Journal of Intelligent Systems and Internet of Things
Vol. 18, No. 02, PP. 341-360, 2026
DOI: https://doi.org/10.54216/JISIoT.180224

A Systematic Review on Classification Techniques of Microorganisms: Challenges and Recommendations – Towards Medical Intelligent Systems

Marwa T. Albayati[1,*], Mohd Ezanee Bin Rusli[1], Moamin A. Mahmoud[1], Aws A. Abdulsahib[2], Mohammed F. Alomari[2], Sallar S. Murad[1]

[1]College of Computing and Informatics, University Tenaga Nasional (UNITEN), Selangor, Malaysia

[2]Dijla University College, Computer Science Department, Baghdad, Iraq

Emails: compmmtt@gmail.com; Ezanee@uniten.edu.my; Moamin@uniten.edu.my; aws.abdulkareem@duc.edu.iq; mohamed.fadhil@iku.edu.iq; sallarmurad@gmail.com

Abstract

Microorganisms are commonly found in our daily living environments and play a crucial role in environmental pollution control, disease prevention, and treatment, as well as food and drug production. To fully utilize the diverse functions of microorganisms, their analysis is essential using Intelligent Systems. Traditional analysis methods can be labor- intensive and time-consuming. As a result, image analysis using Intelligent Systems i.e. machine learning or deep learning have been introduced to improve efficiency. Deep learning networks algorithms such as CNN contain a stack of multi-layer, the first layer and the last are the input and output layers, between them are the hidden layers to extract and learn many features in images, recurrent network algorithms (RNN) combined with convolution neural network (CNN), these networks allow to process a series of images to extract the crucial information from images and also these algorithms help to minimize the size of images and reduce the redundancy in microrganisms images According to previous studies, these algorithms are the most used to classify the images of microorganisms. However, the classification of microorganism images presents several challenges these include the need for robust algorithms due to varying application contexts, the presence of insignificant features, along various analysis tasks that need to be addressed. The research summarizes significant advancements that tackle these challenges through deep learning and machine learning methods. Current obstacles, gaps in knowledge, unresolved issues, limitations, and difficulties in classification techniques are also discussed.

Received: April 10, 2025 Revised: June 21, 2025 Accepted: August 19, 2025

Keywords: Microorganisms; Deep learning; Machine learning; Classification techniques

1. Introduction

Microbes, containing bacteria, yeasts, and molds, play significant roles in human life, particularly in biotechnology, food sciences, medicine, and genetic engineering [1]. While these microbes are necessary for producing antibiotics, hormones, and other therapeutic compounds, they can also present risks to human health and industry through food spoilage and disease transmission [2]. In the United States, there are approximately 48 million cases of foodborne illnesses each year, resulting in around 128,000 hospitalizations and 3,000 deaths. Poor water hygiene contributes significantly to 1.7 million deaths globally each year, with most of these occurring in children in developing countries. Intelligent systems and the Internet of Things monitor the safety of food and water, identify harmful microbiological, and inform people immediately to take action before any risks may cause damage. The pathogens accountable for these illnesses and fatalities include Campylobacter jejuni, enterotoxigenic Escherichia coli, Shigella spp., Vibrio cholerae, Aeromonas spp., en-terotoxigenic Bacteroides fragilis, Clostridium difficile, and Cryptosporidium parvum [3].The genetic and biochemical characteristics of bacteria or viruses determine their potential use in biotechnological processes as well as their ability to cause disease. By characterizing, identifying, and taxonomically classifying biological materials, we can explore industrial applications and enhance treatment options for infections in the future. It is important to recognize that classify and systematics, although often utilized interchangeably, have distinct meanings. Systematics focuses on the diversity, relationships, and interactions among organisms, while taxonomy involves organizing organisms into hierarchical groups based on their common ancestry. Despite appearing similar, individual organisms within a group can exhibit phenotypic variation because of the genetic variation. The greater the genetic differences, the more distant the relationship between organisms. The Internet of Things monitors patients, vital signs like as temperature, heart rate, and oxygen levels. Clinics can take action to reduce any possible infections and eliminate the spread of illnesses if unusual signs emerge, such as a sudden fever or a decrease in oxygen levels. The hierarchical classification system includes division, class, family, genus, species, and strain. Research in classification, systematics, and detection of microorganisms is interconnected and mutually influential. Precise identification is crucial for the classification. Employing classical methods alongside 16S rRNA genes and other molecular markers is fundamental for microbial identification and classification [2]. Accurate microbial identification is vital in many fields, containing clinical microbiology and food production. Microbial identification can be done using both direct and indirect techniques. Manual classification is time-consuming and requires precision due to the growing volume of microbial data. Traditional methods are often costly and labor-intensive, leading to a gradual shift toward machine learning techniques in microbial research. Machine learning and deep learning algorithms such as KNN, SVM, and RF are being used for microorganism classification, though challenges such as image contrast and noise can lead to misclassification [4]. Figure 1 shows examples of different types of microorganisms. Rapid advancements in medical image classification have become increasingly crucial for doctors in diagnosing diseases. Despite previous studies in this area, significant gaps remain, such as data scarcity and poor classification performance. Additionally, there is a lack of diversity in the datasets and testing across diverse datasets. While studies like [8] have contributed to the field of medical image classification, they have not fully addressed issues like misclassification across various types and categories nor the problem of limited data. The review discusses various classification techniques and machine learning methods, including Support Vector Machines (SVM), Random Forest (RF), and deep learning methods such as Recurrent Neural Networks (RNN) to provide a comprehensive understanding of classification techniques towards Medical Intelligent Systems and internet of things using machine learning and deep learning techniques(1) also discussed the challenges faced by researchers and provided overall analysis of different methods and the influence of various ML techniques in microorganism image classification. Finally, highlights key features, challenges, and potential future research directions of the internet of things. This review addresses these gaps by summarizing current research and providing a clearer understanding of medical image classification and the challenges it faces. This will assist researchers in their future endeavors. Moreover, Microscopy modalities are used to collect medical information on small-scale biological objects that cannot be obtained through other means. There are two main kinds of microscopes: electron microscopy and light (optical) microscopy. Electron microscopes, including scanning and transmission electron microscopes, use a beam of electrons to examine the diagnostic object. Light (optical) microscopes, on the other hand, pass light rays through lenses to magnify the object and capture diagnostic medical images. Various techniques, such as Phase Contrast Microscopy (PCM), Bright field Microscopy (BM), fluorescence microscopy, and Dark field Microscopy (DFM), are examples of optical microscopy methods [12]. Figure 3 shows the types of medical images.

The operational strategies of these algorithms are discussed, highlighting the advantages and limitations of deep learning and machine learning algorithms in the context of the internet of things. This analysis helps in selecting the most relevant algorithms and classes for various methods of classification and provides a foundation for future enhancements. Other recent reviews are summarized in Table 1.

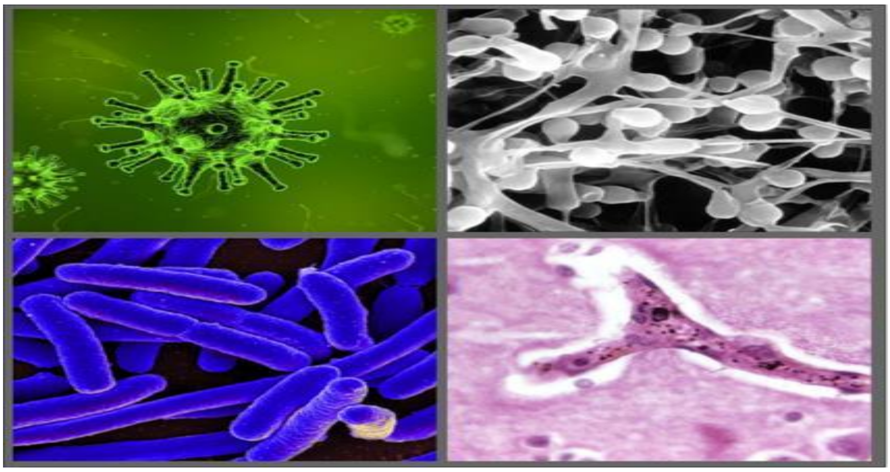

Figure 1. An example of microorganisms types (bacteria, viruses, fungus and parasites

Table 1: Comparison with recent reviews of classification techniques

Authors	Review Scope	Contributions	Recommendations
[5]	A review on image classification for MRI brain stroke lesion.	Developing stroke classification methods enables quicker and more accurate medical decisions and treatments.	• To examine MRI brain and classification algorithms for brain stroke. • To enable rapid diagnosis of stroke .
[6]	A survey on deep learning applied to medical images: from simple artificial neural networks to generative models.	Introducing generative models, such as Generative Adversarial Networks and Autoencoders, which can be used for data augmentation and improved classification algorithms.	Developing hybrid models could benefit from merging diffusion models with other generative techniques.
[7]	A survey of Short-Term Traffic Prediction Using Deep Learning Long Short-Term Memory: Taxonomy, Applications, Challenges, and Future Trends.	• The survey discusses the use of LSTM-based traffic forecasting algorithms developed in the last three years. It focuses on providing a detailed description of the algorithms, rather than just a superficial overview. • The algorithms are categorized based on techniques combined with LSTM, and a comparative analysis is conducted. • The survey also describes the key operations, advantages, and limitations of each algorithm, aiding in algorithm selection for specific road traffic management needs.	• Implementing noise removal techniques, such as filtering image data and adjusting data values to practical ranges, can improve prediction accuracy. • Utilizing transformer architecture permits parallel processing of time-series data, reducing training time and improving prediction accuracy. • Exploring transfer learning techniques can help adapt parameters and reduce computational time in complex traffic networks.
This review	A systematic review on Classification techniques of microorganisms: challenges and recommendations towards medical intelligent system	• Classification of various types of microbes. • Highlighted challenges of classification in machine learning and deep learning algorithms. • Comparison between recent reviews and article • Highlighted challenges of IoT and edge devices	• Using hybrid models to achieve high accuracy. • Developing big datasets for better model performance. • Integrate IoT with machine learning and deep learning to lead to intelligent system

