

Rapid Detection and Classification of Escherichia Coli, Salmonella Typhi, and Staphylococcus Aureus using Smartphone-Based Deep Learning System

Maryam Arain ^{a*}, Sheeraz Ahmed Memon ^a, Attiya Baqai ^b, Mohsin Memon ^c, Kundan Kumar ^d, Sajid Hussain ^a, Zuhaib Siddiqui ^e

^a Institute of Environmental Engineering and Management, Mehran University of Engineering and Technology, Jamshoro, Pakistan

^b Department of Electronics Engineering, Mehran University of Engineering and Technology, Jamshoro, Pakistan

^c Department of Software Engineering, Mehran University of Engineering and Technology, Jamshoro, Pakistan

^d Department of Environmental Science, College of Natural Resources, University of Idaho, 995 MK Simpson Blvd, Idaho Falls, ID 83401, USA

^e School of Engineering, Western Sydney University, NSW, Australia

*Corresponding author: ms.maryam@faculty.muet.edu.pk

Received: 03 March 2026, Revised: 18 April 2026, Accepted: 22 June 2026, Published: 01 July 2026

KEY WORDS	ABSTRACT
Bacterial Classification Bacterial Morphology Deep Convolutional Neural Network Optical Imaging Smartphone Microscopy	The accurate and rapid detection of pathogens is important for maintaining infection control, ensuring food safety, supporting environmental monitoring, and supporting public health surveillance. Relying on traditional methods is time-consuming, which typically takes 24-72 hours to generate results. This study presents a machine-learning-based bacterial detection system that integrates smartphone imaging with convolutional neural networks to classify <i>Escherichia coli</i>, <i>Salmonella typhi</i>, and <i>Staphylococcus aureus</i>. In this study, a portable smartphone-compatible microscope with a magnification of 750x and a resolution of 1 μm was used, enabling high-quality bacterial imaging directly at the sampling site. To train three CNN architectures, namely, ResNet50, VGG16, and InceptionV3, hundreds of microscopic images were obtained by preparing, scanning stained bacterial slides, and then pre-processing them for this study. Furthermore, these models were assessed using various bacterial pair combinations to evaluate the classification accuracy. The results revealed that the system achieved 98.6% accuracy, demonstrating robust classification performance. The developed system was shown to outperform traditional cultural laboratory methods, enabling it to transform an ordinary smartphone into a handheld bioresearch machine that provides diagnostic results within minutes. The novelty lies in the comparative assessment of three CNN architectures across multiple bacterial class combinations, enabling robust assessment of model performance in distinguishing morphologically similar and distinct bacterial species. Through the combination of optical imaging, convolutional neural networks, and adaptive mobile computing, the investigation gains a significant decrease in the turnaround and operational cost of the assay and maintains the same level of accuracy as gold-standard assays.

1. Introduction

The classification of bacteria is crucial in various fields, including clinical diagnosis, food hygiene and environmental monitoring. However, bacterial morphology is highly variable in terms of dimensions,

form, and cellular structure [1]. These characteristics limit and pose challenges in traditional bacterial classification. To overcome the limitations and challenges, a deep learning-based approach is integrated for bacterial classification. Deep artificial neural networks known as convolutional neural

networks (CNNs are employed for object detection, image visualization, similarity clustering, and image classification. Using CNNs to extract feature properties in an interpretable way ensures dependability and enables evaluation of model authenticity based on training data, without being affected by individual intuition. In addition to academic research, deep learning is widely used in sectors that rely heavily on data. [2]. One promising architecture in deep learning is CNN, which has been successfully applied to solve problems related to visual imagery [3]. A smartphone-based fluorescent sensor has been employed to quickly detect various bacterial pathogens, such as *Escherichia coli* and *Staphylococcus aureus*, using portable equipment. [4]. Quick identification and examination of pathogenic organisms are crucial steps in averting health problems, safety issues, and well-being concerns. Early detection of pathogens is crucial for diagnosing and preventing illnesses. An automated identification and classification process can be developed to address these problems. Machine learning-assisted image processing has been extensively used to reduce workload and improve performance precision. Deep learning, a subset of machine learning, has achieved incredible success in a few biomedical applications specifically. Deep learning methods have been successfully employed for bacterial colony detection and identification using lens-free microscopy time-lapse images. A deep learning design, comprising two stages and combining convolutional and recurrent neural networks, has been developed to identify spatiotemporal patterns in unreconstructed lens-free microscopy images. The proposed method achieves high detection and classification accuracy, with a detection rate of 96.0% and classification precision and sensitivity of 93.1% and 94.0%, respectively. The system has been evaluated on a dataset of seven pathogenic bacteria, thereby demonstrating its ability to accurately detect and identify bacteria in real time [5]. Simultaneously, studies in bacterial image analysis employing deep learning (for example, multi-task CNNs) have demonstrated competitive accuracy in categorizing species or morphological states [6]. However, several gaps and challenges remain insufficiently addressed. However, the BIOINTEL system, which uses various cutting-edge CNNs, including VGG16, ResNet50, ResNet100, Inception v3, Lenet5, EfficientNet, and ConvNeXt, offers a significant improvement in bacterial identification. These helped process and analyse microscopy images in real time for applications such as clinical diagnostics and environmental monitoring, where prompt identification is essential [7]. BIOINTEL system

heavily rely on high quality imaging and lack detailed dataset characterization, the proposed study emphasizes multi-model evaluation, portability, and enhancing real-world applicability. However, most studies typically focus only on the detection of single bacterial species rather than on the classification of multiple species. In contrast, distinguishing between more than two species remains less explored. Additionally, Treebupachatsakul et al. (2019) employed the LeNet architecture to classify *Staphylococcus aureus* and *Lactobacillus delbrueckii*, achieving an accuracy of 75% [8]. Abd-Alhalem et al. (2020) investigated the use of CNNs for bacterial taxonomy prediction, focusing on DNA nucleotide sequences (A, T, C, G), and showed that CNNs outperformed SVM classifiers [9]. Kang et al. (2021) applied deep residual and recurrent networks to classify food-borne bacteria in hyperspectral microscopy images, with a 1D-CNN achieving 96.2% classification accuracy [10]. Mhathesh et al. (2021) also used CNNs for 3D fluorescence microscopy images of zebrafish gut bacteria, achieving an accuracy of 95%, outperforming traditional machine learning classifiers [12].

Despite these significant advancements, there remains a need for systematic evaluation of deep learning architectures, specifically for morphological classification of bacteria based on shape characteristics across different bacterial detection applications. Bacteria, based on shape characteristics, i.e., round-shaped (cocci) and rod-shaped (bacilli), serve as primary phenotypic markers in microscopic identification. This study introduces a novel smartphone-based bacterial identification system that integrates high-resolution portable microscopy with deep-learning classification. In this study, we focus on rod-shaped bacilli (*Escherichia coli* and *Salmonella typhi*) and round-shaped cocci (*Staphylococcus aureus*). While two common rod-shaped bacilli, i.e., *Escherichia coli* and *Salmonella typhi*, share similar morphology but differ in staining characteristics and subtle structural details, this makes image classification challenging. In contrast, round-shaped cocci, i.e., *Staphylococcus aureus*, exhibit a distinct morphology, forming grape-like clusters, providing a clear morphological contrast to bacilli. To systematically investigate shape-based discrimination, three binary classification datasets: (i) *Escherichia coli* and *Staphylococcus aureus*, (ii) *Staphylococcus aureus* and *Salmonella typhi*, and (iii) *Escherichia coli* and *Salmonella typhi* were constructed. These data set helped in improving system accuracy under distinct (rod vs. round) and similar (rods vs. rod) morphological conditions. The selected bacteria are of significant environmental,

clinical and public health importance, as they are commonly associated with environmental contamination and waterborne infections, enabling them relevant targets for rapid detection. The study primarily focused on developing a bacterial identification system using smartphones and deep learning. Through a designed system, end users were enabled to accurately determine bacterial morphological characteristics. Three CNN architectures, i.e., InceptionV3, ResNet50, and VGG16, were utilized across different bacterial pair combinations. This evaluation would be useful for enabling users to assess the strength of CNN architectures for classifying bacterial species exhibiting morphologically divergent yet closely related phenotypes. The proposed framework contributes to the enhancement of portable, efficient, and scalable bacterial identification systems.

The next step involved pre-processing the images to improve the quality of the inputs for model training. Then, the study narrowed down to the choice and a comparative analysis of model architectures using three state-of-the-art deep transfer learning models of ResNet50, VGG16, and InceptionV3 to classify bacterial images.

All models were trained, fine-tuned, and evaluated using standard performance metrics, including confusion matrices, learning curves, and accuracy, precision, recall, and F1-score. Each model was then used to assist in identifying and classifying bacteria.

2.3 Portable Smartphone Microscope

A portable smartphone microscope PSM was used to capture Gram-stained bacterial images. PSM comprises three primary components: lens, slide holder, and stage, as depicted in Fig. 2. The stage

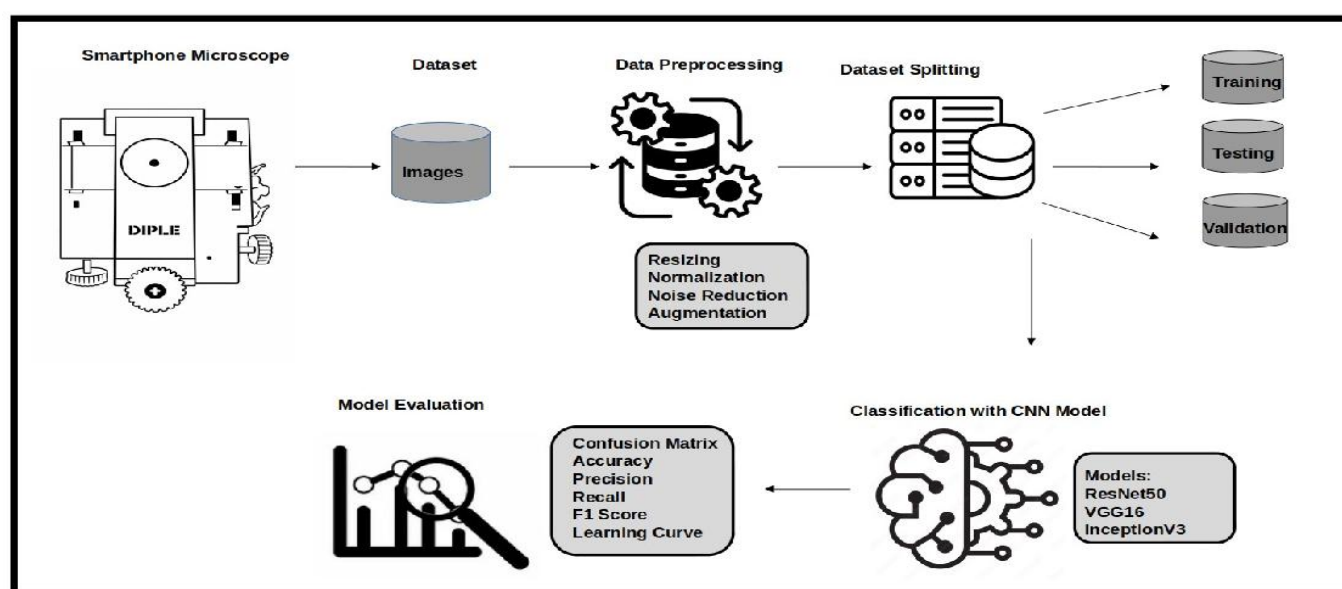


Fig. 1. Proposed methodology and overall working architecture

2. Materials and Methods

2.1 Materials

For bacterial dataset development, the Gram stain technique was used, and pure cultures of *Escherichia coli*, *Salmonella typhi*, and *Staphylococcus aureus* were used. The staining and imaging procedures used included a primary stain (crystal violet), mordant (iodine of gram), ethanol to decolorize followed by a counterstain (safranin) to prepare the slide.

