset, EMDS-6. The proposed model achieved a 97% accuracy on this dataset, in contrast to the 67.62% reported in a related study by [28], further validating its efficacy on datasets of varying sizes. For a broader comparison, Table 5 highlights the performance of several machine learning algorithms on the same task. K-Nearest Neighbors (KNN) yielded an accuracy of 0.25%, SVM achieved 0.59%, Decision Trees achieved 21%, Stochastic Gradient Descent (SGD) achieved 16%, Logistic Regression achieved 45%, and XGBoost achieved 44% in comparison with the proposed model, which significantly outperformed existing methods by achieving an accuracy of 86%, confirming a higher classification capability. The study compares the performance of the foundation models CNN, CNN-LSTM, and CNN-BiLSTM. CNN achieved an accuracy of 58%, CNN-BiLSTM 59%, and CNN-LSTM 37%.

Despite the proposed model classification's general effectiveness, a significant issue lies in the disparity between precision and recall for specific species of microbes. The model attains a recall of 1.00 for Proteus bacteria but only a precision of 0.58, indicating that the model regularly misclassifies the species as Proteus. Papillomavirus attains a recall of 1.00 and a precision of 0.69, showing an inclination towards overprediction. These differences result in enhanced false positive (FP) rates that are specifically about applications of medical diagnostics, in which wrong extraction can lead to improper decisions of the treatment. Addressing these limitations might require adopting a range of strategies to mitigate them. Improving the data through extra data augmentation and class balancing methods might enhance the representation of neglected species like Pseudomonas aeruginosa and Neisseria gonorrhea, hence model bias reduction.

Feature extraction, refining, or integrating dimensionality mitigation methods may improve the ability of the model to differentiate between classes with corresponding characteristics. The improvements of the model-level may also add to advanced classification balance. These include modifying decision thresholds by applying weighting for classes to penalize incorrect categorization of minority classes and analyzing ensemble learning strategies. For viral species, the enhancements are particularly critical where FN might lead to missed diagnoses and the spread of the disease. The proposed model shows a strong classification performance within a wider range of microbial categories, targeting refinements, especially for specific viral and bacterial species, are critical. Improved data representation and strategic model enhancements present promising avenues for enhancing the precision–recall balance and increasing the model's effectiveness in clinical and real-world diagnostic settings.

5. Conclusion and Future Work

This study presents a hybrid model of classification of microorganisms that integrates CNN-BIKSTM-LSTM and an autoencoder. The main contributions include using data augmentation and regularization techniques such as dropout to reduce overfitting and making an evaluation of the benchmark. The results showed that the proposed model achieved high accuracy when handling a large data set and similarity between different classes of microorganisms compared with baseline CNN, CNN-BILSTM, CNN-LSTM, BILSTM, and LSTM. Its robust generalization capabilities suggest potential applicability to other image-based microbiological datasets. The model's rapid inference capability surpasses traditional classification methods such as culturing and manual microscopic analysis, offering a more efficient and accurate diagnostic alternative. This advancement could be valuable in medical settings, where timely and precise identification of microbial agents is critical for determining appropriate treatment and preventing incorrect diagnoses that may endanger patient outcomes. However, there are some limitations of the proposed model that have not yet been validated in a real-world clinical setting. For future work, several directions can be pursued to enhance the study. One important aspect is incorporating explainable AI (Grad-CAM, SHAP), validating on clinical datasets, or exploring transformer-based architectures. Develop a user interface for clinicians for microscope image classification.

Acknowledgement

Universiti Tenaga Nasional (UNITEN) supports the research.

Conflict of Interest

The authors declare that they have no conflict of interest regarding the publication of this paper.

Author Contribution

The authors confirm contribution to the paper as follows: ***study conception and design:*** *Marwa T. Albayati, Moamin A Mahmoud;* ***data collection:*** *Marwa T. Albayati, Mohd Ezanee Rusli;* ***analysis and interpretation of results:*** *Marwa T. Albayati, Muhammed Ibrahim, Salam Omar Alo;* ***draft manuscript preparation:*** *Marwa T. Albayati, Muhammed Ibrahim. All authors reviewed the results and approved the final version of the manuscript.*

References

[1] Rani, P., Kotwal, S., Manhas, J., Sharma, V., & Sharma, S. (2022). Machine learning and deep learning-based computational approaches in automatic microorganisms image recognition: Methodologies, challenges, and developments. Springer Science and Business Media B.V. https://doi.org/10.1007/s11831-021-09639-x