2. Methodology

This review followed the PRISMA 2020 guidelines and conducted a systematic literature review (SLR) with the primary aim of minimizing bias to ensure the study's reliability and precision [9]. Systematic reviews serve several essential functions, such as addressing concerns that individual studies might overlook, identifying limitations in the existing research that need to be addressed in future studies, refining or testing theories about how and why phenomena occur, and providing detailed summaries of what we currently know in a specific domain to guide future research efforts [10].

a. Search strategy

This review conducted a comprehensive literature review using keyword searches on article titles, abstracts, and terms in three databases: Scopus and Web of Science. These databases were chosen because they include many high-impact and high-quality articles [11], and most high-quality articles from other databases are available in Scopus and Web of Science. We selected specific keywords to search these databases to identify relevant publications on the use of AI techniques for smart grid stability to meet our research objectives ((”Microorganism” or ”Microbe” or ”Microscopic organism” or ”Microbial life” or ”Microscopic life forms” or ”Microscopic organisms”) and (”Characteristics” or ”Attributes” or ”Traits” or ”Properties” or ”Qualities” or ”Characteristics and attributes” or ”Features and traits” or ”Microorganism traits and features”) and (”AI” or ”Machine learning” or ”Deep learning” or ”Neural networks” or ”Computational intelligence” or ”AI technologies” or ”Intelligent algorithms”) and (”Classification” or ”Categorization” or ”Taxonomy” or ”Grouping” or ”Sorting” or ”Organization” or ”Clustering” or ”Categorizing”)).

b. Study Selection Process

Figure 2 summarizes all the steps involved, including the criteria for inclusion and exclusion, as well as the stages of the process. To assess each article for various qualities, data was collected and extracted into an Excel spreadsheet and then used to list and categorize the articles. To find articles, the first step was to perform a keyword search in two databases. The initial search query yielded 275 articles from all databases. In the first round, 185 unrelated papers were excluded because they are not relevant to classification techniques. 95 articles not related to AI and edge devices of the Internet of Things. Following a full-text review of the remaining 90 papers, 78 relevant articles were identified based on title and abstract. Two duplicate articles were also removed, resulting in a final set of 76 articles. Subsequently, 56 articles were removed because they are not relevant to AI and the Internet of Things, resulting in a final set consisting of 20, which includes articles that integrated AI with IoT and edge devices

c. Inclusion and Exclusion Criteria

We carefully applied inclusion and exclusion criteria to ensure that the study remained relevant and effectively addressed our research objectives [9]. The criteria were established to select reviewed articles that are comprehensive and focus on using AI techniques. To be included in the study, an article must be written and published in the English language, released between 2019 and 2024, based on classification techniques, address one of the research questions, be attainable, and be a journal or conference paper with results. The review process excludes articles not published in English, those published before 2019, studies that do not involve AI techniques, duplicates, review articles, short articles, and articles with restricted access, as shown in Table 2.

d. Review Objectives and Research Questions

This systematic investigation aims to identify and estimate the algorithms and methods in deep learning to classify microorganisms. The algorithms of deep learning in classification will be evaluated. Some methods need further investigation to classify many types of microorganisms to get a more accurate classification. As shown in Table 3, research questions were to define the methods and algorithms of classification.

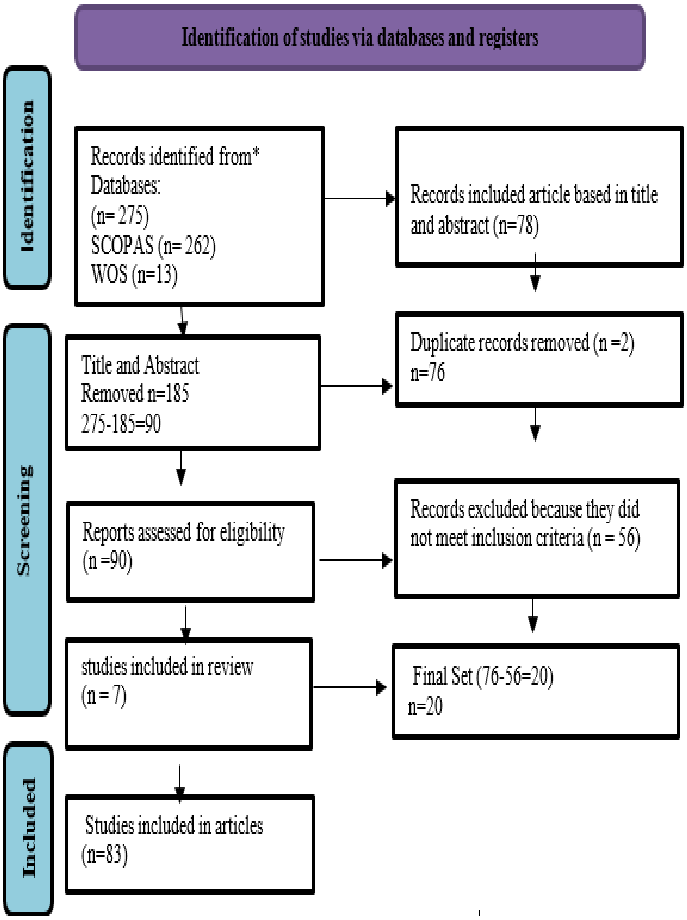

Figure 2. Inclusion and exclusion criteria

Table 2: Inclusion and Exclusion Criteria.

Inclusion Criteria	Exclusion Criteria
Articles published in English language	Articles not published in English
Study published between 2019–2024	Study published before 2019
Studies based on classification techniques	Studies without classification techniques, duplicate and review studies
Including articles	Short articles, duplicated, review articles
Articles that were accessed	Articles with restricted access

Table 3: Research Questions (RQs).

RQs	Research Questions
RQ1	What are the algorithms proposed in the literature review for the classification of microorganisms?
RQ2	What are the types of microorganisms classified using deep learning?
RQ3	What are the algorithms for more accurate classification of microorganisms?

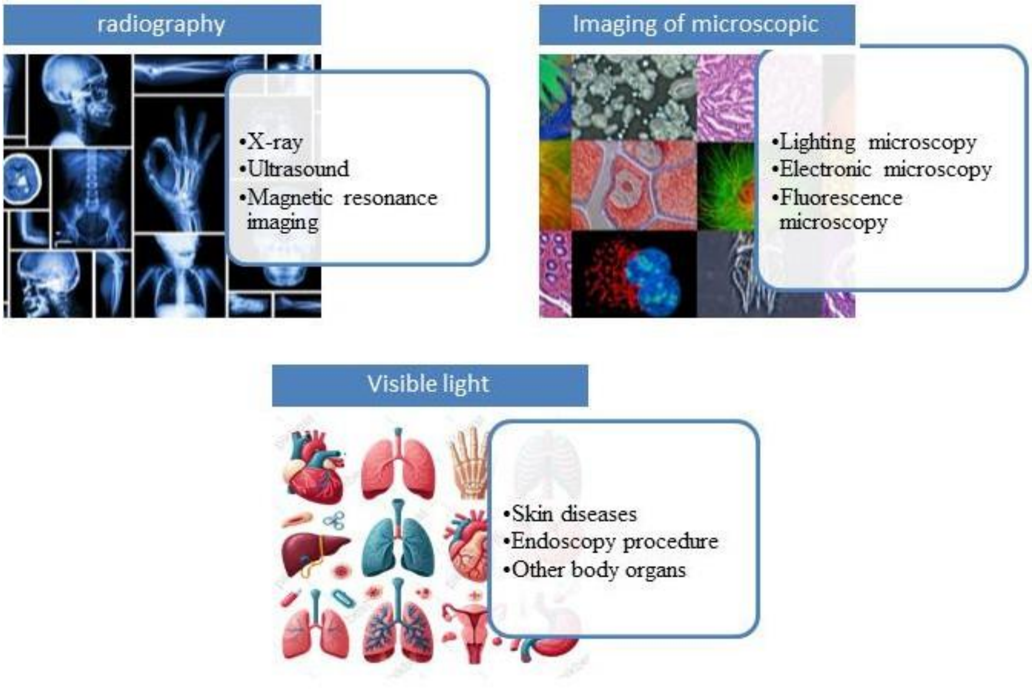

Figure 3. Types of medical images

3. Medical Images Classification studies

Machine learning in medical imaging has turned into a prominent topic in the field of medical image analysis [14] . [13] introduced a technique for automatically classifying CT brain images of various head trauma types using machine learning algorithms. Deep learning has demonstrated promising results in medical image analysis, as shown by [15] with their innovative deep learning model for detecting Alzheimer's disease using brain MRI data. [16] emphasized the importance of data augmentation in enhancing deep learning for image-classification issues. [17] explored the use of deep convolutional neural networks (CNN) for medical image classification, specifically focusing on pneumonia detection in chest X-ray datasets. Additionally, [18] discussed the integration of medical image information and robotic technology in medical surgical robots, highlighting the significance of grasp planning in multi-fingered manipulators. [19] provided a comprehensive review of deep learning models for retinal image analysis, showcasing the rapid advancement of computer vision-based techniques in medical image studying. In the context of medical diagnostic image analysis for COVID-19 detection, [20] proposed a dynamic fusion-based learning approach to enhance performance of model. Finally, [21] addressed the challenge of the limited availability of medical images for training deep neural networks by combining CNN features with SVM for the automated categorization of brain tumors from MRI images. Machine learning algorithms are growingly being used for the classification of bacteria and viruses. [22] introduced the Generative Adversarial Networks (GANs) in biological sequence analysis to address data imbalance issues. Automated bacterial classifications using machine learning can be classified into supervised learning, unsupervised learning, and reinforcement learning [23]. Machine learning techniques have been employed for genus-level classification of bacteria species [24]. A study here compared different pre-trained (CNN) architectures for bacteria classification, demonstrating the effectiveness of machine learning algorithms in microbial classification [25]. The use of machine learning in microbial species and community classification has made significant progress, especially in classification and clustering algorithms [26]. Additionally, machine learning has been used for the rapid and environmentally friendly classification of bacteria using NIR spectroscopy, along with multivariate analysis and machine learning algorithms [27]. ML algorithms have also been applied in bacteriophage research for the classification of phage genomes [28].. Table 4 Comparison review papers on classification. Additionally, medical images classification integrates with IoT devices (e.g., smart microscopes, real-time sensors) by these techniques

- Medical images of blood samples, tissue slides, or microorganisms are easily automatically captured using smart microscopes. These images can be provided directly to a categorization system (locally or in the cloud) via the Internet of Things for a more accurate diagnosis, real-time sensors can continually collect data such as patient vitals, and temperature and adapt it to medical images.
- Real-time sensors gather data regarding the patient's temperature and vital signs. To recognize diseases, identify pathogens, or monitor the effectiveness of treatment, deep learning or machine learning models evaluate the collected

images and related data. Healthcare providers in remote regions can obtain immediate feedback from IoT devices that provide results instantly, allowing prompt diagnosis and treatments without requiring samples to be provided by a distant laboratory. (2)

For all columns, the✓ symbol indicates that overall methods and challenges were proposed in the references, and the • symbol means that the authors in the reference showed their view for several aspects that are not covered at all in the references. The △ symbol means that this aspect is only mentioned briefly but not discussed comprehensively in the reference.