2.2 Methodology

The methodological framework is shown in Fig. 1 and comprises the major steps of systematic bacterial classification using deep learning approaches. It began the work using a portable smartphone microscope to visualize Gram-stained bacteria.

accommodates the slide platform, objective lens (diple grey), light source, battery holder, and switch, and features a manual mechanical system facilitating precise specimen positioning through fine two-axis movement of the slide. The PSM has a magnification of 750x, a resolution of 1 μ m, a field of view of 0.7 mm, and a working distance of 0.6 mm. The overall hardware cost is 50\$.

PSM is designed with a simple, easy-to-use operating system. Working mechanism shown in Fig. 2: Turn on the light, place the sample slide above the illumination hole, place the objective lens (diple grey) in its assigned slot, and then set and align the smartphone camera with the lens. The slide can then be adjusted using the gears (stage wheels), and a clear image is obtained by rotating the focusing screw to achieve the final focus.

2.4 Bacterial Dataset Development

Bacteria were stained using Gram staining and imaged with a Samsung S22 Ultra smartphone under identical lighting and focal conditions to maintain

classification of the bacteria i.e *Escherichia coli*, *Salmonella typhi*, and *Staphylococcus aureus*. The data set consists of three species, which are organized in three subsets of a binary classification that have pairwise: *Escherichia coli* versus *Staphylococcus aureus* (600 images); *Salmonella typhi* versus

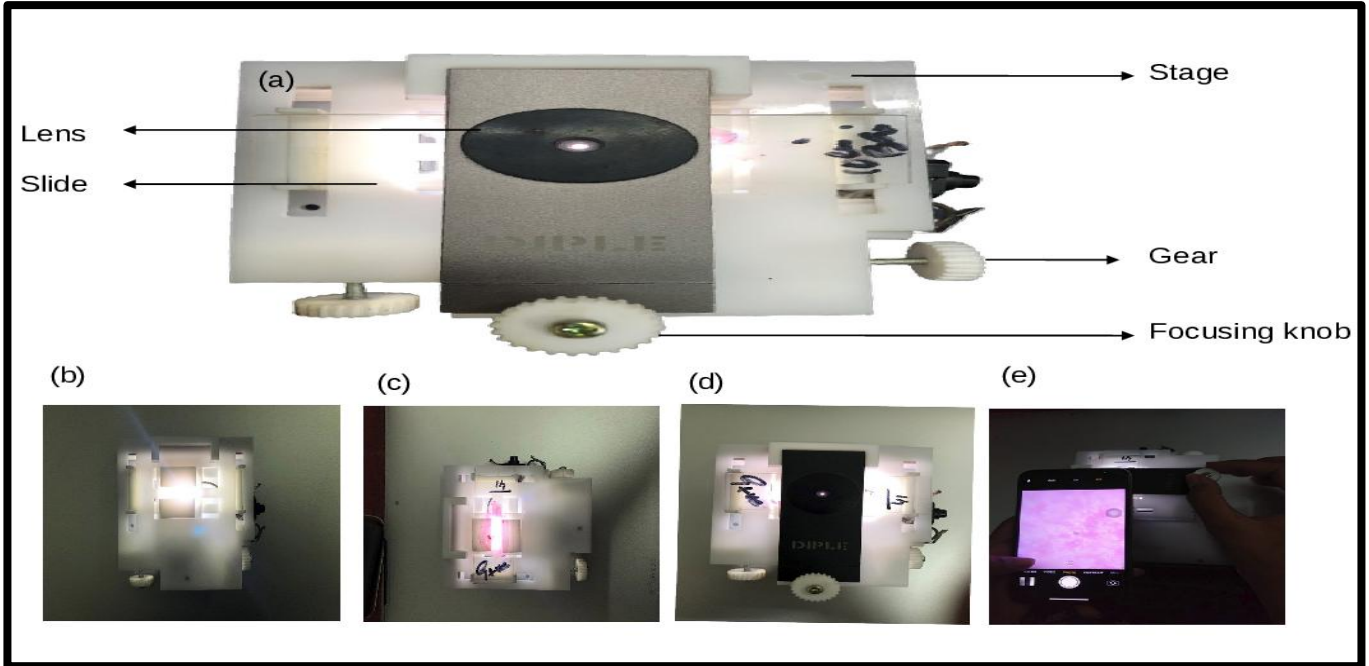


Fig. 2. PSM, (a) supporting stage, slide holder, lens, mechanical system with gears and focus knob, (b) Light source turn on (c) Sample slide placement (d) Objective lens dipple grey fixation (e) Adjustment of mobile phone, lens and slide for image capturing

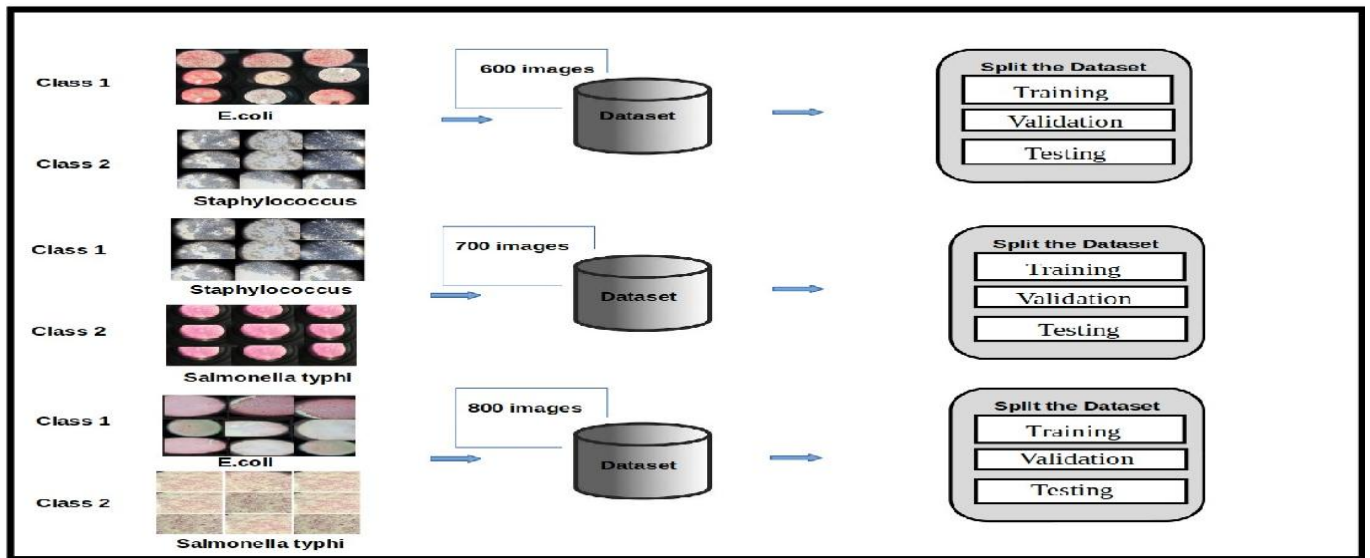


Fig. 3. Splitting of the Bacterial Image Dataset captured through PSM

consistency. For Gram staining, the smear was fixed on glass slide by applying heat, stained for 1 minute with crystal violet, treated for 1 minute with Gram's iodine, decolorized with ethanol for 10–15 seconds, counter stained for 30 seconds with safranin, and observed under oil immersion smartphone microscopy. A feature engineered comprehensive image database of bacterial strains was developed from multiple independent samples of bacterial cultures to ensure dataset variability and support the

© Mehran University of Engineering and Technology 2026

Staphylococcus aureus (700 images); and *Escherichia coli* versus *Salmonella typhi* (800 images). All images were annotated by hand (based on their bacterial class), and the annotations were encoded through a binary labelling scheme of each pairwise comparison. Stratified random sampling was used to divide the dataset into training (80%), validation (10%), and test (10%) subsets and ensure a balanced sample of all classes, as shown in Fig. 3. This measure reduced the leakage of data, and the distribution of data was

balanced when evaluating the model.

2.5 Bacterial Image Pre-processing

The subsequent phase entails pre-processing before being input into the CNN models. Image pre-processing includes resizing to 224 by 224 pixels for ResNet50 and VGG16, and 299 by 299 pixels for InceptionV3. Through contrast enhancement and colour normalization, features such as clustering and cell shape become more distinct. Image tagging, standardization, and augmentation (rotation, brightness adjustment, flipping, and zooming) were used to improve input quality for model training.

2.6 Models' architecture and transfer learning

Three deep transfer learning models, InceptionV3, ResNet50, and VGG16, were used in this research to classify bacterial images. The architectural details of the three models are shown in Fig. 4. The methodology for their operation is as follows.