[2] [2] Ibrahim, M., & Mahmoud, M. A. (2025). Enhancing Smart Grid Stability Using AI Techniques : A Systematic Literature Review. 2025 21st IEEE International Colloquium on Signal Processing & Its Applications (CSPA), February, 50–55. https://doi.org/10.1109/CSPA64953.2025.10933390

[3] Yassin, N. I. R., Omran, S., El Houby, E. M. F., & Allam, H. (2018). Machine learning techniques for breast cancer computer-aided diagnosis using different image modalities: A systematic review. Computer Methods and Programs in Biomedicine, 156, 25–45. https://doi.org/10.1016/j.cmpb.2017.12.012

[4] Zhu, Z. (2024). Systematic optimization of overfitting problem in machine learning. Highlights in Science, Engineering and Technology, 111, 353–359. https://doi.org/10.54097/3TKZRJ84

[5] Rasool, Z. H., & Abdel Amir, M. A. (2024). Comprehensive image classification using hybrid CNN-LSTM model with advanced feature extraction on COCO dataset. Mustansiriyah Journal of Pure and Applied Sciences, 2(2), 28–47.

[6] Sudha, V. K., & Kumar, D. (2023). Hybrid CNN and LSTM network for heart disease prediction. SN Computer Science, 4(2). https://doi.org/10.1007/S42979-022-01598-9

[7] El-Sayed, R. S. (2023). A hybrid CNN-LSTM deep learning model for classification of the Parkinson disease. IAENG International Journal of Applied Mathematics, 53(4).

[8] Muttaqin, K., Fadillah, N., Fajri, R., & Mursyidah, M. (2025). Optimized CNN-RNN architecture for rapid and accurate identification of hazardous bacteria in water samples. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4961159

[9] Ranjan Sahoo, A., & Chakraborty, P. (2024). Hybrid CNN Bi-LSTM neural network for hyperspectral image classification.

[10] Priyadarshini, I. (2025). An explainable autoencoder-based feature extraction combined with CNN-LSTM-PSO model for improved predictive maintenance. Computers, Materials & Continua, 83(1).

[11] Mohseni, P., & Ghorbani, A. (2024). Exploring the synergy of artificial intelligence in microbiology: Advancements, challenges, and future prospects. Computational and Structural Biotechnology Reports, 100005.

[12] Zhang, Z., Wang, H., Wang, S., Wei, Z., Zhang, Y., Wang, Z., & Liang, Q. (2022). Deep learning-based classification of infectious keratitis on slit-lamp images. Therapeutic Advances in Chronic Disease, 13. https://doi.org/10.1177/20406223221136071

[13] Shen, S., Wang, X., Mao, F., Sun, L., & Gu, M. (2022). Movements classification through sEMG with convolutional vision transformer and stacking ensemble learning. IEEE Sensors Journal, 22(13), 13318–13325. https://doi.org/10.1109/JSEN.2022.3179535

[14] Kholerdi, H. A., Taherinejad, N., & Jantsch, A. (2018). Enhancement of classification of small data sets using self-awareness – An Iris Flower case-study. Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS), 2018-May. https://doi.org/10.1109/ISCAS.2018.8350992

[15] Tewes, T. J., Welle, M. C., Hetjens, B. T., Tipatet, K. S., Pavlov, S., Platte, F., & Bockmühl, D. P. (2023). Understanding Raman spectral-based classifications with CNNs using examples of fungal spores and carotenoid-pigmented microorganisms. AI, 4(1), 114–127. https://doi.org/10.3390/ai4010006

[16] Nneji, G. U., Cai, J., Jianhua, D., Chikwendu, I. A., Oluwasanmi, A., James, E. C., & Mgbejime, G. T. (2021). Enhancing low-quality radiograph datasets using wavelet transform CNN and GAN for COVID-19 identification. 2021 4th International Conference on Pattern Recognition and Artificial Intelligence (PRAI), 146–151. IEEE. https://doi.org/10.1109/PRAI53619.2021.9551043

[17] Wang, K., Chen, L., Ma, X., Ma, L., Chou, K. C., Cao, Y., & Lu, X. (2020). Arcobacter identification and species determination using Raman spectroscopy combined with neural networks. Applied and Environmental Microbiology, 86(20), 1–16. https://doi.org/10.1128/AEM.00924-20

[18] Mansour, R. F., Escorcia-Gutierrez, J., Gamarra, M., Gupta, D., Castillo, O., & Kumar, S. (2021). Unsupervised deep learning based variational autoencoder model for COVID-19 diagnosis and classification. Pattern Recognition Letters, 151, 267–274. https://doi.org/10.1016/j.patrec.2021.08.018

[19] Dev, A., Fouda, M. M., Kerby, L., & Md Fadlullah, Z. (2024). Advancing malaria identification from microscopic blood smears using hybrid deep learning frameworks. IEEE Access, 12, 71705–71715. https://doi.org/10.1109/ACCESS.2024.3402442

[20] Madhu, G., VNS, P. J., & Karthik, A. S. (2025, February). Hybrid deep neural network for classification of malaria parasites in microscopic images. 2025 International Conference on Artificial Intelligence and Data Engineering (AIDE) (pp. 253–261). IEEE.