Table 4: A Comparison of Existing Reviews on Microorganisms Classification

Authors	**Key Technologies**	**Challenges**
[29]	△	△
[30]	△	✓
[31]	✓	•
[32]	•	△
[33]	•	•
This Review	✓	✓

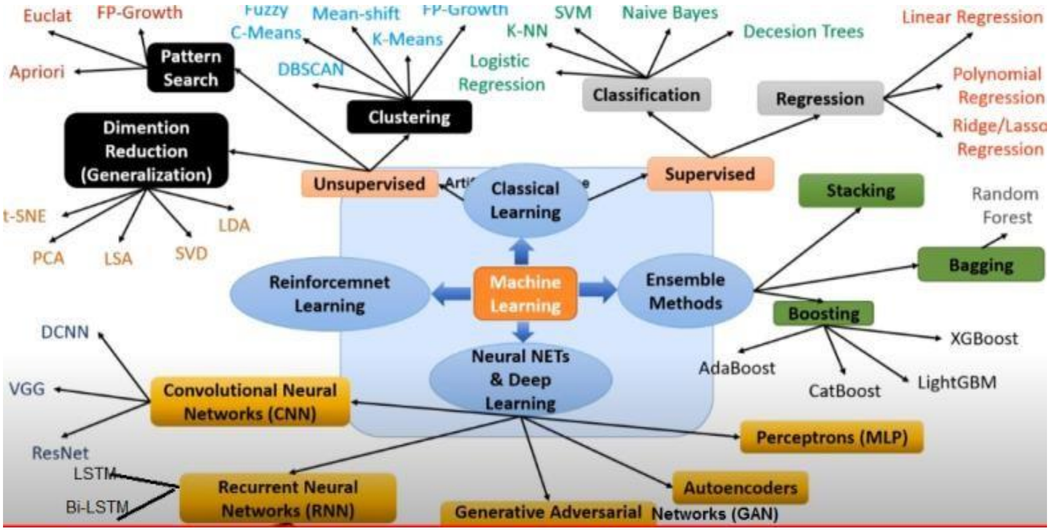

Figure 4. An Examples of techniques of classification

4. Classification in Machine Learning Techniques (remove paragraph)

- **Support Vector Machine Algorithm**

The study by [34] [35] aimed to classify pathogenic bacteria at the genus level using images captured by an optical sensor system. This research concentrated on identifying patterns in four common bacterial species. The first step involved enhancing image quality using the Contrast Limited Adaptive Histogram Equalization (CLAHE) method to generate clearer images of the bacteria. Next, features were extracted and trained utilizing the Sup- port Vector Machine (SVM) method, incorporating texture analysis and Zernike Moment Invariant (ZMI). The trained features were then classified utilizing two species of SVM kernels: linear and Radial Basis Function (RBF). The study focused on four pathogenic bacteria—Escherichia, Listeria, Salmonella, and Staphylococcus the results showed that utilizing the SVM with the RBF kernel led to superior classification performance compared to the linear kernel, achieving an accuracy of 90.33%. However, a drawback of this study was its use of only one type of microorganism. In contrast, the study by [36] introduced a new method to automate microorganism identification. This method proposed a segmentation mechanism using a convolution filter (Kirsch) and a pixel-clustering algorithm (Otsu) to extract microorganism bodies. Twenty-five features were selected to map microbe characteristics and morphology, with mutual information-based models showing the best performance for feature selection. Hyperparameter tuning was conducted for various machine-learning models, with the SVM radial kernel requiring further improvement for maximum accuracy. The study also faced issues with data bias due to abundance. A comparative analysis between SVM and an improved version (ISVM) showed that ISVM outperformed SVM in part of accuracy (98.2%), precision (98.2%), recall (98.1%), and F1 score (98.1%). Furthermore, the study [37] developed a computer-based method to recognize digital images of bacterial cells. This method can automatically identify twenty genera and species of bacteria. In the Support Vector Machine algorithm, the data is distributed based on the training of the model, but the data is not distributed within the algorithm randomly; all the data is used to get the optimal margin the nearest points are called support vectors only these points are saved .These methods of classification are summarized in Table 5, and Figure 5 shows SVM workflow for microorganism classification.

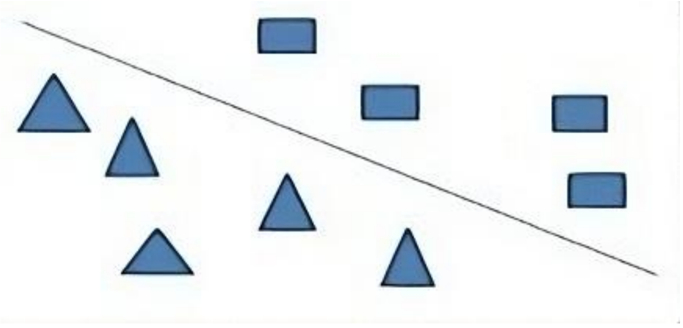

Figure 5. SVM workflow for microorganism classification

- **Decision Tree Algorithm**

Decision tree learning is a predictive modeling approach that utilizes a decision tree to relate a product to conclusions about the item's target value. This algorithm is commonly usage in statistics, data mining, and machine learning. The target is to create a model that detect the value of a target variable based on different input variables. Decision trees are rule-based decision tools widely employed in pattern recognition, offering an efficient training procedure and model construction [38]. The study by [39] focused on extracting features from MRI images, which is crucial for subsequent classification. The extracted features represent the characteristics of normal and abnormal tumors (both benign and malignant). Two classification algorithms were used: Naive Bayesian and Decision Tree. The Decision Tree classifier is particularly highlighted for its ability to classify images based on these features. However, a limitation of this paper is that although the Decision Tree classifier showed promising results, it may not be the best choice for all data types. Decision trees can may lead to overfitting, particularly with complex datasets, which could affect the model's generalizability. The results demonstrated an accuracy of 96% and a sensitivity of 93% for the Decision Tree algorithm. In contrast, this study [40] used a decision tree algorithm to detect dengue-infected patients based on blood smear images captured under microscope at a magnification of 400x. The study also focused on classifying dengue infections based on WBC characteristics, with the WBC classification model achieving 92.2% accuracy and the dengue classification model achieving 72.3% accuracy. Figure 6 illustrates the decision tree algorithm.

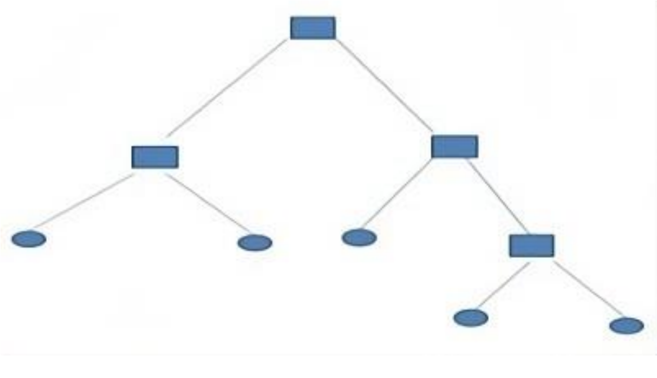

Figure 6. Dataset in Decision tree algorithm

- **Random Forest Algorithm**

The general idea behind this algorithm is to combine multiple decision trees into a single ensemble model. The expectation is that multiple models will perform better than a single model. The ’random’ aspect of the algorithm involves selecting input variables and observations from the example data to build each decision tree. This results in different trees producing different classification results. The outcome is the average of the results for regression and a majority-rules vote for classification. The advantage of this algorithm is that it addresses the limitations of individual decision trees and automatically identifies the importance of variables in solving the problem. The algorithm is relatively easy to use, as there are only two parameters to optimize: the number of trees and the number of independent variables used at each split. However, a disadvantage is that high prediction accuracy comes at the expense of explain ability [41]. The paper [41] presented two algorithms for skin disease classification: K-Nearest Neighbors (KNN) and Random Forest (RF). KNN classifies an input based on the ’k’ nearest data points in the feature space and determines the most common class among them. In this study, a challenge with KNN was its low accuracy due to the inappropriate choice of ’k’ and the lack of consideration for distance measures. Furthermore, RF algorithm selects a random subset of features from all available features, denoted as” k” from” m” total features. The algorithm then calculates the best split for the nodes based on parameters like entropy, which helps determine how the nodes branch during classification. The process involves generating daughter nodes and repeating the steps until the nodes reach a terminal state. The study found that RF algorithm achieved a testing accuracy of 94.22% and an F1 score of 95.94%, demonstrating its effectiveness in classifying skin diseases. The paper also compared the performance of both algorithms, noting that KNN had a testing accuracy of 95.23% and an F1 score of 95.98%, which was slightly higher than that of the Random Forest algorithm. Furthermore, the research by [42] presented a method for detecting and classifying bacteria of pneumonia in chest X-ray images using a decision tree algorithm. The study used a dataset of 5,856 chest X-ray images, including 4,273 cases of pneumonia and 1,583 normal cases. Figure 7 shows the Random Forest algorithm.