(3×3 convolutions) to detect medium-level features such as cell groupings and bacterial boundaries, and (5×5 convolutions) to extract broad patterns, such as background morphology or colony density. Maximum and average pooling layers are used to reduce redundant information and summarize spatial relationships. The outputs from these convolutions are concatenated, enabling the model to analyse bacterial images at multiple spatial scales simultaneously, which is crucial for distinguishing subtle variations in shape and staining between bacterial types. Successive inception blocks are used to obtain hierarchical feature learning. The first layers isolate the primitive visual representations, like colour intensity of Gram stain images, outlines of shapes, and edges. Later intervening steps recognize bacterial morphology, such as whether they are rods, as in the bacterium *Escherichia coli*, or cocci, as in the bacterium *Staphylococcus aureus*. These spatial patterns are combined into a more detailed layer to model more

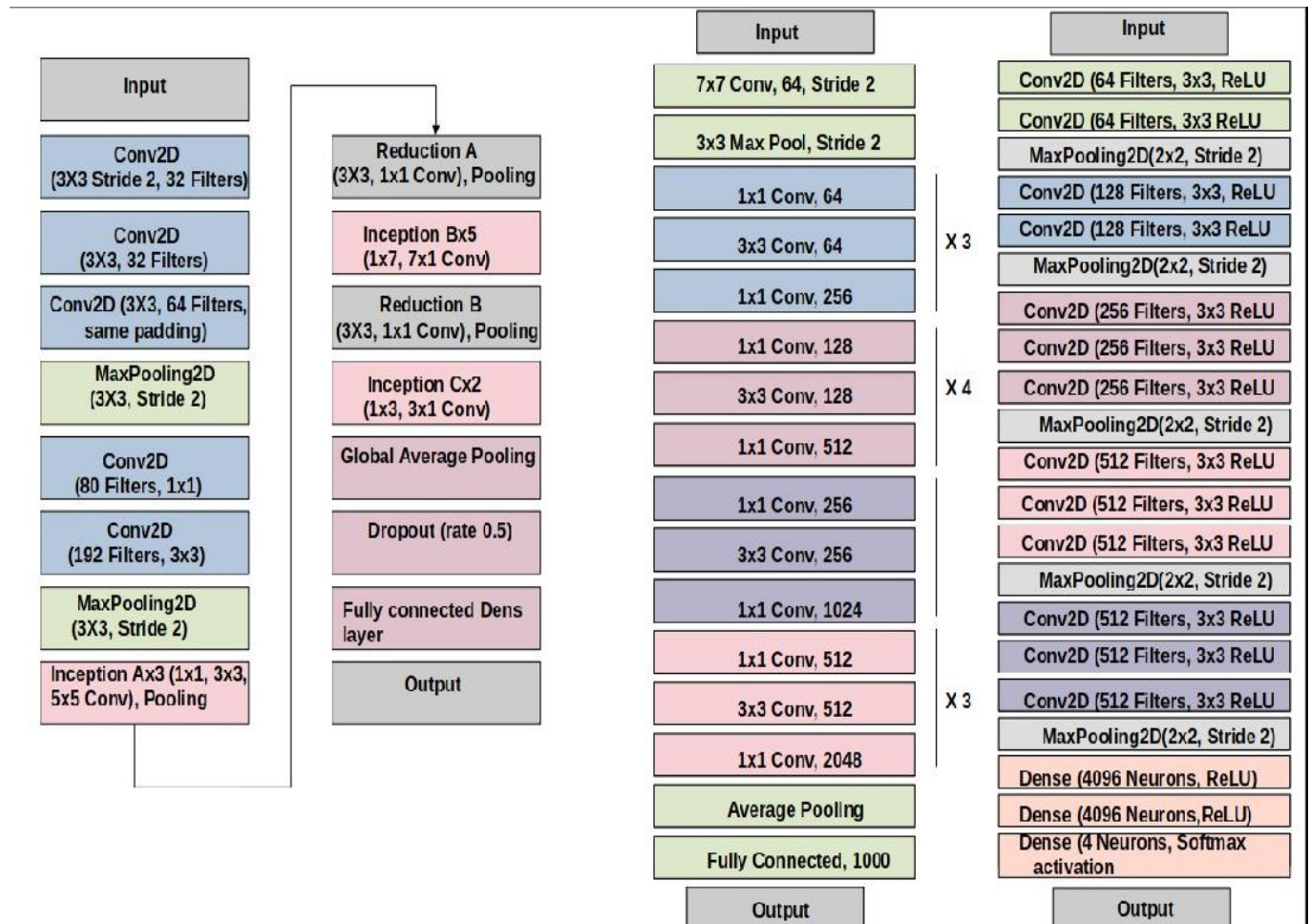


Fig. 4. InceptionV3, ResNet50 and VGG16 architecture

2.6.1 Inception V3

The InceptionV3 architecture uses a series of specialized Inception modules to extract large- scale and fine-grained morphological features from strained bacterial images. Each Inception module performs the following in parallel: (1×1 convolutions) to capture texture patterns and local colour in stained regions,

complex abstractions, such as colony organization and textural variation, thereby enabling accurate differentiation across species. The resulting convolutional features are pooled using global average pooling, which helps reduce overfitting by replacing large dense layers with averaged feature maps. The resulting pooled vector is then sent to a fully connected

(dense) layer, with a SoftMax activation, which produces probabilistic scores of each bacterial class.

2.6.2 ResNet50

The ResNet50 architecture contains 50 layers with residual connections that prevent vanishing gradients. Each residual block learns to refine bacterial features, such as *Escherichia coli* (rod-shaped bacteria), *Staphylococcus aureus* (clustered cocci), and *Salmonella typhi* (short rods with specific staining). Using initial convolutional and pooling layers, low-level features, such as colour gradients and edges, were extracted. Deeper residual layers identify complex bacterial morphology, such as spatial patterns and texture. Global average pooling converts feature maps into compact vectors. A fully connected layer followed by a softmax function outputs probability scores for each bacterial class. The highest-probability label is predicted.

2.6.3 VGG16

The VGG16 model consists of 13 convolutional layers grouped into five blocks. Each convolutional layer applies multiple 3×3 filters that slide over the input image to extract high- and low-level features. Blocks 1 and 2 capture low-level features, such as colours, edges, and textures. Blocks 3–5 capture complex bacterial morphology, including colony boundaries, shape, and internal structures. After each convolution, Rectified Linear Unit (ReLU) activation is applied to introduce nonlinearity, followed by max pooling to retain important features and reduce spatial dimensions. After feature extraction, the 3D feature maps are flattened into a 1D vector. This vector is passed through three dense (fully connected) layers. The first two layers consist of neurons each, applying learned weights to combine features. The final dense layer outputs a probability distribution across bacterial classes using a softmax activation function.

2.7 Models Training and Evaluation

For training, the Lambda Tensor book was used, a high-performance Linux-based laptop designed for deep learning. The specifications of the Lambda Tensor book are Processor (Intel Core i7-11800H, 11th Gen, 8 cores, 16 threads), GPU (NVIDIA GeForce RTX 3080 Max-Q), Memory (64 GB DDR4), and Storage (2 TB). All three models were trained using TensorFlow, NumPy, Keras, Pandas, and Seaborn. Important training parameters include categorical cross-entropy (Loss Function) and the Adam optimizer with an adaptive learning rate. Typically trained with a batch size of 32 for 50 epochs. Batch normalization and dropout were applied to enhance generalization. During training, backpropagation adjusts the network weights to

maximize bacterial classification accuracy and minimize prediction errors. The classification performance is evaluated using confusion matrices, learning curves, accuracy, precision, recall, and F1-score metrics.

3. Results and Discussion

3.1 *Escherichia coli* and *Staphylococcus aureus*

As shown in the confusion matrix in Fig. 5 (a), when the model was used to classify between the two bacteria, namely *Escherichia coli* and *Staphylococcus aureus*, it achieved high classification accuracy. Twenty-two *Escherichia coli* images were correctly identified, with only 1 instance incorrectly classified as *Staphylococcus aureus* out of the total test samples. In the case of the *Staphylococcus aureus* class, there were no false-negative encounters, indicating that all samples were properly classified. These findings suggest that the separation of the samples into classes is excellent, and the rates of misclassification are incredibly low, even in situations where morphological distinctions between the pictures under the microscope may arise.

As shown in Fig. 5 (b), training and validation loss curves overlap very quickly in the initial epochs. In the initial few epochs, the training loss plummeted to near zero, indicating effective representation of features learning by InceptionV3. The validation loss continues to decrease throughout the training process, indicating strong generalization. There is a periodic deviation in the loss around mid-training epochs, but no sustained separation between the training and validation losses. This behaviour has been interpreted to indicate little overfitting, and the argument that the InceptionV3 architecture will learn well when presented with bacterial images captured with a portable smartphone microscope. The model is robust as shown by the training and validation accuracy curves in Fig. 5 (c). The training accuracy increases from about 65% in the early epochs to 100% on validation, which remains constant and best in the last epochs, 98 to 99%. The slight difference between the validation and training accuracies indicates strong generalisation power. During the training period, a temporary decrease in the middle of the training is followed by a rapid recovery; this is due to changes in the optimisation process, not instability. Besides, the model has high predictive consistency and learning efficiency on unknown data.

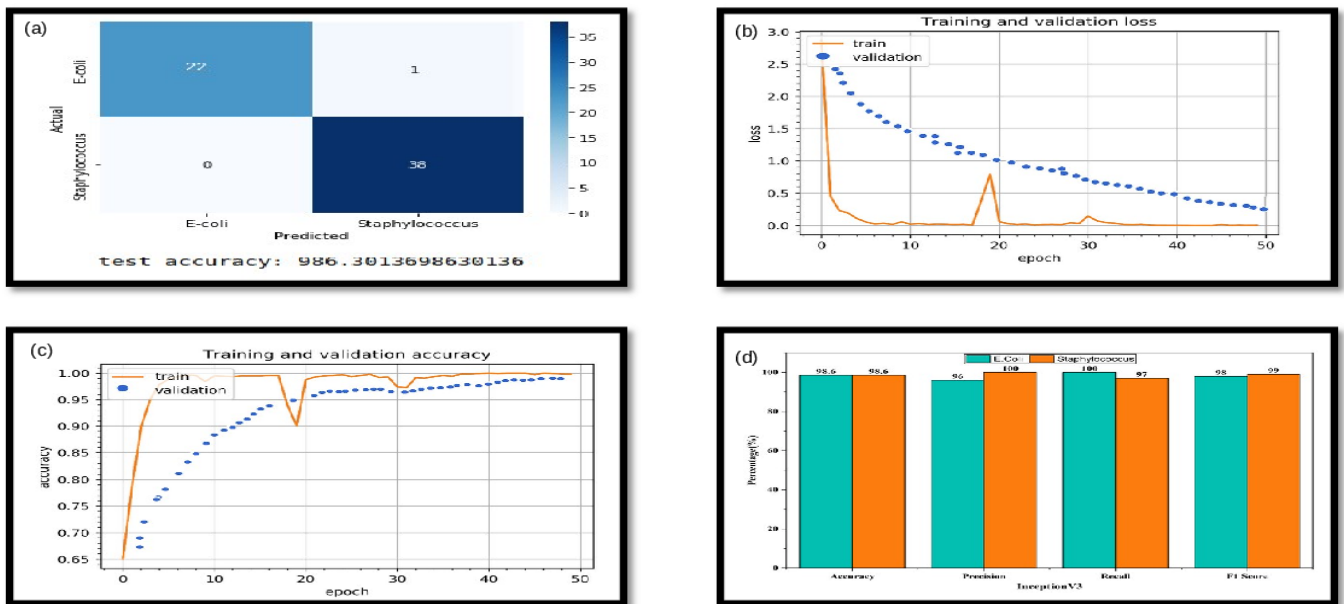


Fig. 5. (a) Confusion matrix, (b) training and validation loss curves, (c) training and validation accuracy curves, and (d) evaluation matrices for PSM and InceptionV3 of two bacterial classes, i.e., *Escherichia coli* and *Staphylococcus aureus*

Fig. 5 (d) presents quantitative performance metrics that demonstrate the effectiveness of the systems. For either class, the combined test accuracy is about 98.6%. In the case of *Escherichia coli*, InceptionV3 had an F1-score of 96, with 100% precision, 98% recall, and 98% F1-score, respectively. In the case of *Staphylococcus aureus*, the accuracy is 100%, the recall is 97%, and the F1-score is 99%, indicating a high ability to detect and a very low false-positive rate. The values of F1-score are high and balanced, indicating the best trade-offs between recall and accuracy. The combination of InceptionV3 with smartphone microscopy makes it a fast, dependable, and highly precise method for classifying individual

bacterial species, making it suitable for real-time field data collection and microbial diagnosis.

