African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 120-152
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 15, 2025
Peer reviewed: December 20, 2025
Revised: February 26, 2026
Published: April 2026

DETECTION OF PNEUMONIA FROM CHEST X-RAY IMAGES USING ENSEMBLE DEEP LEARNING WITH A VOTING MECHANISM

Bager Dahoos, Z.[1], Elbai, S.[2], Alghrawy, M. A.[3], Haitham Abdul Amir, A.[4], Hussien, G. H.[5], Elayati, K.[6], and Alaidany, A. A.[7]
[1,5] Department of Computer Science, College of Computer Science and Information Technology, University of Basra, Basra, Iraq.
[2] Department of Genetics Engineering, Biotechnology Research Centre (BTRC), Libya.
[3] Department of Electrical Engineering, College of Engineering, University of Misan, Al-Amarah, Misan, Iraq.
[4,7] Department of Computer Science, Shatt Al-Arab University College, Basra, Iraq.
[6] Department of Microbiology, Biotechnology Research Centre, Libya.
[1]zainab.dahoos@uobasrah.edu.iq

ABSTRACT

Purpose: The study aims to perfect the accuracy and reliability of pneumonia detection in chest X-ray pictures, as pneumonia remains a major cause of respiratory-related deaths, particularly among children and the elderly, while manual diagnosis is often time-consuming and prone to variability and errors.
Design/Methodology/Approach: The study proposes an integrated hybrid detection framework that combines unsupervised clustering techniques such as K-means with traditional deep learning models, where deep neural networks extract high-level features, and clustering identifies hidden data patterns, and the outputs are integrated using soft voting and majority voting techniques, with hyperparameter tuning applied to reduce overfitting and enhance performance.
Findings: The proposed FBMV model achieved superior results, reaching 100% training accuracy, 99.51% validation accuracy, and 99.38% precision with the lowest loss values, while MobileNet demonstrated the fastest execution time suitable for real-time applications, and VGG16 showed the lowest memory usage, making it suitable for well-constrained environments.
Research Limitation: The study is limited by its reliance on a specific high-quality chest X-ray dataset, which may not fully represent real-world variability, and further validation is required across diverse clinical settings and populations.
Practical Implication: The proposed model can support healthcare professionals in achieving faster, more accurate pneumonia diagnoses, reducing errors and improving clinical decision-making, especially in limited-resource environments.
Social Implication: Enhancing pneumonia detection contributes to early treatment, reduced mortality rates, and improved healthcare outcomes, particularly for vulnerable groups such as children and the elderly.

ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 120-152
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 15, 2025
Peer reviewed: December 20, 2025
Revised: February 26, 2026
Published: April 2026

Originality/Value: The study presents a novel hybrid range that integrates unsupervised clustering with deep learning and ensemble voting techniques (FBMV), providing a more stable and accurate diagnostic approach and adding value through the efficiency of MobileNet and the low-resource capabilities of VGG16.

Keywords: *Chest x-ray. classification. deep learning. healthcare. neural network.*

INTRODUCTION

Pneumonia is a severe infectious disease that poses a significant risk to life and is among the foremost causes of serious illness and mortality globally, especially affecting children, the elderly, immunocompromised individuals, and those with pre-existing chronic conditions(Çınar et al., 2021). Pneumonia is an infection of the lungs caused by bacteria, viruses, or fungi. Pneumonia reduces oxygen levels in the body, hindering an individual's ability to breathe. Patients with severe cases are typically hospitalised, and in instances of extreme severity, they may necessitate a ventilator for respiratory assistance(Chakraborty et al., 2022). Chest X-rays are the predominant modality for diagnosing and monitoring this condition, attributed to their affordability and accessibility. Nonetheless, manual interpretation from these images is labour-intensive and exhibits diversity in diagnostic outcomes among clinicians, potentially resulting in misdiagnoses (Baik et al., 2024).