5. Classification in Deep Learning

- **Convolution Neural Network**

[43] used CNN to classify E. coli and V. cholera bacteria in this research. The results achieved 93.01% accuracy in classification and 97.0% in counting. The limitation of this study was that Gaussian noise enhancement decreased CNN accuracy. [44]. this study analyzed bacterial classification using deep learning. The results show that the bacterial identification system using the InceptionV3 architecture, along with k-fold cross-validation and augmentation, prepares the highest average accuracy of the three experiments conducted. This system can accurately classify five species of bacteria: Acinetobacter baumanii, Escherichia coli, Neisseria gonorrhoeae, Propionibacterium acnes, and Veionella. When using the InceptionV3 architecture with k-fold cross-validation and augmentation, along with mini batch size 5, max epoch 5, initial learning rate 0.0001, and validation frequency 3 training configurations, the system achieved an average validation accuracy of 94.42% and an average training time of 7 minutes and 27 seconds However, this study contain some drawbacks which is only classified five types of bacteria, which may limit its generalizability. The system has the potential to be expanded to classify more types of bacteria or other objects to address various issues. The laborious process of identifying bacteria may pose challenges in the Scalability and efficiency of the classification system. While this research by [45] emphasizes that identifying and categorizing bacteria is critical in medicine for disease diagnosis, infection handling, and epidemic realization, manual identification and categorization of bacteria require significant time and effort. With technological advancements, computer systems can now identify images captured by digital electron microscopes. Additionally, modern Deep Learning (DL) methods have significantly improved image classification. This research explores the use of a DL model to automate the classification and identification of bacteria. To develop the DL model, they utilized a dataset of over 600 images of 33 distinct bacteria captured with a microscope. We also employed the ’transfer learning’ technique. Transfer learning models such as GoogLeNet and AlexNet were utilized in this study. The DL classification accuracy was evaluated utilizing 20% of randomly chosen images from the dataset. The experimental findings revealed that GoogLeNet achieved an accuracy of approximately 98.67%, and both transfer learning models successfully classified all 33 bacterial kinds with high success rates. A limitation of this research is that it focused on only one species of bacteria; additional classification of other microorganisms is needed. In a study [8], a computerized Deep Learning (DL) technique was developed to quickly and accurately identify bacterial genera. Traditional methods such as culture, staining, and morphological characteristics are commonly used but are time-consuming. The developed technique uses a convolutional neural network (CNN) as part of machine learning to discriminate between bacterial genera such as Streptococcus, Staphylococcus, Escherichia, Salmonella, and Corynebacterium. This technique was improved using Python and the Keras API with the TensorFlow ML/DL framework. A final of 200 microscopic images, with 40 images for each of the mentioned genera, were used in this research. The improved technique achieved a high accuracy rate in identifying and distinguishing microscopic images of Streptococcus, Staphylococcus, Escherichia, Salmonella, and Corynebacterium, with a high accuracy of 92.20% for Staphylococcus and the low of 77.40% for Salmonella. Staphylococcus had the highest accuracy range among the five epochs. The results indicate that the DL method is effective for predicting bacterial genera. The study suggested that while the improved DL technique shows promise for identifying bacterial genera, further improvement is needed to expand its use to include more bacterial genera, especially those with similar morphology. The limitation of the study was its focus on only one type of microorganism. However, there is potential for further enhancements to accommodate a wider range of bacterial genera. Figure 8 shows the structure of CNN.Moreover, [46] presented a deep learning approach using convolutional

neural networks classification of bacterial types based on the morphology of bacterial colonies, achieving a high accuracy of 97.19% for three bacterial colonies using a training set of 5,000 augmented images. Some limitations in this paper include the time-consuming process of determining bacterial species, the requirement for long-term experience from bi- ology specialists, the costly specialized machines used for analysis, and reliance on a training set of 5,000 augmented images deriving from only 40 original photos, which may limit generalizability. Additionally, the study was conducted in a specific laboratory setting (Hanoi Medical University in Vietnam), which may not be generalized to other settings.

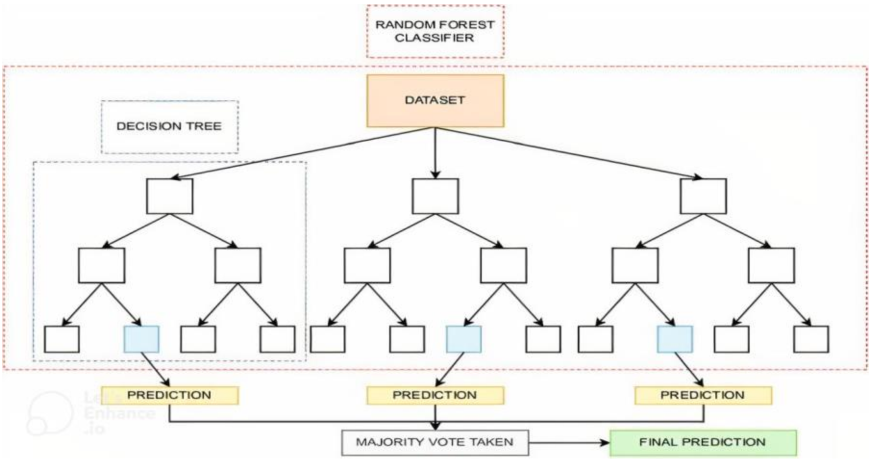

Figure 7. Structure of Random Forest algorithm.

Additionally, [47] used Raman spectroscopy for accurate microorganism identification with predictive models, focusing on fungal spores and carotenoid-pigmented microorganisms for classification. The results showed 94% accuracy, but the study had several limitations: it did not provide specific details on the limitations of current prediction models, lacked explanation on accuracy, focused narrowly on baselines, only considered a limited scope of data types, overlooked methodological limitations, assumed benefits of data pretreatment without evidence, and neglected practical implementation issues. It did not address challenges faced by the models, explore accuracy in depth, consider alternative spectral features, or discuss method limitations, potential drawbacks of data pretreatment or practical challenges in real-world implementation.

Another study by [48] presented a multi-channel image processing approach to automatically extract features and identifying pneumonia from chest X-ray (CXR) images. This method aims to improve the accuracy of pneumonia detection in CXR images, which can often be of low quality. Three different channels of CXR images, Local Binary Pattern (LBP), Contrast Enhanced Canny Edge Detection (CECED), and Contrast Limited Adaptive Histogram Equalization (CLAHE) are processed utilized deep neural networks. While features from CLAHE images are obtained through a pre-trained Inception-V3 model and features from CECED images are extracted utilizing a pre-trained MobileNet-V3 model. The final features from three channels are combined, and a softmax classification is used to obtain the presence of pneumonia. The introduced network demonstrated high accuracy in classifying pneumonia, with experimental results showing an accuracy of 98.3%, a sensitivity of 98.9%, and a specificity of 99.2%. Compared to other models, the proposed network performs comparably well.

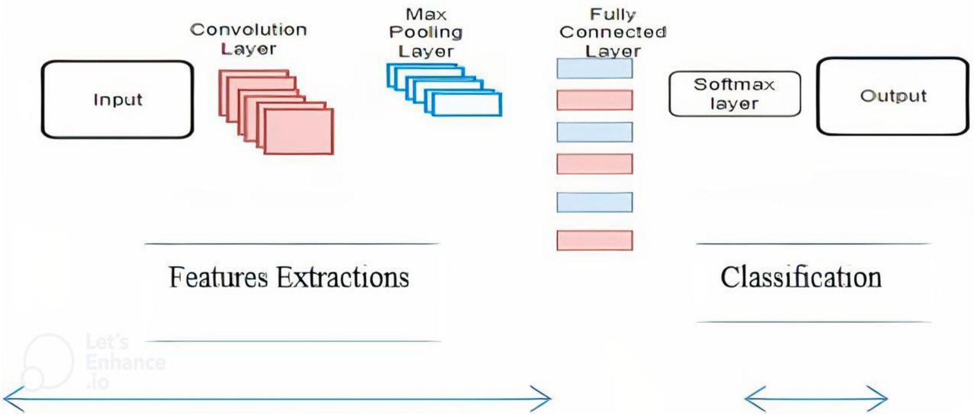

Figure 8. The structure of CNN

Furthermore, the study by [8] aimed to develop a deep learning-based auxiliary diagnostic model for early infectious keratitis (IK) diagnosis. The study collected 4,830 slit-lamp images from patients diagnosed with infectious keratitis between June 2010 and May 2021. The breakdown of diagnoses included 1,490 (30.8%) bacterial keratitis (BK), 1,670 (34.6%) fungal keratitis (FK), 600 (12.4%) herpes simplex keratitis (HSK), and 1,070 (22.2%) Acanthamoeba keratitis (AK). The combined model, called KeratitisNet, which integrates ResNext101 32x16d and DenseNet169, achieved the highest accuracy at 77.08%. The accuracy of KeratitisNet for diagnosing BK, FK, AK, and HSK was 70.27%, 77.71%, 83.81%, and 79.31%, respectively. The area under the curve (AUC) for diagnosing BK, FK, AK, and HSK was 0.86, 0.91, 0.96, and 0.98, respectively. A limitation of this research is that the model has difficulty distinguishing between bacterial and fungal keratitis.

On the other hand, the study [49] developed a fast, precise, and robust method for identifying *Arcobacter* bacteria species employing Raman spectroscopy. When combined with a CNN model, this approach can analyze the biochemical signatures of various *Arcobacter* strains and achieve a high identification accuracy of 97.2% at the species level. Furthermore, it can predict the relative ratio of a specific *Arcobacter* species within a bacterial mixture. This method is faster and more cost-effective compared to nucleic acid amplification-based methods currently in use and has comparable or even better accuracy. As a result, it holds promise for this study of *Arcobacter* and can directly detect *Arcobacter* species in practical settings, such as in food and environmental samples. In the future, implementing Raman spectroscopy could help create a standardized central warning system by constructing a shared database among surveillance systems for the relevant bacterial strains of interest. The limitations of this study include the underestimation of *Arcobacter* in agro-ecosystems due to detection limitations and the limited availability of detection and characterization techniques for *Arcobacter* species.