Fig. 6 (a) illustrates the ResNet50 model's classification performance, as shown in the confusion matrix for distinguishing between *Escherichia coli* and *Staphylococcus aureus*. The correct identification of 21 of 61 test images resulted from the presence of 21 images of *Escherichia coli*, and the incorrect identification of 2 images as *Staphylococcus aureus*. There were no missed samples; each of the 38 samples with *Staphylococcus aureus* was named correctly, and the accuracy is 96.72% in total. The matrix has high true positive values across the classes, and the

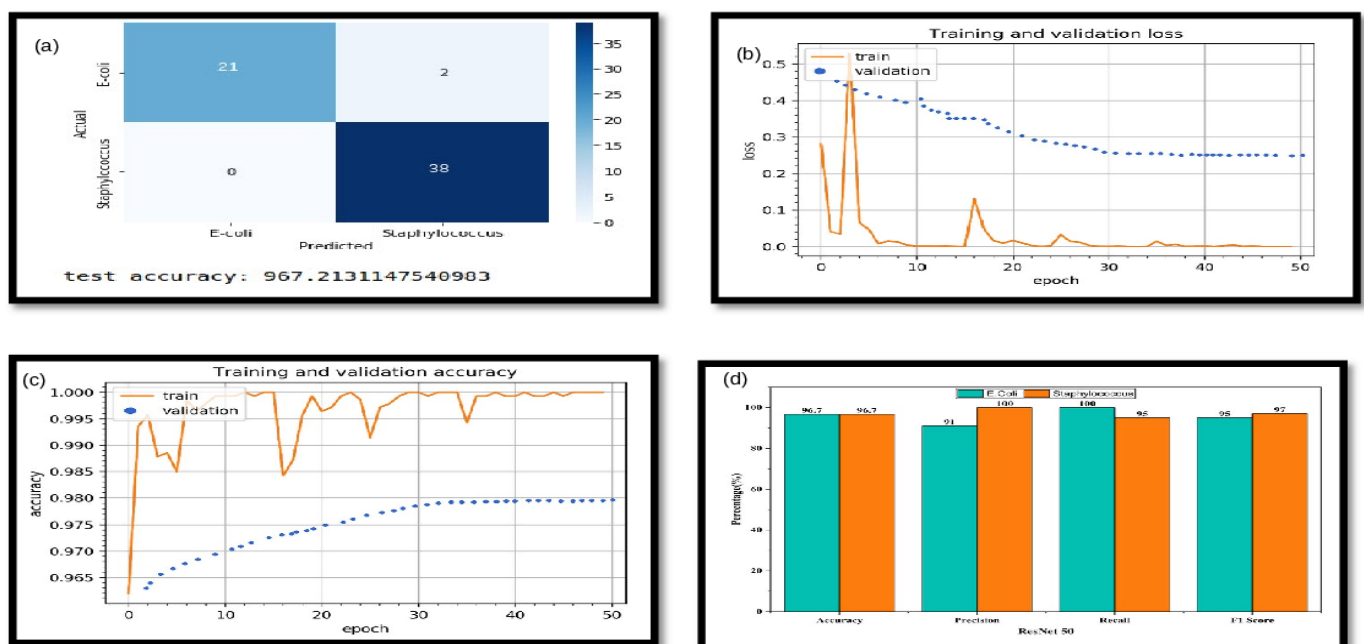


Fig. 6. (a) Confusion matrix, (b) training and validation loss curves, (c) training and validation accuracy curves, and (d) evaluation matrices for PSM and ResNet50 of two bacterial classes i.e., *Escherichia coli* and *Staphylococcus aureus*

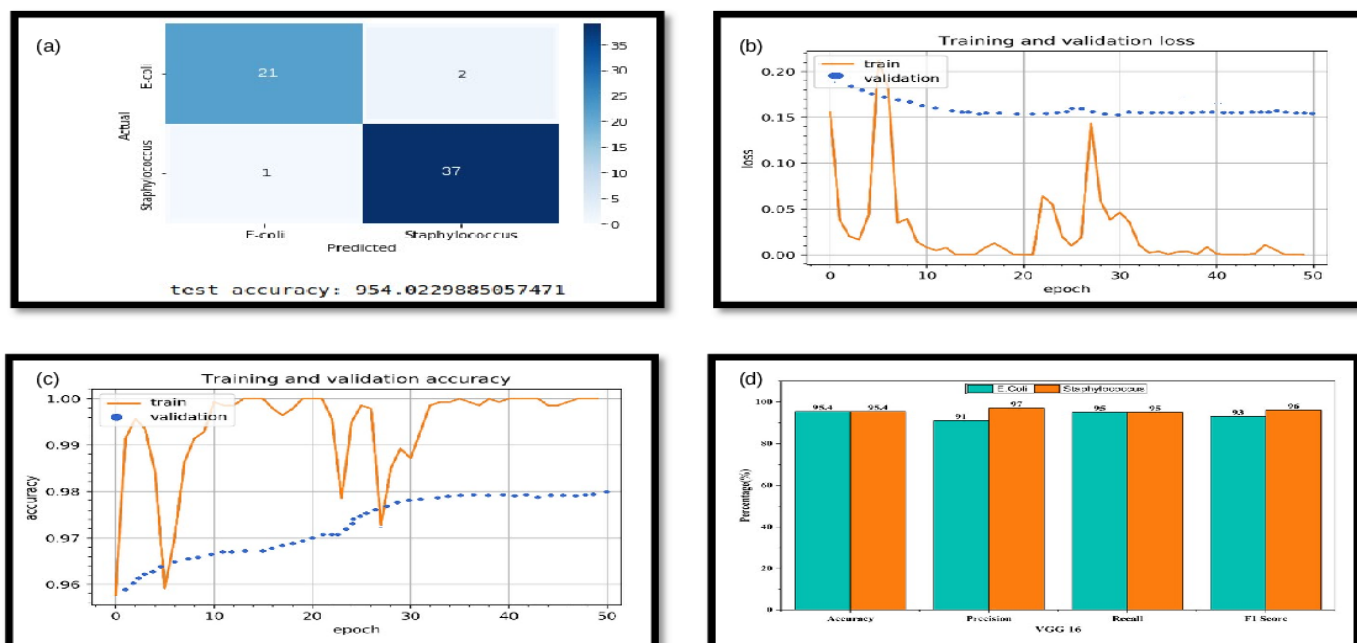


Fig. 7. (a) Confusion matrix, (b) training and validation loss curves, (c) training and validation accuracy curves, and (d) evaluation matrices for PSM and VGG16 of two bacterial classes i.e., *Escherichia coli* and *Staphylococcus aureus*

diagonal presence is very high.

In Fig. 6 (b), the validation loss is reduced steadily, with no sharp spikes, whereas the training loss is driven to zero quickly. The small generalisation gap shows that ResNet50 is an effective extractor of textural and structural characteristics of bacterial images and has a strong generalisation accuracy.

As illustrated in Fig. 6 (c), ResNet50 training is stable and smooth, enabling easy gradient propagation, which can be attributed to the residual learning structure. The validation accuracy decreases

gradually but stabilises at 98%, while the training accuracy reaches 100%.

Fig. 6 (d) evaluation metrics supporting the reliability and robustness of the proposed system. For *Escherichia coli*, the ResNet50 model achieved a precision of 91%, a recall of 100%, and an F1-score of 95%, indicating excellent specificity with few false negatives. The precision rate for *Staphylococcus aureus* was 100%, recall was 95%, and F1-score was 97%, demonstrating extremely low false positive rates and perfect sensitivity. The well-balanced and high F1-scores for both bacterial classes show a successful

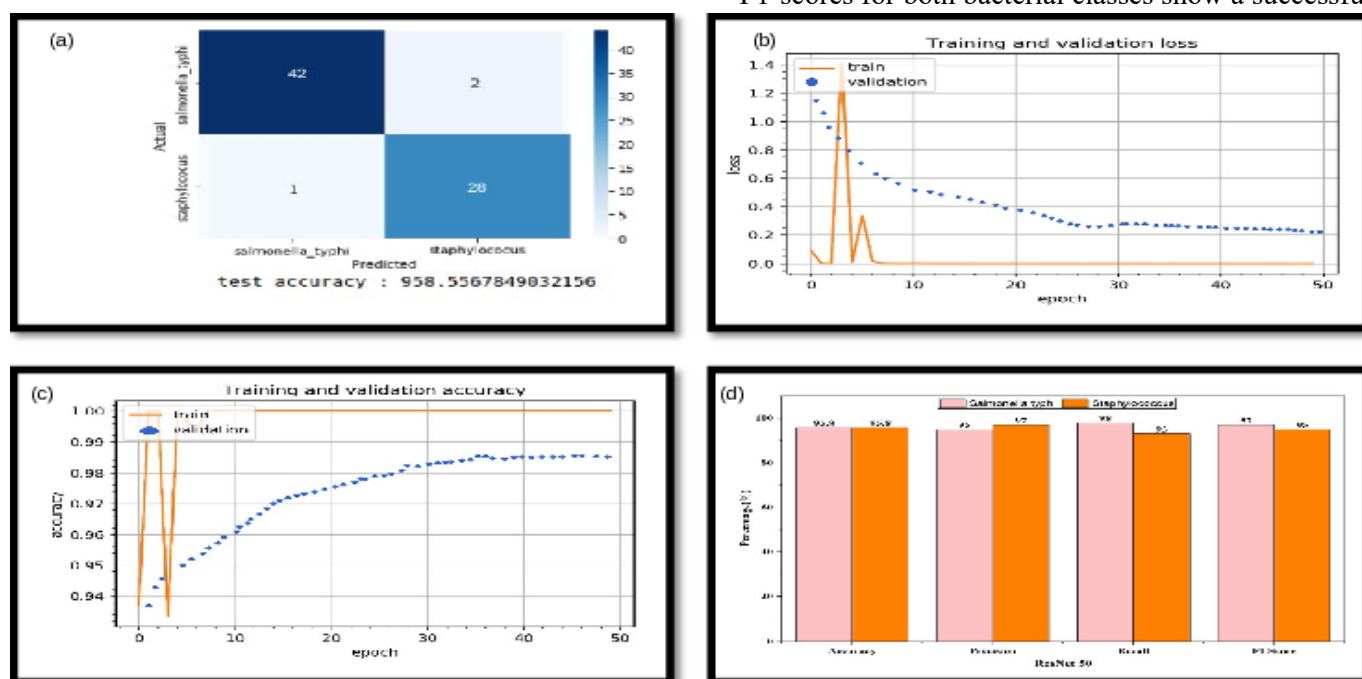


Fig. 8. (a) Confusion matrix, (b) training and validation loss curves, (c) training and validation accuracy curves, and (d) evaluation matrices for PSM and InceptionV3 of two bacterial classes i.e., *Salmonella typhi* and *Staphylococcus aureus*

trade-off between recall and precision. These results demonstrate that integrating a portable smartphone microscope with ResNet50 enables precise and accurate bacterial classification and confirms its applicability for field-based rapid microbial diagnostics.