In recent years, the field of medical image analysis has witnessed significant development thanks to deep learning techniques, particularly convolutional neural networks, which have proven highly efficient for classifying, detecting, and segmenting chest diseases, including pneumonia. Computer-aided diagnosis systems based on deep learning provide significant support to physicians by reducing the likelihood of error and improving the accuracy of clinical decisions (Rana & Gautam, 2023).

Numerous studies have examined the use of multiple architectures, such as VGG16, ResNet152V2, DenseNet121, MobileNetV2, and InceptionV3, achieving classification accuracies exceeding 90% in most cases. The integration of transfer and ensemble learning techniques has also significantly improved model performance. Despite these achievements, challenges remain with these models, including high computational complexity, poor clinical interpretability, and limited generalizability across different clinical settings. Therefore, recent research focuses on designing lightweight, reliable, and practical real-time models (Byeon, 2024).

The model proposed in this research aims to develop a deep learning model that accurately detects pneumonia in chest X-rays. This study compares and contrasts the performance of VGG16 and DenseNet-121, two successful CNN-based architectures from previous medical

ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 120-152
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 15, 2025
Peer reviewed: December 20, 2025
Revised: February 26, 2026
Published: April 2026

imaging research, with that of other successful models, such as DenseNet-169 and MobileNet. The main goal of the study is to develop an intelligent detection system that performs well in healthcare settings with limited resources.

propose a model that integrates these models with sophisticated data augmentation techniques will enhance detection accuracy and robustness beyond previously documented outcomes. The strategy will involve architectural alterations, such as reducing specific layers to mitigate overfitting, incorporating dense layers to improve feature representation, and prolonging training epochs to enhance model convergence. Ensemble detection techniques utilising hard voting and soft voting will also be implemented. Model evaluation will utilise comprehensive metrics, including detection accuracy, precision, recall, and F1-score. A meticulously curated, standardised, and high-quality collection of chest X-ray images from pneumonia patients will be used to ensure database consistency and repeatability.

LITERATURE REVIEW

This part discusses a previous study on the classification and early detection of pneumonia from chest x-rays. This paragraph aims to briefly summarise several surveys on this topic, placing the current research within the context of the available literature.

A recent study employed deep learning to analyse X-ray pictures and develop a system capable of predicting pneumonia (Jakhar & Hooda, 2018). The study developed a model classification utilising Deep Convolutional Neural Network (DCNN) techniques, achieving of 100% accuracy in distinguishing between pneumonia patients and healthy individuals. The proposed model was compared to standard machine learning algorithms like Naive Bayes, Decision Tree, Logistic Regression, Support Vector, AdaBoost, and Random Forest. There were 5,863 X-ray pictures in the dataset.

A 10-fold K-Fold cross-validation technique was used to ensure the model was stable after cleaning and prepping the data (scaling, eliminating bad photographs). Recall, accuracy, mistake rate, F1-score, and AUC were among the measures utilised to evaluate the outcomes. With an F1-score of 77%, TP rate of 92%, and FP rate of 11%, the DCNN model outperformed the other models. Its accuracy was 84%. Conversely, more conventional models underperformed. To illustrate the point, Random Forest achieved 82% accuracy and Neural Network 81%.

ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 120-152
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 15, 2025
Peer reviewed: December 20, 2025
Revised: February 26, 2026
Published: April 2026

A number of strengths of the work include its use of deep learning techniques in a huge data processing setting, its extensive use of comparison algorithms, and K-Fold cross-validation. A number of issues plagued the study's use of a deep model. These included a dataset that was unequally distributed between the normal and pneumonia groups, and a lack of clarity on the relationship between memory utilisation and execution time(Jakhar & Hooda, 2018).

Sheu et al. (2022) proposed a system called MDA-PSP that aims to predict the recovery status of pneumonia patients during the first 3 days of hospitalisation and determine whether they will be discharged within 7 days or require a longer stay. This was achieved by employing deep and machine learning algorithms to fuse and analyse vital signs data and chest X-ray (CXR) images. The deep learning model was compared with machine learning algorithms: Support Vector Machine, Random Forest, XGBoost, and Decision Tree.