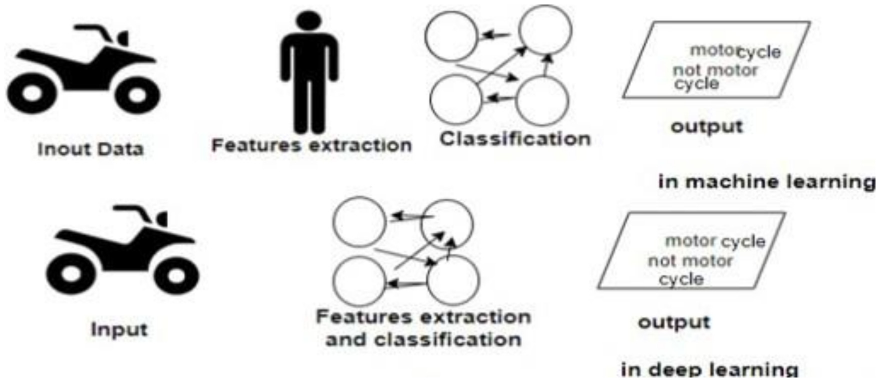

Figure 9. Shows the differences between deep learning and machine learning in classification

For instance, in the research [50], a new gene sequence was represented with Genomic Signal Processing for classifying SARS-CoV-2. The study achieved high accuracy and precision in viral classification using deep learning, with results showing an accuracy of 98.35%, 99.08%, and 99.69% for different viral signature sizes. The method obtained 99.95% precision for vectors with a size of 256. Figure 9 shows the differences between deep learning and machine learning in classification.

- **Long Short-term Memory**

Long Short-Term Memory (LSTM-RNN) models have been successfully used in the classification of medical images. For example, [59] developed a deep learning method based on an attention mechanism that integrates LSTM recurrent neural networks for lung image classification. The proposed model achieved an overall classification accuracy of 95.93%, outperforming other deep learning networks. However, further validation and testing on various datasets are very important to ensure the generalizability and robustness of the model. The study specifically focuses on lung image classification, and expanding the research to include other medical imaging tasks or diseases could provide a more comprehensive evaluation of the model's performance. Additionally, the potential computational complexity of implementing this deep learning model may be a limitation in real-world applications that require efficient processing of large amounts of data. Similarly, [60] used LSTM models for the classification of micro aneurysms in fundus images for diabetic retinopathy diagnosis. Their model achieved sensitivity scores of 66.6% for the e-aphtha dataset and 60.5% for the ROC dataset, highlighting the challenge of the study. Furthermore, Shaker and [59] introduced LSTM-RNN for lung image classification with end-to-end training to reduce model training overhead. While researchers are increasing network complexity for better classification performance, this model focuses on a lightweight network with an attention mechanism CBAM for extracting lung image features. Experiments on medical datasets demon- started good classification performance, with plans to optimize the model further for improved accuracy in a smaller network. Additionally, research is ongoing in image preprocessing to address the noise in lung images. The experiment involved selecting lung X-ray images from a public medical dataset, including both normal lung images and images of patients with lung diseases. The training set consisted of a 5:5 ratio of images, and the model was successfully trained, achieving an accuracy of 95.93%. Additionally, the authors [61] employed deep learning techniques to predict virus mutations, specifically codon mutations, in six types of coronaviruses:

Table 5: provides a critical studies of classification techniques, including DL and ML algorithms

Authors	Research Aims	Research Problem/ Gaps	Scope	Method	Features extract	Dataset	Accuracy	Software/Tool
[51]	Explore how CNNs utilize Raman spectra for microorganism classification	Gaps in understanding CNNs' use of Raman spectra	One class (Fungus)	CNN	Edges, Texture	Public	94%	MATLAB R2021b
[36]	Develop a system for identifying microorganisms with high accuracy	Bias in data and over-fitting	Identification of microorganisms	SVM, KNN, MLP, QDA	Unique characteristics, shape, color	Public	98.2%	Python
[52]	Identify bacterial pneumonia and COVID-19 virus	Not mentioned	COVID-19, virus, and bacteria pneumonia classification	CNN, CECED, Inception-v3, MobileNet3	Texture	Public	98.3%	Python
[53]	Develop an early diagnostic model for infactious keratitis	Unbalanced data, misclassification between BK and FK	Bacteria, fungal, acanthamoeba, herpes keratitis	CNN (ResNet18, ResNet50, DenseNet121, DenseNet169, EfficientNet)	Color, Shape	Public	70.27–83.81%	MATLAB
[49]	Identify Arcobacter species using Raman spectroscopy	Underestimation of Arcobacter due to detection limitations	One class (Bacteria)	Raman spectroscopy	Nucleic acids, proteins in bacterial cells	Public	97.2%	MATLAB 2018b
[54]	Enhance cellular image analysis in micro-biology (focus on parasites)	Morphological variability in parasites due to life cycle changes	One class (Parasites)	Deep-cycle transfer learning (DCTL), VGG-Net,	Shape characteristics	Public	95.7%	Python
[55]	Identify microbial keratitis	Small sample size, non-pathogenic elements affect accuracy	One class (Bacterial microbe)	Random Forest composition	Bacteria	Public	96.25%	PICRUSt2
[56]	Identify gut microbiome	Understanding gut microbiome metabolism is challenging	One class (Bacteria)	Random Forest imbalance in the gut	Bacterial	MetaHI, HMP	Not mentioned	Python, MATLAB
[57]	Classify and identify environmental microorganisms	Not mentioned	21 environmental microorganism classes	SVM, Random Forest	Not mentioned	Public	55%	Python 3.6
[58]	Recognize and classify pneumonia pathogens	Imbalanced data	One class (Bacteria)	Deep Diagnostic Agent Forest (DDAF)	Not mentioned	Single institution dataset	77.08%	Python

6. Internet of things/ edge devices with MLand DL

Internet of Things (IoT) technologies have proliferated in recent years and have ingrained themselves into many facets of our lives. IoT devices, for example, are often used to improve environmental monitoring, energy-smart wearable technologies, enable smart city solutions, and strengthen healthcare systems. IoT devices generated enormous volumes of data in each of these situations, which may include numerical data from several sensors or images and videos captured by cameras. Processing such amounts of data at data centers has became more difficult, particularly in situations where internet connectivity is unstable. A computing architecture known as edge computing has started to emerge. Bringing processing units closer to end devices is the concept underlying edge computing allowing data to be to processed rather than sending them to a centralized service. These points provide additional benefits, such as enhancing the response times, conserving network bandwidth, increasing scalability by reducing reliance on centralized services, and improving data privacy, which will not be necessary to send all data to the cloud or data center for additional processing. The processes of data on edge devices are still computationally expensive and rely on centralized servers, making it more difficult to apply machine learning at the edge. In the same way, the deep learning techniques that employ multi-layered artificial neural networks (AANs), which are less reliant on human intervention than typical machine learning models, automate feature extraction, and frequently learn from larger data sets, making it difficult to apply them on edge, these challenges of edge devices(6)

- Limited Devices and Computation Offloading

One of the major limitations of edge computing is that computations now being done by computing boards with inferior CPUs, low memory, and low disk space than machines performing in cloud data centers, wherefore the edge computing Workloads must be improved and adapted to the computing capabilities of the chosen edge device. This may effect of accuracy and speed

- Safety and privacy

It is necessary to make sure that each device has a limited set of authorization to avoid a compromised edge device might being utilized to obtain access to crucial system components, wherefore protecting data privacy is crucial. To solve this problem, data sets are subjected to anonymization to remove or replace personally identifiable information.

- Managing of energy consumption

One of the main challenges in edge computing is managing energy consumption, especially for devices that operate solely on batteries when they do not have a stable power source. This problem becomes more critical when the workload requires high computing power or needs additional hardware accelerators, such as onboard GPUs, which are commonly used in machine learning and deep learning applications. To handle these challenges, algorithms must be enhanced at the edge devices to balance performance and energy efficiency. Researchers are actively working on improving this balance. Another technique to reduce energy consumption is computation offloading. This practice consists of transferring certain tasks to other edge devices that are better equipped, such as those with GPUs, or to devices with more available energy, or even directly to the cloud.

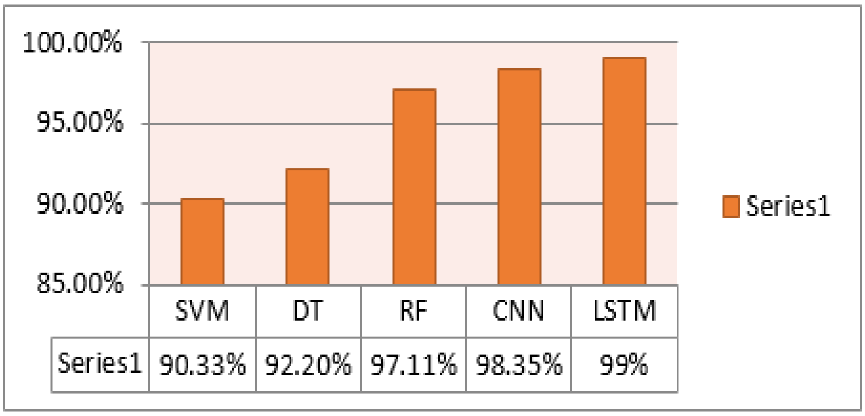

Figure 10. Percentage of performance metrics reported in studies reviewed for SVM, DT, RF, CNN, and LSTM

Table 6: Critical studies of long-term short memory.