[21] He, W., Li, J., Tang, Z., Wu, B., Luan, H., Chen, C., & Liang, H. (2020). A novel hybrid CNN-LSTM scheme for nitrogen oxide emission prediction in FCC unit. Mathematical Problems in Engineering, 2020, 8071810. https://doi.org/10.1155/2020/8071810

[22] Anki, P., Bustamam, A., Al-Ash, H. S., & Sarwinda, D. (2020). High accuracy conversational AI chatbot using deep recurrent neural networks based on BiLSTM model. 2020 3rd International Conference on Information and Communications Technology (ICOIACT). IEEE.

[23] Kylberg, G., Uppström, M., Hedlund, K. O., Borgefors, G., & Sintorn, I. M. (2012). Segmentation of virus particle candidates in transmission electron microscopy images. Journal of Microscopy, 245(2), 140–147. https://doi.org/10.1111/j.1365-2818.2011.03556.x

[24] Zieliński, B., Plichta, A., Misztal, K., Spurek, P., Brzychczy-Włoch, M., & Ochońska, D. (2017). Deep learning approach to bacterial colony classification. PLoS ONE, 12(9), e0184554. https://doi.org/10.1371/journal.pone.0184554

[25] Zieliski, B., Sroka-Oleksiak, A., Rymarczyk, D., Piekarczyk, A., & Brzychczy-Woch, M. (2020). Deep learning approach to describe and classify fungi microscopic images. PLoS ONE, 15(6), e0234806. https://doi.org/10.1371/journal.pone.0234806

[26] Li, S., Yang, Q., Jiang, H., Cortés-Vecino, J. A., & Zhang, Y. (2020). Parasitologist-level classification of apicomplexan parasites and host cell with deep cycle transfer learning (DCTL). Bioinformatics, 36(16), 4498–4505. https://doi.org/10.1093/bioinformatics/btaa513

[27] Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. http://www.robots.ox.ac.uk/

[28] YA, M. A., Yaw, C. T., Koh, S. P., Tiong, S. K., Chen, C. P., Yusaf, T., & Raj, A. A. (2023). Detection of corona faults in switchgear using 1D-CNN, LSTM, and 1D-CNN-LSTM methods. Sensors, 23(6), 1–19. https://doi.org/10.3390/s23063108

[29] Dwivedi, K., Dutta, M. K., & Pandey, J. P. (2024). EMViT-Net: A novel transformer-based network utilizing CNN and multilayer perceptron for the classification of environmental microorganisms using microscopic images. Ecological Informatics, 79, 102451. https://doi.org/10.1016/j.ecoinf.2023.102451

[30] Marjani, M., Mahdianpari, M., & Mohammadimanesh, F. (2021). A comparative study of deep learning classification methods on a small environmental microorganism image dataset (EMDS-6): From convolutional neural networks to visual transformers. arXiv. http://arxiv.org/abs/2107.07699

[31] Marjani, M., Mahdianpari, M., & Mohammadimanesh, F. (2024). CNN-BiLSTM: A novel deep learning model for near-real-time daily wildfire spread prediction. Remote Sensing, 16(8). https://doi.org/10.3390/rs16081467

[32] Shen, S., Wang, X., Mao, F., Sun, L., & Gu, M. (2022). Movements classification through sEMG with convolutional vision transformer and stacking ensemble learning. IEEE Sensors Journal, 22(13), 13318–13325. https://doi.org/10.1109/JSEN.2022.3179535

[33] Dastider, A., Sadik, F., & Fattah, S. A. (2021). An integrated autoencoder-based hybrid CNN-LSTM model for COVID-19 severity prediction from lung ultrasound. Computers in Biology and Medicine, 132, 104296. https://doi.org/10.1016/j.compbiomed.2021.104296

[34] Zhou, H. Y., Li, Y., Li, J., Meng, J., & Wu, A. (2025). Unleashing the potential of artificial intelligence in infectious diseases. National Science Review, 12(3), nwaf004.