Fig. 7 (a) Confusion matrix of the VGG16 model demonstrates strong but slightly less classification performance than Inceptionv3 and ResNet50 models. For *Escherichia coli*, out of 23 predictions, 22 were correctly classified, while 2 samples were incorrectly classified as *Staphylococcus aureus*. For *Staphylococcus aureus*, 39 samples were evaluated, with 38 accurately identified and 1 incorrectly identified, attaining an overall classification accuracy of 95.4%. The classification performance of model VGG16 indicates a high but relatively moderate discriminating capability. Despite inter-class confusion, the confusion matrix reveals strong diagonal dominance.

Fig. 7 (b) The validation loss fell to a very low and stable level, and the training loss approached near zero rapidly, indicating high classification confidence. The small difference between the validation and training metrics confirms that VGG16 performed well, even with its relatively simple architecture.

Fig. 7 (c) shows that at the beginning of the learning process, the model VGG16 achieved a steady increase in validation accuracy, stabilizing at approximately 98%. Although slightly slower convergence than

InceptionV3, the validation curve indicates smooth progression with minimal fluctuations, showing reliable learning behavior.

Fig. 7 (d) evaluation metrics, which provide additional support regarding the effectiveness of the VGG16 model. For *Escherichia coli*, the model achieved a precision of 91%, a recall of 95%, and an F1-score of 93%, with some false negatives indicating good predictive reliability. The precision rate for *Staphylococcus aureus* was 97%, while the recall reached 95% and the F1-score was 96%, demonstrating strong sensitivity and well-balanced performance. For both classes, the closely aligned, relatively high F1-scores indicate a stable balance between precision and recall. Overall, VGG16 delivers good classification performance, with 95.4% accuracy, slightly lower than the other two models, which Fig. 7. (a) Confusion matrix, (b) training and validation loss curves, (c) training and validation accuracy curves, capture finer, more detailed features for improved bacterial identification.

3.2 *Salmonella typhi* and *Staphylococcus aureus*

Fig. 8 (a) The confusion matrix produced by the InceptionV3 model exhibits excellent classification performance in distinguishing *Salmonella typhi* and *Staphylococcus aureus*. Forty-three *Salmonella typhi* samples were accurately classified (predicted total = 44), with only 1 incorrectly identified. The results indicate that all *Salmonella typhi* samples were accurately classified, with a low false-positive rate. In

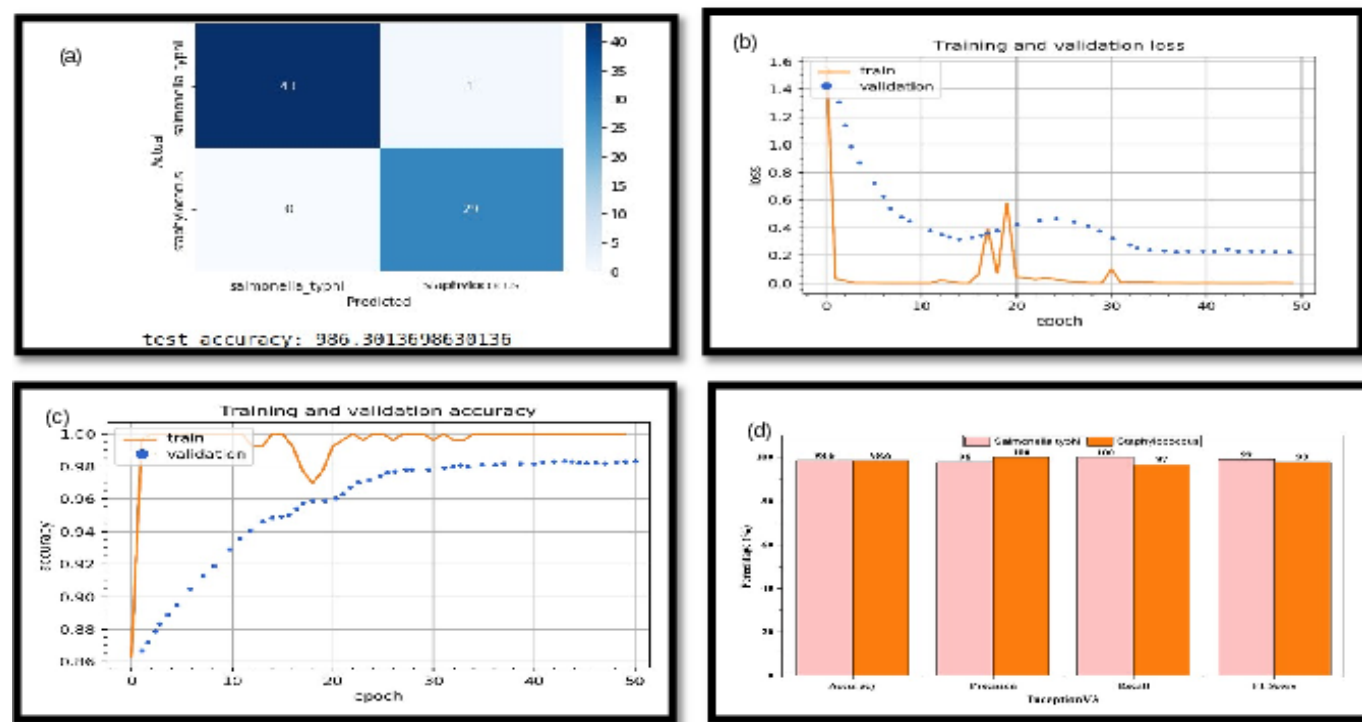


Fig. 9. (a) Confusion matrix, (b) training and validation loss curves, (c) training and validation accuracy curves and (d) evaluation matrices for PSM and ResNet50 of two bacterial classes i.e., *Salmonella typhi* and *Staphylococcus aureus*

Staphylococcus aureus, 29 out of 30 were correctly classified with only 1 incorrect classification, in the confusion matrix. The confusion matrix of the model Inceptionv3 shows strong diagonal dominance and minimal inter-class confusion.

Fig. 8 (b) shows that within the first few epochs, the training loss decreased sharply to near zero. In contrast, the validation loss decreased steadily with minimal fluctuations throughout the mid-training stage. However, a slight difference between training and validation loss is observed. Still, validation loss stabilizes at a low value, suggesting that InceptionV3 effectively learned distinguishing features from images of *Salmonella typhi* and *Staphylococcus aureus*.

Fig. 8 (c) the training accuracy of InceptionV3 increased quickly and approached near 100% within the initial epoch and remained stable throughout the training process. A steady upward trend is shown in the validation accuracy gradually converging to approximately 98%. In later epochs, the small gap between training and validation accuracy indicates minimal over-fitting and good generalization.

Fig. 8 (d) shows that the model achieved an overall classification accuracy of 98.6%, as indicated by the evaluation metrics, confirming the reliability and robustness of the InceptionV3 model. For *Salmonella typhi*, the F1-score of 99% suggests an optimal balance between precision (98%) and recall (100%). The F1-score for *Staphylococcus aureus* was 98%, indicating

equally strong performance with 100% perfect precision and near-perfect recall of 97%. The InceptionV3 architecture effectively extracts key morphological features from microscopic images captured on smartphones, as indicated by these high and balanced performance metrics. The results show that a portable smartphone microscope integrated with InceptionV3 provides highly accurate and reliable classification of *Salmonella typhi* and *Staphylococcus aureus* and validates its usability for rapid, field-deployable microbial diagnostics.

Fig. 9 (a) The confusion matrix of the ResNet50 model indicates strong classification performance, but slightly lower than InceptionV3. Forty-three of 44 predictions were correctly classified, with only 1 incorrectly classified. These results show that all true *Salmonella typhi* samples were successfully identified with a minimal number of false positives. In the *Staphylococcus aureus* samples, 29 out of 30 were correctly classified, and 1 was incorrectly classified. The matrix shows strong diagonal dominance, with slightly higher inter-class confusion than InceptionV3.

Fig. 9. (b) Initially, the training loss approached zero, and the validation loss indicates a smooth downward trend with minimal fluctuation. The small gap between the training and validation curves indicates that, despite limited overfitting, the ResNet50 model achieved robust generalization, making it well-suited for bacterial classification.