Authors	**Research Aims**	**Method/Technique**	**Limitations/Gaps**	**Accuracy**
[51]	To classify lung disease caused by microorganisms	LSTM and Deep Learning (VGG16, VGG19)	Lack of generalization and limited data.	95.93%
[52]	To classify microaneurysms in fundus image, LSTM	classifier	Limited sensitivity.	66.6% (e-ophtha) 60.5% (ROC)
[53]	To predict virus mutations	LSTM	The study did not address the generalizability of the model to other contexts.	99%
[54]	To classify Coronavirus	LSTM	The study's results may not fully address the key issues in health representation.	RMSE: 0.03 (training) RMSE: 0.00 (testing)
[55]	To identify disease subtypes caused by microorganisms	Backpropagation (BP)method and LSTM	Adding more layers to an RNN can cause the gradient of the loss function to approach zero.	Not mentioned
[56]	To classify Covid-19	Bi-LSTM	Not mentioned.	94.88%

Table 7: Advantages and limitations of classification techniques

Classification Techniques	Advantages	Limitations
Support Vector Machine	• Can be utilized for both linear and non-linear purposes. • Includes powerful kernels such as linear, Gaussian, and polynomial kernels. • Effective at identifying and providing correct results. • Performs accurately even if the training images contain noise.	• Training process is time-consuming. • Structure of the algorithm is difficult to understand. • Works with only two classes.
Random Forest Algorithm	• Handles missing data well and maintains accuracy despite of significant loss • Can be used for both regression and classification. • Does not require extensive design and training. • Easy to implement and offers good computational efficiency.	• Complex process involving creation and combination of numerous trees. • Time-consuming.
Convolution Neural Network	• Effective for medical image classification. • Can leverage pre-trained weights, reducing training time. • Includes powerful architectures like LeNet, AlexNet, VGG, ResNet, and GoogLeNet. • Convolutional layers allow extraction of crucial features from images.	• Requires a large number of training samples. • Training can be time-consuming. • Overfitting may occur.
Long Short-Term Memory	• Efficient in handling sequential data and long-term predictions. • Adaptable to different input sequences.	• Complexity of the model. • Higher risk of overfitting when the training dataset is not sufficiently large.

7. Current challenges and knowledge gaps

a. Challenges in machine learning

Some machine learning models in medical imaging suffer from unclear explanation mechanisms, which makes it challenging to assess their scientific validity and affects the trust of specialized users such as oncologists or market traders. Therefore, there are some challenges in machine learning [67]:

Bias and data fairness: Significantly affect machine-learning models, undermines the cultural and social perception of machine learning and artificial intelligence as a whole. In addition, these shortcomings can be handled with introduce in scientific literature. For instance, explanation frameworks can be integrated with "black box model through the use of decision trees or polynomial regression. In addition, the bias was detected and reduced using pre-processing, in-processing, and post- processing algorithms. Moreover, techniques such as reinforcement learning (RL) and imitation learning (IL) suggest a stronger similarity in the way machine learning models and humans learn, potentially indicating a more durable theoretical foundation for artificial intelligence. Figure 12 shows the challenges of machine learning.

- **Training samples:** Machine learning algorithms like support vector machine (SVM), random forest (RF), decision tree, and k-nearest neighbor (KNN) often struggle with overfitting on training samples, performing well only on test samples from the same dataset. This issue is particularly problematic in medical imaging applications due to scanner variability, scan settings, subject demographics, and disease heterogeneity. Therefore, it is essential to interpret trained networks and validate the significant features they have learned. Reporting testing results with an external dataset not used for training is important.
- **Limited dataset:** Availability of datasets can pose challenges. However, some research failed to investigate how the model performs on a broader and more diverse dataset.
- **Accuracy:** The accuracy is often inadequate, especially in complex tasks.
- **Feature engineering:** In machine learning, feature extraction is often based on manual techniques.

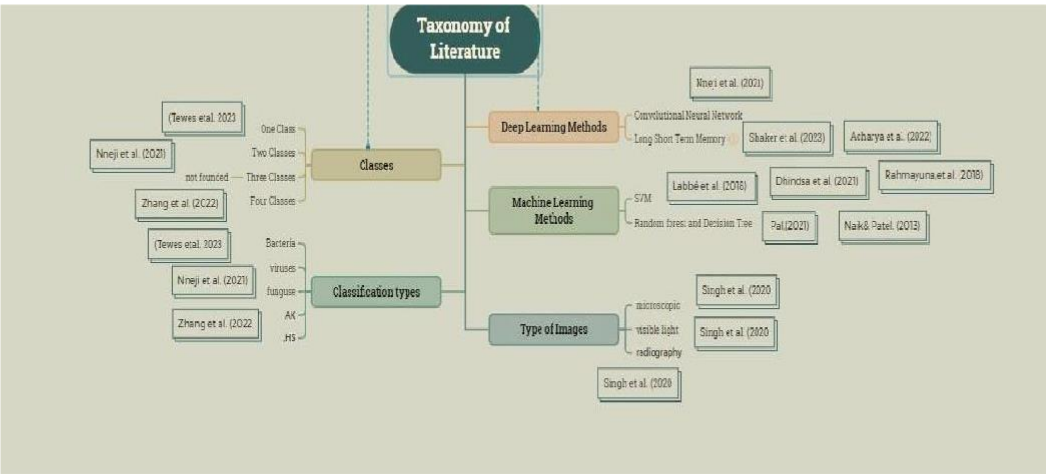

Figure 11. Showed the taxonomy of literatures

Figure 12. Challenges in machine learning.

b. Challenges in Deep Learning

Deep learning models require a huge number of training samples in medical images for effective training, as seen in models trained on large datasets like ImageNet. There are some challenges in deep learning:

i. **Large training dataset required:** It remains unclear whether deep learning models can effectively work with smaller datasets, such as medical images, due to the unique characteristics of medical images.

ii. **Errors in feature selection:** There is ongoing uncertainty in deep learning about whether models are genuinely identifying discriminative features or merely overfitting to specific ones within a given dataset. Deep learning algorithms can process raw image data without manual feature constructing, with convolutional neural networks (CNNs) responsible for uncovering the relevant features from the data. CNNs have simplified the process of encoding raw features into a latent space. However, it is crucial to determine if these features are transferable across different datasets. This limitation is particularly problematic in medical imaging due to scanner variability, scan settings, subject demographics, and disease heterogeneity [68]. Therefore, it is essential to interpret trained networks and validate the significant features they have learned. Reporting testing results with an external dataset not used for training is important, although limited dataset availability can pose challenges.

Feature extraction involves capturing distinct patterns from raw pixel values to differentiate between categories. This process is done without supervision, meaning the extracted information is not explicitly related to image classes. In CNNs, the convolution layer is crucial for feature extraction as it allows parameter sharing. When features are extracted, a classifier is trained using the images and their associated labels (e.g., logistic regression, random forests, decision trees, support vector machines, etc.). However, a drawback of this pipeline is that feature extraction cannot be adapted based on the classes and images, which may compromise the model's accuracy if the chosen features do not provide sufficient information to distinguish categories. Researchers have explored combining various feature extractors to improve feature extraction, but this approach requires considerable manual effort to adjust settings based on the domain. The core idea in deep learning is that there is no specific way to extract features from data (no hard-coding). Some research has not investigated how the model performs on broader and more diverse datasets.

a. **Complexity of models:** The models can be limited in generalizing data and may suffer from misclassification due to limited datasets. The ultimate goal is to surpass human-level performance, known as Artificial General Intelligence. However, obstacles such as a lack of labeled images, high labeling costs, and disagreements among experts on label validation hinder progress. Previous studies have been plagued by inadequate data provision, leading to inaccurate classification and low accuracy. Additionally, there has been a lack of data generalization, causing overfitting. To address these challenges, reliable data augmentation methods and the generations of samples with known ground truths are essential considerations.

b. **Training time:** Due to the complexity of models, the training process often suffers from long time consumption.

c. **Overfitting:** Deep learning models may suffer from overfitting due to insufficient datasets.

d. **Hardware requirements:** Due to the complexity of models, high hardware requirements are necessary. This complexity often results in lengthier training times compared to models that are more straightforward.

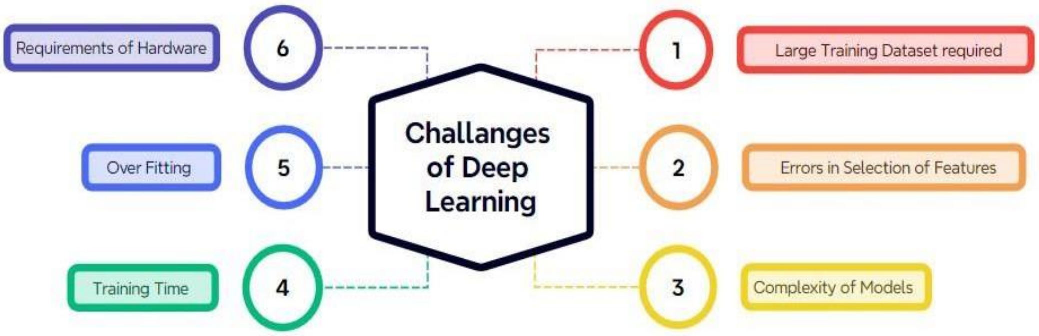

Figure 13. The challenges in deep learning.

8. Conclusion and Future Work

Studying microorganisms using intelligent image recognition is a complex and challenging task. However, advancements in machine learning (ML) and deep learning (DL) algorithms, particularly Convolutional Neural Networks (CNNs), have shown promise in developing accuracy in microorganism image classification. Deep learning algorithms often outperform traditional machine learning algorithms in this task, highlighting the importance of using the appropriate techniques and models for specific tasks to obtain the best results. Despite promising results, challenges remain in microorganism image recognition. Factors such as data quality and quantity can significantly affect model accuracy, and the variability of microorganisms' characteristics can make accurate identification and classification difficult. Further advancements in ML and DL algorithms are needed to address these limitations and enhance accuracy rates in microorganism image recognition. Methods like data augmentation and transfer learning have the potential to improve model accuracy by allowing algorithms to learn from a larger and various dataset. The study of microorganisms through intelligent image recognition is a rapidly evolving field with the potential to advance in various areas such as clinical microbiology, agriculture, medical science, and food production. utilized of ML and DL algorithms, specifically CNN models, has shown promise in developing accuracy rates in microorganism image detection, and continued advancements in these techniques could lead to even greater development in accuracy and effectiveness, such as enhanced data augmentation and providing more comprehensive datasets. One of the key strengths of this review is its in-depth examination of the literature, particularly focusing on classification methods and the challenges encountered in previous studies. The authors adeptly integrate diverse studies, providing a critical assessment that deepens our understanding. Advancements in ML and DL algorithms are critical for overcoming challenges and enhancing accuracy in the classification of microorganism images and connecting them with the Internet of Things and edge devices (10). Techniques like data augmentation and transfer learning have the potential to improve model performance by enabling them to learn from a larger and more diverse dataset. The field of intelligent image recognition in microbes is rapidly progressing and has the potential to revolutionize various industries, including clinical microbiology, agriculture, medical science, and food production. The use of ML and DL algorithms, particularly CNN models, has shown promise in boosting accuracy rates in microorganism image recognition. The recommendations of this research are greater improvements in accuracy and effectiveness by providing more data, alongside specialized hardware for training deep learning models at the network edge. Some methods minimize the memory used by the edge device and accelerate training on low-resource devices. Alongside specialized hardware for training deep learning lightweight models at the network edge, four main techniques are commonly used to compress pre-trained models: pruning, quantization, knowledge distillation, and low-rank decomposition.