The number of patients from whom data was collected was 3,972 (mean age 71 ± 17 years), and the types of data included CXR images, vital signs (pulse, respiration, blood pressure, temperature, etc.), and laboratory data (WBC, CRP, glucose, etc.). The patients were divided into two groups: the first group included those discharged within 7 days, and the second group included those discharged beyond 7 days. The results showed that the Dense-BN model had the highest prediction accuracy, achieving Accuracy = 0.77, F1-score = 0.76, Precision = 0.79, and Recall = 0.57. The advantages of the study include the efficient integration of clinical data and medical images into an intelligent predictive system. The disadvantages of the study include the lack of K-fold cross-validation during training and validation. Details of the model's memory consumption and computational time are not provided(Byeon, 2024).

Manan Pruthi et al. (2023) conducted a comparative study to evaluate the effectiveness of machine and deep learning algorithms for classifying chest X-ray images of patients with pneumonia. The Random Forest was used as a representative of traditional machine learning, while the VGG-16 and Inception V3 convolutional neural network (CNN) models were used. The database was split into 80% for training, 10% for validation, and 10% for testing, with data augmentation used to reduce imbalance between the two classes (pneumonia/normal). The Random Forest achieved a relatively low accuracy (75.00%), while the VGG-16 model achieved an accuracy of 91.99%, demonstrating the superiority of deep learning models in this field.

The advantages of the study included comparing CNNs and ML algorithms and using overtraining mitigation strategies such as Early Stopping and Dropout. Disadvantages of the study included the lack of cross-validation techniques, such as K-Fold cross-validation, to

ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 120-152
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 15, 2025
Peer reviewed: December 20, 2025
Revised: February 26, 2026
Published: April 2026

improve generalisation. The comparison between machine learning and deep learning was limited to accuracy only, without mentioning other metrics such as F1-score, precision, and recall, and the study did not compare computational resource consumption or the possibility of deploying the model into real-world environments(Baik et al., 2024).

Vetrithangam et al. (2023) presented a study aimed at early and accurate detection of pneumonia using chest X-ray images by designing a deep learning model based on the ResNet152V2 architecture. The model architecture was optimised using weight adjustment and data pre-normalisation techniques to reduce computational time while maintaining high classification accuracy. This study used the RSNA Pneumonia Detection Challenge database available on Kaggle, which contains 1485 X-ray images. The images were resized to 224 × 224 pixels, and data augmentation techniques (such as rotation, cropping, and reflection) were applied, along with data normalisation.

The model results were evaluated using a confusion matrix, classification accuracy, recall, precision, and F1 coefficient. The model achieved accuracy = 99.77%, recall = 99.86%, specificity = 95.4%, precision = 99.86%, and F1-score = 99.86%, and achieved a shorter training time. The advantages of the study include improving the ResNet152V2 architecture without the need to train the model from scratch. As for the disadvantages of the study, K-Fold Cross Validation was not mentioned in the evaluation (Rana & Gautam, 2023).

Rani Puspita et al. (2024) published a study aimed at predicting pneumonia in chest x-ray images using deep learning techniques to automatically analyse chest x-ray pictures and determine whether they are infected with pneumonia. The research used deep learning techniques using the VGG16 and DenseNet121 neural network models to classify chest x-ray pictures. The image data was collected from Kaggle.com and divided into three sets: training data (5,216 images), testing data (624 images; 234 normal and 390 infected), and validation data (16 images; 8 normal and 8 infected).