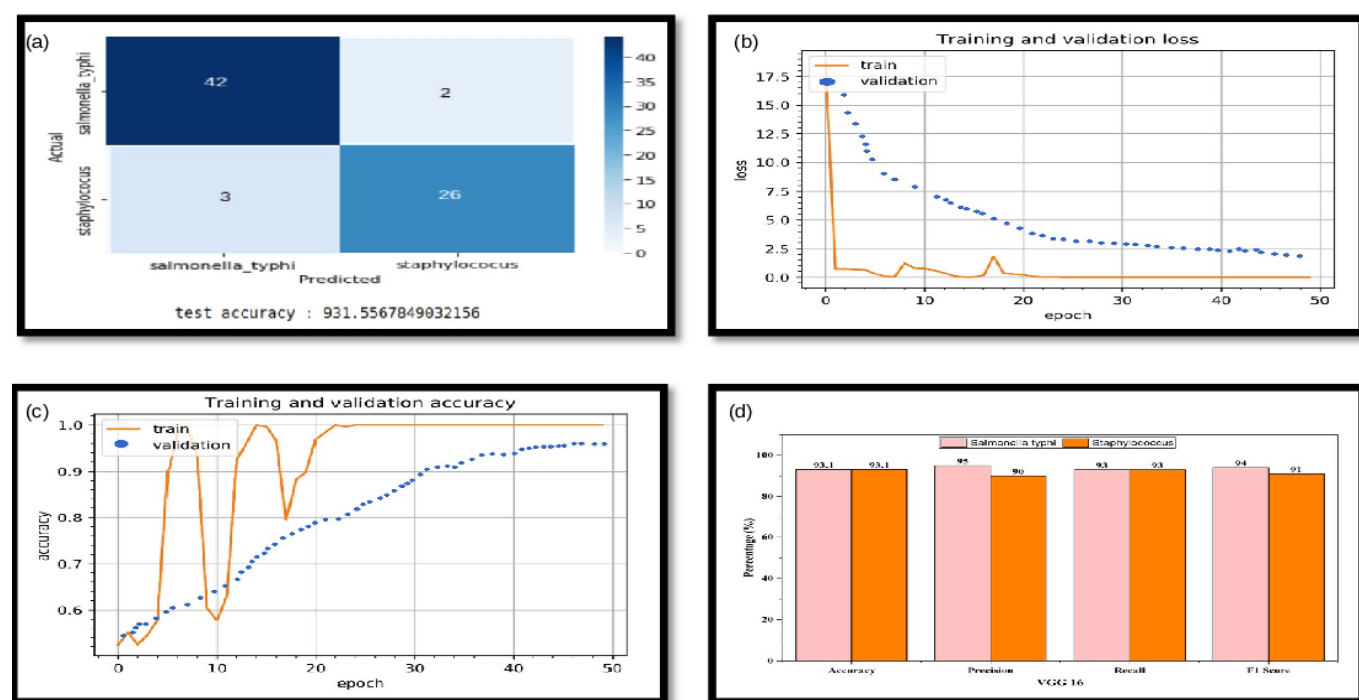


Fig. 10. (a) Confusion matrix, (b) training and validation loss curves, (c) training and validation accuracy curves, and (d) evaluation matrices for PSM and VGG16 of two bacterial classes i.e., *Salmonella typhi* and *Staphylococcus aureus*

Fig. 9 (c) shows that ResNet50 also performed strongly. The training accuracy rapidly reaches 100% and remains constant, while validation accuracy continues to improve and stabilizes at 98%. The convergence pattern shows stable learning dynamics resulting from residual connections that enable gradient flow.

Fig. 9 (d) shows the reliability of the ResNet50 model as measured by the evaluation metrics. For *Salmonella typhi*, the F1-score of 97% demonstrates a balanced trade-off between precision (95%) and recall (98%). The F1-score for *Staphylococcus aureus* was 95%, presenting high predictive reliability despite a slight reduction in recall. The overall accuracy of 95.8% indicates that ResNet50 captures and extracts more discriminating, fine morphological variations. For rapid results, the ResNet50 model is well-suited to field-based microbial classification applications.

Fig. 10 (a) shows the confusion matrix of VGG16 for the classification of *Salmonella typhi* and *Staphylococcus aureus*, which exhibits a robust capability with a minimal rate of incorrect classification. The matrix indicates that 42 *Salmonella typhi* samples were accurately classified, while only 2 were misclassified; for *Staphylococcus aureus* samples, 26 were correctly predicted, with 3 instances of misclassification. Approximately 93.15% overall test accuracy was achieved, indicating that the VGG16 model performed well at differentiating between the two morphological distinct bacterial classes; however,

Fig. 10 (b) shows the training and validation loss curves, which substantiate the model's robustness. The training loss declines sharply in the first epoch and approaches nearly zero by the final epoch, indicating successful convergence and feature learning. The validation loss gradually diminishes from a high initial value and stabilizes around 2 in later epochs. Moreover, a gap exists between the training and validation losses; the steady decline in validation loss indicates minimal overfitting and stable generalization.

Fig. 10 (c) shows steady performance improvement, as evidenced by the training and validation accuracy curves. The training accuracy increased to nearly 100%, demonstrating the impressive learning ability of the InceptionV3 model. Validation accuracy consistently progressed across epochs, eventually reaching approximately 95-96% by the final epoch. In the later stages, the close alignment between training and validation accuracy indicates reliable model generalization to unseen data.

Fig. 10 (d) shows evaluation metrics demonstrate balanced but slightly reduced performance. For *Salmonella typhi*, the F1-score was 94%, indicating a strong balance between precision (95%) and recall (93%). The F1-score for *Staphylococcus aureus* was 91%, influenced by lower precision (90%) despite stable recall (93%). These results show that VGG16 effectively extracts relevant morphological features from smartphone-based microscopic images; it is

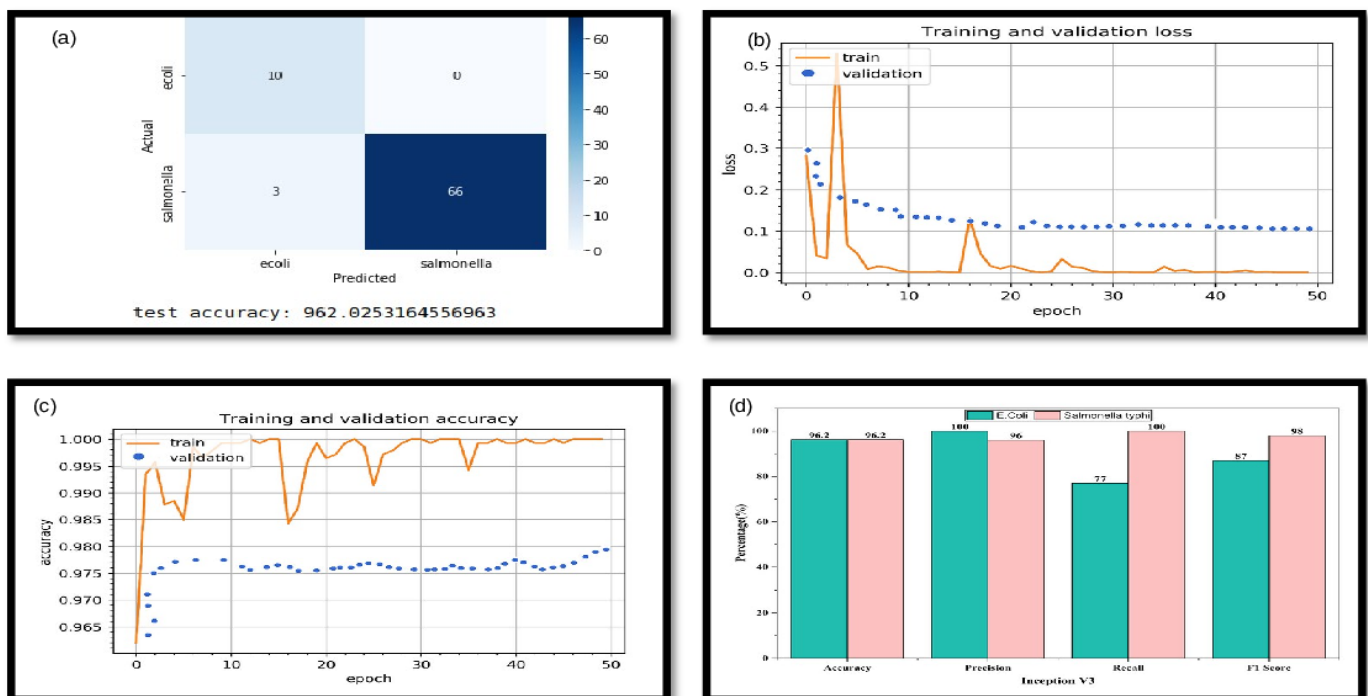


Fig. 11. (a) Confusion matrix, (b) training and validation loss curves, (c) training and validation accuracy curves, and (d) evaluation matrices for PSM and InceptionV3 of two bacterial classes i.e., *Escherichia coli* and *Salmonella typhi*

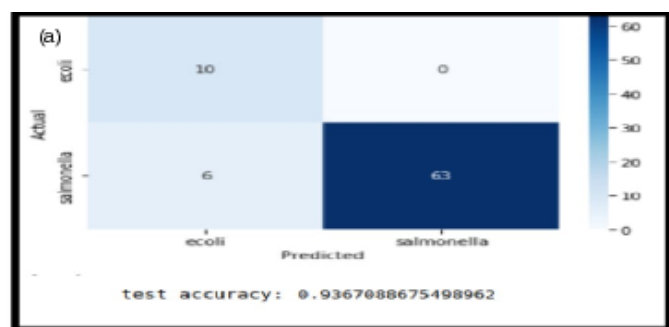
a small degree of inter-class confusion persisted.

somewhat less effective at capturing structural similarities between the two bacterial species. Overall,

achieving 93.1% reliable classification accuracy, these findings imply that more complex and deeper architectures may offer enhanced feature representation and higher diagnostic reliability for field-based microbial identification.

3.3 *Escherichia coli* and *Salmonella typhi*

Fig. 11 (a) The confusion matrix shows the strong



transfer learning. After a progressive increase, the validation accuracy stabilized at 98%.

Fig. 11 (d) further confirms the robustness of the model through the evaluation metrics. The model achieves an overall balanced high accuracy of 96.2%. For the *Escherichia coli* class, the precision was 100%, indicating no false-positive predictions. The

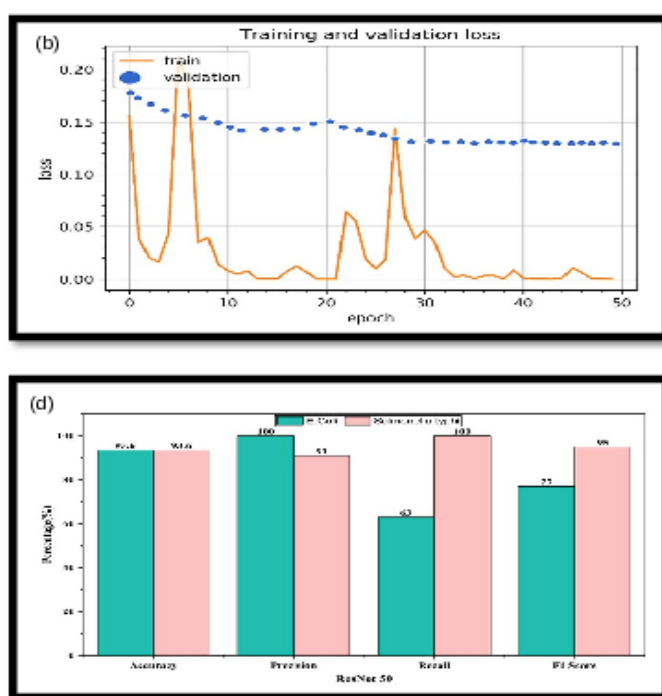


Fig. 12. (a) Confusion matrix, (b) training and validation loss curves, (c) training and validation accuracy curves, and (d) evaluation matrices for PSM and ResNet50 of two bacterial classes i.e., *Escherichia coli* and *Salmonella typhi*

identification capability of InceptionV3 in *Escherichia coli* and *Salmonella typhi*. Out of 13 samples of *Escherichia coli*, 10 were accurately identified, whereas 3 were incorrectly classified as *Salmonella typhi*. Sixty-six *Salmonella typhi* samples were evaluated, and 69 were predicted, indicating some false positives. The overall accuracy of 96.2% for InceptionV3 indicates minimal inter-class confusion and high model reliability.