References

[1] Mendez-Vilas, *Microbes in Applied Research: Current Advances and Challenges*. Malaga, Spain, 2011, pp. 1–677, doi: 10.1142/8462.

[2] O. M. Prakash, R. Verma, and V. Sharma, "Polyphasic approach of bacterial classification—An overview of recent advances," *Indian J. Microbiol.*, vol. 47, no. 2, pp. 98–108, Jun. 2007, doi: 10.1007/s12088-007-0022-x.

[3] T. Ramamurthy, A. Ghosh, G. P. Pazhani, and S. Shinoda, "Current perspectives on viable but non-culturable (VBNC) pathogenic bacteria," *Front. Public Health*, vol. 2, p. 103, Jul. 2014, doi: 10.3389/fpubh.2014.00103.

[4] J. Wang *et al.*, "An image pre-processing improvement procedure for standardizing multiple colour variations in the astrocytoma immunohistochemical staining images," *J. Phys.: Conf. Ser.*, vol. 2071, p. 012036, 2021, doi: 10.1088/1742-6596/2071/1/012036.

[5] N. Saad, N. Noor, and A. A. Abdullah, "A review on image segmentation techniques for MRI brain stroke lesion," *J. Telecommun. Electron. Comput. Eng.*, vol. 13, no. 4, pp. 1–8, 2021. [Online]. Available: http://jtec.utem.edu.my/jtec/article/view/6145

[6] P. Celard *et al.*, "A survey on deep learning applied to medical images: from simple artificial neural networks to generative models," *Neural Comput. Appl.*, vol. 35, no. 3, pp. 2291–2323, 2023, doi: 10.1007/s00521-022-07953-4.

[7] Khan *et al.*, "Short-term traffic prediction using deep learning long short-term memory: Taxonomy, applications, challenges, and future trends," *IEEE Access*, vol. 11, pp. 89714–89740, 2023, doi: 10.1109/ACCESS.2023.3309601.

[8] Zhang, X. Zhang, and D. Tu, "A set of comprehensive evaluation system for different data augmentation methods," *Mobile Inf. Syst.*, vol. 2022, Art. no. 8572852, 2022, doi: 10.1155/2022/8572852.

[9] Kitchenham *et al.*, "Systematic literature reviews in software engineering—a tertiary study," *Inf. Softw. Technol.*, vol. 52, no. 8, pp. 792–805, Aug. 2010, doi: 10.1016/j.infsof.2010.03.006.

[10] J. Webster and R. T. Watson, "Analyzing the past to prepare for the future: Writing a literature review," *MIS Quart.*, vol. 26, no. 2, pp. xiii–xxiii, Jun. 2002, doi: 10.2307/4132319.

[11] M. Kumpulainen and M. Seppänen, "Combining Web of Science and Scopus datasets in citation-based literature study," *Scientometrics*, vol. 127, no. 10, pp. 5613–5631, Oct. 2022, doi: 10.1007/s11192-022-04475-7.

[12] M. A. Azam *et al.*, “A review on multimodal medical image fusion: Compendious analysis of medical modalities, multimodal databases, fusion techniques and quality metrics,” *Comput. Biol. Med.*, vol. 144, Art. no. 105253, May 2022, doi: 10.1016/j.compbiomed.2022.105253.

[13] T. Gong *et al.*, “Classification of CT brain images of head trauma,” in *Proc. Pattern Recognit. Bioinf. (PRIB)*, 2007, pp. 401–408, doi: 10.1007/978-3-540-75286-8_38.

[14] M. Srinivas, R. Bharath, P. Rajalakshmi, and C. K. Mohan, “Multi-level classification: A generic classification method for medical datasets,” in *Proc. 17th Int. Conf. E-Health Netw., Appl. Serv. (HealthCom)*, Oct. 2015, pp. 262–267, doi: 10.1109/HEALTHCOM.2015.7454509.

[15] J. Islam and Y. Zhang, “A Novel Deep Learning Based Multi-class Classification Method for Alzheimer’s Disease Detection Using Brain MRI Data,” in *Lect. Notes Comput. Sci.*, vol. 10654, 2017, pp. 213–222, doi: 10.1007/978-3-319-70772-3_20.

[16] Mikołajczyk and M. Grochowski, “Data augmentation for improving deep learning in image classification problem,” in *Proc. Int. Interdiscipl. PhD Workshop (IIPhDW)*, Jun. 2018, pp. 117–122, doi: 10.1109/IIPHDW.2018.8388338.

[17] S. S. Yadav and S. M. Jadhav, “Deep convolutional neural network based medical image classification for disease diagnosis,” *J. Big Data*, vol. 6, no. 1, pp. 1–18, Dec. 2019, doi: 10.1186/S40537-019-0276-2.

[18] J. Hu *et al.*, “Probability analysis for grasp planning facing the field of medical robotics,” *Measurement*, vol. 141, pp. 227–234, Jul. 2019, doi: 10.1016/j.measurement.2019.03.010.

[19] T. A. Soomro *et al.*, "Deep learning models for retinal blood vessels segmentation: A review," *IEEE Access*, vol. 7, pp. 71696–71717, Jun. 2019, doi: 10.1109/ACCESS.2019.2920616.

[20] W. Zhang *et al.*, “Dynamic-fusion-based federated learning for COVID-19 detection,” *IEEE Internet Things J.*, vol. 8, no. 21, pp. 15884–15891, Nov. 2021, doi: 10.1109/JIOT.2021.3056185.

[21] S. Deepak and P. M. Ameer, “Automated categorization of brain tumor from MRI using CNN features and SVM,” *J. Ambient Intell. Humaniz. Comput.*, vol. 12, no. 5, pp. 8357–8369, Aug. 2021, doi: 10.1007/s12652-020-02568-w.

[22] T. Murad, S. Ali, and M. Patterson, “Exploring the potential of GANs in biological sequence analysis,” *Biology*, vol. 12, no. 6, p. 854, 2023, doi: 10.3390/biology12060854.

[23] S. Kotwal, P. Rani, J. Manhas, and V. Sharma, “Automated bacterial classifications using machine learning-based computational techniques: Architectures, challenges, and open research issues,” *Arch. Comput. Methods Eng.*, vol. 29, no. 4, pp. 2469–2490, Jun. 2022, doi: 10.1007/s11831-021-09660-0.

[24] J. A. Wani, S. Sharma, and M. M. Muzamil, “Machine Learning and Deep Learning Based Computational Techniques in Automatic Agricultural Diseases Detection: Methodologies, Applications, and Challenges,” *Arch. Comput. Methods Eng.*, vol. 29, no. 1, pp. 641–677, Jan. 2022, doi: 10.1007/s11831-021-09588-5.

[25] Y. Wu and S. A. Gadsden, “Machine learning algorithms in microbial classification: a comparative analysis,” *Front. Artif. Intell.*, vol. 6, Art. no. 1200994, 2023, doi: 10.3389/FRAI.2023.1200994.

[26] Y. Jiang *et al.*, “Machine learning advances in microbiology: A review of methods and applications,” *Front. Microbiol.*, vol. 13, Art. no. 925454, May 2022, doi: 10.3389/fmicb.2022.925454.

[27] L. R. Farias *et al.*, “Rapid and green classification method of bacteria using machine learning and NIR spectroscopy,” *Sensors*, vol. 23, no. 17, Art. no. 7336, Aug. 2023, doi: 10.3390/s23177336.

[28] Y. Nami, N. Imeni, and B. Panahi, “Application of machine learning in bacteriophage research,” *BMC Microbiol.*, vol. 21, no. 1, p. 193, Jun. 2021, doi: 10.1186/s12866-021-02256-5.

[29] K. Qu *et al.*, “Application of machine learning in microbiology,” *Front. Microbiol*, vol. 10, Art. no. 827, Apr. 2019, doi: 10.3389/fmicb.2019.00827.

[30] R. Franco-Duarte *et al.*, "Advances in chemical and biological methods to identify microorganisms—from past to present," *Microorganisms*, vol. 7, no. 5, Art. no. 130, May 2019, doi: 10.3390/microorganisms7050130.

[31] M. M. Ramzan *et al.*, “Recent studies on advance spectroscopic techniques for the identification of microorganisms: A review,” *Arabian J. Chem.*, vol. 15, no. 9, Art. no. 104521, 2022, doi: 10.1016/j.arabjc.2022.104521.

[32] K. M. Maraz and R. A. Khan, “An overview on impact and application of microorganisms on human health, medicine and environment,” *GSC Biol. Pharm. Sci.*, vol. 16, no. 1, pp. 89–104, Jul. 2021, doi: 10.30574/gscbps.2021.16.1.0200.

[33] K. Ayhan *et al.*, “Advance methods for the qualitative and quantitative determination of microorganisms,” *Microchem. J.*, vol. 166, Art. no. 106188, 2021, doi: 10.1016/j.microc.2021.106188.

[34] Q. Lv, S. Zhang, and Y. Wang, “Deep Learning Model of Image Classification Using Machine Learning,” *Adv. Multimedia*, vol. 2022, Art. no. 3351256, Jan. 2022, doi: 10.1155/2022/3351256.