Fig. 11 (b) shows the validation loss steadily decreasing and then plateauing at a low value. The relatively small gap between training and validation performance demonstrates minimal overfitting and good generalization. The strong discriminative capability and stable optimization of InceptionV3 are reflected in the smooth trends of both the accuracy and loss curves.

Fig. 11 (c) for InceptionV3 shows training and validation curves indicating well-balanced, stable learning behavior for classifying *Escherichia coli* and *Salmonella typhi*. In early epochs, the training loss converges to near zero, and the training accuracy rapidly approaches 100%, demonstrating fast convergence and efficient feature extraction through

recall of 77% yielded an F1-score of 87% for detecting all *Escherichia coli* samples, indicating moderate sensitivity. For *Salmonella typhi*, precision was 96%, recall 100%, F1-score of 98%, minimal false negatives, and excellent classification stability. Overall, the results show that InceptionV3 efficiently extracts morphological features from smartphone-acquired microscopic images and maintains high precision for *Escherichia coli* and excellent performance in detecting *Salmonella typhi*.

Fig. 12 (a) represents the strong classification performance of ResNet50 for *Escherichia coli* and *Salmonella typhi*. With zero incorrect classification for *Escherichia coli*, 10 actual samples were correctly classified. Sixty-three *Salmonella typhi* samples were accurately predicted, with 6 incorrectly classified as *Escherichia coli*. An overall test accuracy of 93.67% represents high discrimination between these two bacteria, whereas in the *Salmonella typhi* class, minor confusion was observed.

Fig. 12 (b) shows that ResNet50 achieved perfect training accuracy, confirming its strong capability to distinguish and classify two bacterial classes, with a training loss of zero. Validation loss remains

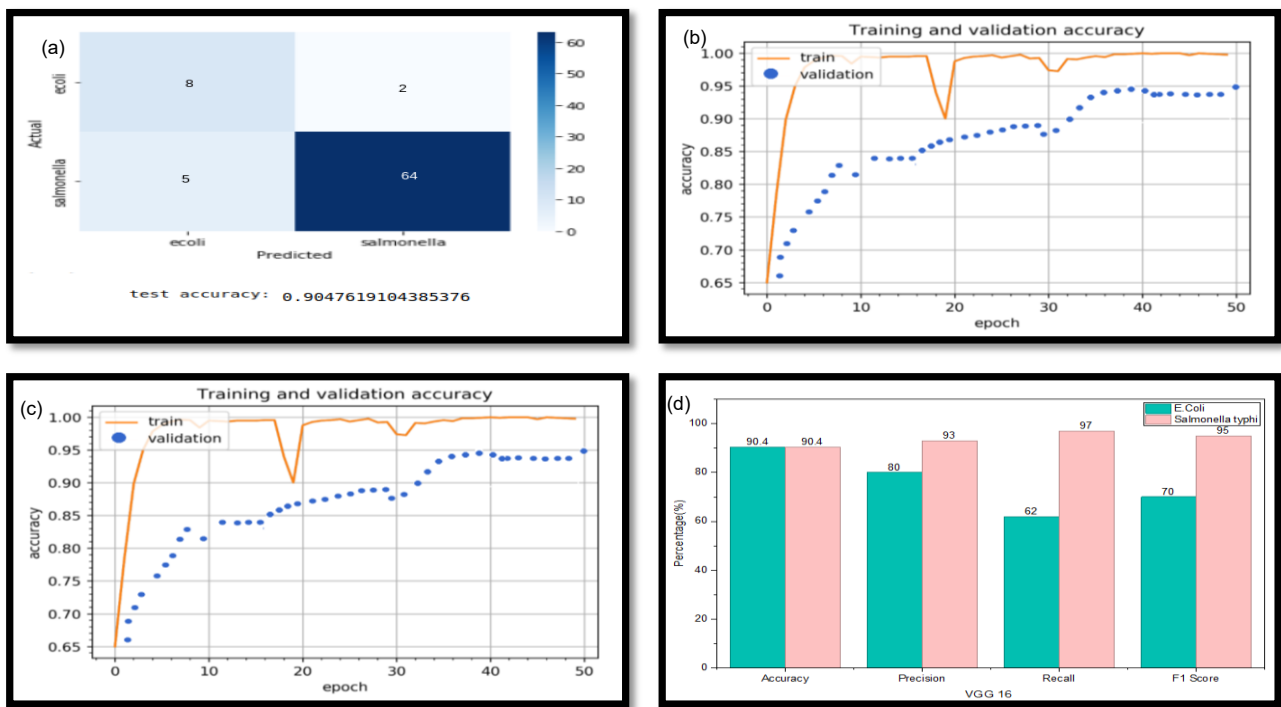


Fig. 13. (a) Confusion matrix, (b) training and validation loss curves, (c) training and validation accuracy curves, and (d) evaluation matrices for PSM and VGG16 of two bacterial classes i.e., *Escherichia coli* and *Salmonella typhi*

moderately higher, and validation accuracy is lower than InceptionV3 and stabilized at 95%. Fig. 12 (c) shows the observed fluctuations in training, which reflect dataset variability and sensitivity to learning rate adjustments. The moderate gap between the training and validation curves indicates overfitting; however, the model still exhibits good generalization performance. The residual connections in ResNet50 facilitate robust feature learning and efficient gradient flow from smartphone microscope images.

The detailed quantification of model performance is presented in the evaluation metrics in Fig. 12 (d). For *Escherichia*, 100% precision yielded no false-positive predictions, whereas 63% recall indicates that some *Salmonella typhi* samples were predicted as *Escherichia coli*. The 77% F1-score for *Escherichia coli* indicates a moderate balance between precision and recall. *Salmonella typhi* achieved 100% recall and 91% precision, indicating that all *Salmonella typhi* samples were detected with a small number of false positives. For *Salmonella typhi*, the F1-score was 95%, indicating consistent, strong classification performance. Overall, 93.6% accuracy and F1-scores confirm that ResNet50 provides reliable bacterial differentiation, with robust *Salmonella typhi* detection, but slight recall limitations are observed for *Escherichia coli*.

Fig. 13 (a) The confusion matrix of VGG16 represents a noticeable class imbalance in predictive strength with moderate classification performance for *Escherichia coli* and *Salmonella typhi*. Of 13 actual *Escherichia coli* samples, 10 were correctly classified,

yielding 3 incorrect classifications. Sixty-six *Salmonella typhi* samples were evaluated, and 69 were predicted, indicating some false positive but good detection capability. The overall model accuracy was 90.4%, implying good performance and uneven class-wise discrimination.

Fig. 13 (b) and (c) show that the VGG16 architecture achieves 100% training accuracy and a very small training loss, which is like other models, further suggesting that the architecture is learning and that the training data are correctly distributed. However, validation accuracy remained relatively low, below 95%, and validation loss remained high throughout training. The high deviation between the training and validation curves is an indication of increased tendency toward overfitting. Unlike the convergence behavior of InceptionV3 and ResNet50, the VGG16 model showed slower validation convergence. There are no inceptions or residual modules, and the large number of parameters in VGG16 makes the model prone to poor generalization on the current dataset. VGG16, among the three analyzed architectures, showed comparatively lower validation metrics.

Fig. 13 (d) reports the evaluation measures and further explains why the class imbalance observed is the case. In the case of *Escherichia coli*, precision of 80%, recall of 62%, and an F1-score of 70% highlight how challenging it is to reliably identify all true cases of *Escherichia coli*, underscoring sensitivity in this class. By contrast, *Salmonella typhi* was classified to 93%, recalled to 97%, and F1-scored to 95% with a

statistically consistent outstanding detection, and a false-negative rate with a very high low value, respectively. The difference between the inter-classes suggests that VGG16 has better efficacy in detecting morphological attributes of *Salmonella typhi* than of *Escherichia coli*. The general model has reasonable accuracy, although additional feature discrimination for *Escherichia coli* is needed to achieve balanced diagnostic accuracy.

Comparative analysis shows that the classification of bacteria with discrete morphological features, compared to that of morphologically similar bacteria, in the considered models of InceptionV3, ResNet50, and VGG16, achieved higher overall accuracy. A comprehensive comparison with existing smartphone based diagnostic systems is presented. Although much of the available literature has focused on bacterial species detection, rather than multi-species classification, species differentiation using morphological patterns has been studied little. Many researchers have therefore used convolutional and deep convolutional neural networks in detecting pathogens and images. A convolutional neural network and a smartphone application was used in a study for the detection of *Escherichia coli* in water samples with an effective accuracy of around 96% [12]. T. Ahmed (2019) used the InceptionV3 structure in combination with a support vector machine to identify seven bacterial species with an accuracy of 96%. More robust decision boundaries and decreased reliance on large fully connected layers or big softmax classifiers are made easier by the hybrid InceptionV3 and SVM approach that pairs deep feature extraction with a classical machine-learning classifier [13]. Narin (2021) introduced an automated deep-learning COVID-19 detection system for chest X-ray images using several convolutional neural network backbones, such as ResNet50, ResNet101, ResNet152, InceptionV3, and Inception-ResNetV2. The proposed approach provided a detection rate of 99.73% with ResNet50 on various datasets but failed to perform as well when challenged by noisy and low-resolution X-ray images as well as heterogeneous images [14]. Arnob, A.S. (2025) used VGG16, InceptionV3, and ResNet50 to classify diseases affecting cauliflower plants in tropical environments, achieving 90.85% accuracy with ResNet50 [15]. ForouzeshFar, P. (2024). Used AlexNet, ResNet-50, VGG-16, and VGG-19: A CNN-based diagnostic model for the automatic detection of dental caries from bitewing radiographic images. The best-performing model, VGG19, achieves a high diagnostic accuracy of approximately 94% [16]. Gupta et al (2025) A comparison of three deep transfer-learning models is conducted for the purpose of classifying cotton leaf

diseases using ResNet50, VGG16, and InceptionV3, having 95.28% accuracy with InceptionV3, but ResNet50's poor performance suggests that more complex or deeper models may not always produce better outcomes in domain-specific data sets[17]. Nie et al. (2015) used a CNN with multiple convolutional layers, a deep learning framework for classification and segmentation, to classify bacterial colonies, achieving accuracy ranging from 52.5% to 78.47% [18]. Lopez et al. (2017) used a CNN to automatically classify light field smear microscopy images to detect *Mycobacterium tuberculosis*, achieving an accuracy of 96% [19]. Turra et al. (2017) used a 1D CNN on hyperspectral imaging (HSI) data to identify bacterial hyperspectral signatures, achieving an accuracy of 99.7% [20]. Zieliński et al. (2017) used this approach to classify bacterial colonies using deep learning, achieving 97.24% accuracy on the DIBaS data set for 33 bacterial species [21]. Panicker et al. (2018) detected tuberculosis bacilli using a deep learning algorithm with a CNN, achieving a recall of 97.13% and an F1-score of 86.76% [22]. In 2019, Traore et al, used CNN (for the recognition of *Vibrio cholera* and *Plasmodium falciparum* images) for the classification of epidemic pathogens, achieving 94% accuracy [23]. Hay et al. (2019) used 3D light-sheet fluorescence microscopy (CNN) to identify gut bacteria in larval zebrafish, achieving 90% accuracy on 3D microscopy datasets [24]. Mithra et al. (2019) Automated identification of *Mycobacterium bacilli in sputum images for tuberculosis diagnosis using a deep belief neural network*, achieving a detection accuracy of 98.23% [25]. Amri et al. (2023) evaluated and compared deep CNN architectures for bacterial colony classification, including a baseline CNN, a gated CNN (GCNN), ResNet-18, and VGG-1. Moreover, VGGNet consistently outperformed the other models (98.12% on the *Escherichia coli* data set and 96.15 % on the DIBaS data set) [26].

Although significant advancements have been made in deep learning applications for microbiology, limitations and drawbacks have emerged in the classification and detection of bacterial images. Advancements and studies were based on small-scale or laboratory-specific data sets with limited microscopy images. This limited their ability to characterize some bacterial species and strains. VGG16, ResNet50, and InceptionV3 were among the commonly employed models used under controlled magnification and lighting to compile the datasets. These controlled parameters render these models less versatile in bacterial classification applications. Studies such as those by Zieliński et al. (2017) and Amri et al. (2023), with accuracy exceeding 94%, still encountered issues such as dependence on expensive

microscopes, computational complexity, and data overfitting [21,26]. Studies by Lopez et al. (2017) and Panicker et al. (2018), which focused on tuberculosis detection, performed well but could not adapt to different pathogens or stain types [19,22]. Although advanced architecture and hyperspectral imaging were employed in studies by Nie et al. (2015) and Turra et al. (2017), they were limited to real-time functionality due to the need for specialized equipment and longer processing times [18,20]. Thus, previous studies were based on non-portable applications, data set limitations (small scale data sets), controlled imaging conditions, and low-resource settings, or on-site detection applications. Additionally, existing smartphone-based bacterial detection systems typically focus on single-species detection using fluorescence or microfluidic platforms, offering limited resolution and limited computational integration.

For bacterial image classification using the Portable Smartphone Microscope data set, InceptionV3 consistently outperformed ResNet50 and VGG16 across both similar and distinct morphological bacterial pairs. InceptionV3 achieved 98.6% accuracy for *Escherichia coli* and *Staphylococcus aureus*, followed by ResNet50 at 96.7% and VGG16 at 95.4%. Similarly, for *Salmonella typhi* and *Staphylococcus aureus*, InceptionV3 98.6%, ResNet50 95.8%, and VGG16 93.1%. The models reduced their performance by a small margin for *Escherichia coli* and *Salmonella typhi*, with InceptionV3 recording 96.2%, ResNet50 93.6% and VGG16 90.4%. InceptionV3 performed better on feature-extraction images and discriminating density for similar and different bacterial morphology, followed by ResNet50, even though VGG16, though achieving relatively worse results, could also be reliable. These results highlight the effectiveness of the designed system for rapid, field-deployable bacterial detection.

Conclusion

This research presents a portable and precise bacterial classification system that utilizes a smartphone microscope (1 μm , 750X magnification) with deep convolutional neural networks. Three architectures (InceptionV3, ResNet50, and VGG16) were tested to classify bacteria that are of environmental and clinical significance, namely *Escherichia coli*, *Salmonella typhi*, and *Staphylococcus aureus*. A framework with the highest accuracy (98.6%) was obtained with InceptionV3, which was higher than ResNet50 and VGG16 in discriminating morphological difference (*Escherichia coli* vs *Staphylococcus aureus*, *Staphylococcus aureus* vs *Salmonella typhi*) and morphologically similar (*Escherichia coli* vs

Salmonella typhi) species. The classification results of morphological different bacteria were found to be more accurate, and well-performing on morphological related bacteria highlights the strength of feature extraction and its generality.

The system can make microbial diagnostics available, low-cost, and accessible at the point of care by converting a smartphone into a miniaturized microscopy platform. This system offers a portable solution for rapid bacterial detection, making it highly suitable for real-world applications such as public health surveillance, environmental monitoring, and resource-limited settings. In future the system can be extended to multi-class classification for additional bacterial species, for rapid and on-site detection, developing a real-time smartphone-based application and incorporating more diverse and larger data sets to improve model generalization and robustness. To evaluate model robustness across different imaging conditions and smartphone microscopy systems cross-device validation will be performed. Although smartphone images and deep learning can be scaled to enable fast pathogen identification, they have a broad range of uses in environmental monitoring, water-quality sensing, food safety, and remote healthcare.

References

- [1] S. Visitsattapong, M. Bunkum, C. Pintavirooj, and M. P. Paing, "A deep learning model for bacterial classification using big transfer (BiT)," *IEEE Access*, vol. 12, pp. 15609–15621, 2024.
- [2] Z. Wang *et al.*, "Deep neural networks offer morphologic classification and diagnosis of bacterial vaginosis," *Journal of Clinical Microbiology*, vol. 59, no. 2, pp. 10–1128, 2021.
- [3] S. Patel, "Bacterial colony classification using atrous convolution with transfer learning," *Annals of the Romanian Society for Cell Biology*, vol. 25, no. 4, pp. 1428–1441, 2021.
- [4] P. Yin *et al.*, "A smartphone-based fluorescent sensor for rapid detection of multiple pathogenic bacteria," *Biosensors and Bioelectronics*, vol. 242, p. 115744, 2023.
- [5] P. Paquin *et al.*, "Spatio-temporal based deep learning for rapid detection and identification of bacterial colonies through lens-free microscopy time-lapses," *PLOS Digital Health*, vol. 1, no. 10, p. e0000122, 2022.
- [6] S. Y. Chin *et al.*, "Bacterial image analysis using multi-task deep learning approaches for clinical microscopy," *PeerJ Computer Science*, vol. 10, p. e2180, 2024.

- [7] H. Ramesh *et al.*, “Biointel: Real-time bacteria identification using microscopy imaging,” in *Proc. IEEE Int. Symp. Biomed. Imaging (ISBI)*, 2024, pp. 1–4.
- [8] T. Treebupachatsakul and S. Poomrittigul, “Bacteria classification using image processing and deep learning,” in *Proc. 34th Int. Tech. Conf. Circuits/Systems, Computers and Communications (ITC-CSCC)*, JeJu, Korea (South), 2019, pp. 1–3.
- [9] S. M. Abd-Alhalem *et al.*, “Bacterial classification with a convolutional neural network based on different data,” 2019.
- [10] R. Kang, B. Park, M. Eady, Q. Ouyang, and K. Chen, “Single cell classification of foodborne pathogens using hyperspectral microscope imaging coupled with a deep learning framework,” *Sensors and Actuators B: Chemical*, 2020.
- [11] T. S. R. Mhathesh, J. Andrew, K. Martin Sagayam, and L. Henesey, “A 3D convolutional neural network for bacterial image classification,” in *Intelligence in Big Data Technologies—Beyond the Hype*, New York, NY, USA: Springer, 2021, pp. 419–431.
- [12] F. M. Khan, R. Gupta, and S. Sekhri, “A convolutional neural network approach for detection of E. coli bacteria in water,” *Environmental Science and Pollution Research*, vol. 28, no. 43, pp. 60778–60786, 2021.
- [13] T. Ahmed, M. F. Wahid, and M. J. Hasan, “Combining deep convolutional neural network with support vector machine to classify microscopic bacteria images,” in *Proc. Int. Conf. Electrical, Computer and Communication Engineering (ECCE)*, Feb. 2019, pp. 1–5.
- [14] A. Narin, C. Kaya, and Z. Pamuk, “Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks,” *Pattern Analysis and Applications*, vol. 24, no. 3, pp. 1207–1220, 2021.
- [15] A. S. Arnob *et al.*, “Comparative result analysis of cauliflower disease classification based on deep learning approach VGG16, InceptionV3, ResNet, and a custom CNN model,” *Hybrid Advances*, vol. 10, p. 100440, 2025.
- [16] P. ForouzeshFar *et al.*, “Dental caries diagnosis from bitewing images using convolutional neural networks,” *BMC Oral Health*, vol. 24, no. 1, p. 211, 2024.
- [17] S. Gupta *et al.*, “Cotton disease prediction using deep transfer learning: Comparative analysis of ResNet50, VGG16, and InceptionV3 models,” *International Journal of Engineering Science and Technology (IJESTY)*, vol. 6, no. 5, pp. 78–88, 2023.
- [18] D. Nie, E. A. Shank, and V. Jovic, “A deep learning framework for bacterial image segmentation and classification,” in *Proc. 6th ACM Conf. Bioinformatics, Computational Biology and Health Informatics*, 2015, pp. 306–314.
- [19] Y. P. Lopez *et al.*, “Automatic classification of light field smear microscopy patches using convolutional neural networks for identifying Mycobacterium tuberculosis,” in *Proc. CHIL-EAN Conf. Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON)*, Pucon, 2017, pp. 1–52.
- [20] G. Turra, S. Arrigoni, and A. Signoroni, “CNN-based identification of hyperspectral bacterial signatures for digital microscopy,” in *Lecture Notes in Image Analysis and Processing (ICAP)*, Springer, vol. 10485, 2017.
- [21] B. Zielinski *et al.*, “Deep learning approach to bacterial colony classification,” *PLoS ONE*, 2017.
- [22] R. O. Panicker, K. S. Kalmady, J. Rajan, and M. K. Sabu, “Automatic detection of tuberculosis bacilli from microscopic sputum smear images using deep learning methods,” *Biocybernetics and Biomedical Engineering*, 2018.
- [23] B. B. Traore, B. Kamsu-Foguem, and F. Tangara, “Deep convolution neural network for image recognition,” *Ecological Informatics*, vol. 48, pp. 257–268, 2018.
- [24] E. A. Hay and R. Parthasarathy, “Performance of convolutional neural network for identification of bacteria in 3D microscopy datasets,” *PLoS Computational Biology*, vol. 14, no. 12, p. e1006628, 2018.
- [25] K. S. Mithra and W. R. S. Emmanuel, “Automated identification of Mycobacterium bacilli from sputum images for tuberculosis diagnosis,” *Signal, Image and Video Processing*, vol. 13, pp. 1585–1592, 2019.
- [26] M. F. Amria *et al.*, “Bacterial classification using deep structured convolutional neural network for low resource data,” *Journal Elektronika dan Telekomunikasi (JET)*, vol. 23, no. 1, pp. 47–54, 2023.