[35] N. Rahmayuna *et al.*, "Pathogenic Bacteria Genus Classification using Support Vector Machine," in *Proc. Int. Semin. Res. Inf. Technol. Intell. Syst. (ISRITI)*, Nov. 2018, doi: 10.1109/ISRITI.2018.8864478.

[36] Dhindsa, S. Bhatia, S. Agrawal, and B. S. Sohi, “An improvised machine learning model based on mutual information feature selection approach for microbes classification,” *Entropy*, vol. 23, no. 2, Art. no. 257, Feb. 2021, doi: 10.3390/e23020257.

[37] Plichta, “Recognition of species and genera of bacteria by means of the product of weights of the classifiers,” *Int. J. Appl. Math. Comput. Sci.*, vol. 30, no. 3, pp. 463–473, Sept. 2020, doi: 10.34768/amcs-2020-0034.

[38] D. Babenko *et al.*, “Ability of procalcitonin and C-reactive protein for discriminating between bacterial and enteroviral meningitis in children using decision tree,” *BioMed Res. Int.*, vol. 2021, Art. no. 5519436, Aug. 2021, doi: 10.1155/2021/5519436.

[39] J. Naik and S. Patel, “Tumor detection and classification using decision tree in brain MRI,” *Int. J. Comput. Sci. Netw. Secur.*, vol. 14, no. 1, pp. 63–69, 2014. [Online]. Available: https://www.rjwave.org/ijedr/papers/IJEDR1301010.pdf

[40] S. Tantikitti, S. Tumswadi, and W. Premchaiswadi, “Image processing for detection of dengue virus based on WBC classification and decision tree,” in *Proc. 13th Int. Conf. ICT Knowl. Eng. (ICTKE)*, Nov. 2015, pp. 84–89, doi: 10.1109/ICTKE.2015.7368476.

[41] O. K. Pal, “Skin Disease Classification: A Comparative Analysis of K-Nearest Neighbors (KNN) and Random Forest Algorithm,” in *Proc. Int. Conf. Electron., Commun. Inf. Technol. (ICECIT)*, 2021, doi: 10.1109/ICECIT54077.2021.9641120.

[42] Akgundogdu, “Detection of pneumonia in chest X-ray images by using 2D discrete wavelet feature extraction with random forest,” *Int. J. Imag. Syst. Technol.*, vol. 31, no. 1, pp. 82–93, Mar. 2021, doi: 10.1002/IMA.22501.

[43] T. Irani, H. Amiri, and S. Azadi, “Use of a convolution neural network for the classification of E. Coli and V. cholera bacteria in wastewater,” *Environ. Res. Technol.*, vol. 5, no. 1, pp. 101–110, 2022, doi: 10.35208/ert.969400.

[44] Nurtanio *et al.*, “Multi classification of bacterial microscopic images using Inception V3,” *Ilkom J. Ilmiah*, vol. 14, no. 1, pp. 80–90, Apr. 2022, doi: 10.33096/ilkom.v14i1.1120.80-90.

[45] Singh *et al.*, “Transfer learning approach on bacteria classification from microscopic images,” in *Proc. 5th Int. Conf. Contemp. Comput. Inform. (IC3I)*, Nov. 2022, pp. 982–987, doi: 10.1109/IC3I56241.2022.10072818.

[46] M. Amano *et al.*, “Deep learning approach for classifying bacteria types using morphology of bacterial colony,” in *Proc. 44th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, 2022, pp. 1234–1237, doi: 10.1109/EMBC48229.2022.9870986.

[47] T. J. Tewes *et al.*, “Understanding Raman spectral based classifications with convolutional neural networks using practical examples of fungal spores and carotenoid-pigmented microorganisms,” *AI*, vol. 4, no. 1, pp. 114–127, Mar. 2023, doi: 10.3390/ai4010006.

[48] G. U. Nneji *et al.*, “Enhancing low quality in radiograph datasets using wavelet transform convolutional neural network and generative adversarial network for COVID-19 identification,” in *Proc. 4th Int. Conf. Pattern Recognit. Artif. Intell. (PRAI)*, Aug. 2021, pp. 146–151, doi: 10.1109/PRAI53619.2021.9551043.

[49] Wang *et al.*, “Arcobacter identification and species determination using Raman spectroscopy combined with neural networks,” *Appl. Environ. Microbiol*, vol. 86, no. 20, Oct. 2020, Art. no. e00924-20, doi: 10.1128/AEM.00924-20.

[50] C. de Souza *et al.*, “New proposal of viral genome representation applied in the classification of SARS-CoV-2 with deep learning,” *BMC Bioinformatics*, vol. 24, no. 1, Art. no. 92, Mar. 2023, pp. 1–19, doi: 10.1186/s12859-023-05188-1.

[51] T. J. Tewes *et al.*, “Understanding Raman spectral based classifications with convolutional neural networks using practical examples of fungal spores and carotenoid-pigmented microorganisms,” *AI*, vol. 4, no. 1, pp. 114–127, 2023, doi: 10.3390/ai4010006.

[52] G. U. Nneji *et al.*, “Enhancing low quality in radiograph datasets using wavelet transform convolutional neural network and generative adversarial network for COVID-19 identification,” in *Proc. 4th Int. Conf. Pattern Recognit. Artif. Intell. (PRAI)*, Aug. 2021, pp. 146–151, doi: 10.1109/PRAI53619.2021.9551043.

[53] Z. Zhang *et al.*, “Deep learning-based classification of infectious keratitis on slit-lamp images,” *Ther. Adv. Chronic Dis.*, vol. 13, 2022, doi: 10.1177/20406223221136071.

[54] S. Li, R. Feng, and Y. Zhang, “A knowledge-integrated deep learning framework for cellular image analysis in parasite microbiology,” *STAR Protocols*, vol. 4, no. 3, Art. no. 102452, Sep. 2023, doi: 10.1016/j.xpro.2023.102452.

[55] H. Dong, “Profiling of the Conjunctival Bacterial Microbiota Reveals the Feasibility of Utilizing a Microbiome-Based Machine Learning Model to Differentially Diagnose Microbial Keratitis and the Core Components of the Conjunctival Bacterial Interaction Network,” *Front. Cell. Infect. Microbiol*, vol. 12, Art. no. 860370, Apr. 2022, doi: 10.3389/FCIMB.2022.860370.

[56] H. L. Nielsen, “Machine learning for data integration in human gut microbiome,” *Microb. Cell Fact*, vol. 21, no. 1, Art. no. 241, Dec. 2022, doi: 10.1186/s12934-022-01973-4.

[57] F. Kulwa *et al.*, “A new pairwise deep learning feature for environmental microorganism image analysis,” *Environ. Sci. Pollut. Res.*, vol. 29, no. 34, pp. 51909–51926, Mar. 2022, doi: 10.1007/s11356-022-18849-0.

[58] W. Chen *et al.*, “Deep diagnostic agent forest (DDAF): A deep learning pathogen recognition system for pneumonia based on CT,” *Comput. Biol. Med.*, vol. 141, Art. no. 105143, Feb. 2022, doi: 10.1016/j.compbiomed.2021.105143.

[59] M. Shaker and S. Xiong, “Lung Image Classification Based On Long-Short Term Memory recurrent neural network,” *J. Phys.: Conf. Ser.*, vol. 2467, p. 012007, 2023, doi: 10.1088/1742-6596/2467/1/012007.

[60] R. Acharya and N. B. Puhan, “Long short-term memory model based microaneurysm sequence classification in fundus images,” in *Proc. IEEE Int. Conf. Signal Process. Commun.*, 2022, doi: 10.1109/SPCOM55316.2022.9840789.

[61] K. Al-Thiabi and A. J. D. Al-Alwani, “The Prediction of COVID-19 Virus Mutation Using Long Short-Term Memory,” in *Proc. 8th Int. Conf. Contemp. Inf. Technol. Math. (ICCITM)*, 2022, pp. 113–118, doi: 10.1109/ICCITM56309.2022.10032012.

[62] O. I. Obaid, M. A. Mohammed, and S. A. Mostafa, “Long Short-Term Memory Approach for Coronavirus Disease Prediction,” *J. Inf. Technol. Manage.*, vol. 12, pp. 11–21, Dec. 2020, doi: 10.22059/JITM.2020.79187.

[63] H. Okut, “Deep Learning for Subtyping and Prediction of Diseases: Long-Short Term Memory,” in *Deep Learning Applications*. IntechOpen, Feb. 2021, doi: 10.5772/INTECHOPEN.96180.

[64] D. A. Pustokhin *et al.*, “An effective deep residual network based class attention layer with bidirectional LSTM for diagnosis and classification of COVID-19,” *J. Appl. Stat.*, vol. 50, no. 3, pp. 477–494, Feb. 2023, doi: 10.1080/02664763.2020.1849057.

[65] Laranjeira, A. Lacerda, and E. R. Nascimento, “On modeling context from objects with a long short-term memory for indoor scene recognition,” in *Proc. 32nd Conf. Graph., Patterns Images (SIBGRAPI)*, Oct. 2019, pp. 249–256, doi: 10.1109/SIBGRAPI.2019.00041.

[66] Alnuaimi, “An overview of machine learning classification techniques,” *BIO Web Conf.*, vol. 97, p. 00133, 2024, doi: 10.1051/bioconf/20249700133.

[67] E. Barbierato and A. Gatti, “The Challenges of Machine Learning: A Critical Review,” *Electronics*, vol. 13, no. 2, p. 416, Jan. 2024, doi: 10.3390/ELECTRONICS13020416.

[68] Zieliński *et al.*, “Deep learning approach to describe and classify fungi microscopic images,” *PLOS ONE*, vol. 15, no. 6, p. e0234806, June 2020, doi: 10.1371/JOURNAL.PONE.0234806.

[69] Alshamrani, “IoT and artificial intelligence implementations for remote healthcare monitoring systems: A survey,” *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 8, pp. 4687–4701, 2022, doi: 10.1016/j.jksuci.2021.06.005.

[70] Chandy, "A review on IoT based medical imaging technology for healthcare applications," *J. Innov. Image Process.*, vol. 1, no. 1, pp. 51–60, Sep. 2019, doi: 10.36548/jiip.2019.1