After running experiments, the VGG16 model achieved a test accuracy of 90%, and the DenseNet121 model achieved a test accuracy of 88%, thus outperforming the DenseNet121 model. The use of real data from the trusted Kaggle medical website is one of the advantages of this research, along with the use of two neural network models, which are among the most powerful classification models for medical images. The validation database is small, which is a disadvantage in this research and may affect the evaluation's accuracy. Furthermore, the lack of techniques such as K-Fold, which enhance the reliability of the evaluation, and of

ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 120-152
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 15, 2025
Peer reviewed: December 20, 2025
Revised: February 26, 2026
Published: April 2026

comparisons of computational resource consumption during model execution, is important information for determining the practical applicability of the model (Vetrithangam et al., 2023).

Seung Min Baik et al. (2024) presented a study aimed at developing an early mortality prediction model for pneumonia patients using laboratory test results. The study used clinical and laboratory data from 1,065 patients, comprising 80,940 medical records. The data were processed using cleaning and standardisation techniques, with the data split into of 80% train and 20% of test. Machine learning was used: XGBoost, CatBoost, LightGBM (LGBM), Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbour (KNN), and Deep Learning Multilayer Perceptron (MLP). The Ensemble model was built using two methods: soft voting using XGBoost, CatBoost, and LGP, and hard voting using XGBoost, CatBoost, and RF. Model performance was improved by modifying the cut-off to enhance the F1-score. The study used the SHAP and Feature Importance interpretability tools to identify factors influencing mortality. The results showed that the best model was the Ensemble model, which achieved AUROC = 0.9006, Accuracy = 90%, and F1-score = 81%. In contrast, the XGBoost model was the best individual model, achieving AUROC = 0.8989 and F1-score = 80% after optimisation. The use of the Ensemble method to combine powerful algorithms to improve accuracy is one of the advantages of the study, in addition to the use of advanced interpretation tools such as SHAP. The disadvantages of the study include the lack of benefits from integrating clinical data with imaging data (Sheu et al., 2022).

Haewon Byeon (2024) proposed a hybrid model that combines the segmentation accuracy of U-Net and the classification power of DenseNet to process medical image features more efficiently and improve the detection accuracy of pneumonia in chest x-ray images. The model used 256×256-pixel normalised chest x-ray images and applied data augmentation techniques (such as rotation, horizontal flipping, and zooming) to increase data diversity and improve generalisation. The proposed model (U-Net + DenseNet) outperformed conventional models, achieving a classification accuracy of 96.4%, F1-score of 94.8%, Recall of 95.2%, and precision of 94.3%.

The model also demonstrated its ability to identify infected areas with high visual accuracy from segmentation maps. The research aims to enhance the accuracy of the results by combining a segmentation model with a classification model. Disadvantages of the study include not specifying the database type or size in detail, not using cross-validation techniques such as K-Fold Cross Validation, and not comparing training time or resource consumption (Agard et al., 2025).

ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 120-152
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 15, 2025
Peer reviewed: December 20, 2025
Revised: February 26, 2026
Published: April 2026

Rajawat et al. (2025) presented a model for early pneumonia diagnosis using deep and machine learning algorithms. The study uses deep imaging and deep learning techniques to analyse symptoms (e.g., cough, chills, changes in respiratory rate, and fever). A hybrid model combining convolutional neural networks (CNNs) and long short-term memory (LSTM) networks was proposed to classify video images or X-rays and accurately identify symptoms of pneumonia.

The deep learning algorithms used were Long Short-Term Memory Networks (LSTMs) and an Attention Mechanism. The dataset used was 224×224 grayscale X-ray images with data augmentation, FLIR thermal imaging, and audio sources. The proposed CNN+LSTM model achieved Accuracy = 98.02%, F1-score = 94.9%, Precision = 94.0%, Recall = 95.8%, and AUC-ROC = 96.1. In contrast, the Random Forest algorithm achieved an accuracy (65.8%), the KNN algorithm (84.91%), the SVM algorithm (84.71%), and the Deep Belief Networks (85.71%). One of the advantages of the study is the combination of CNNs and LSTMs with an attention layer, which enhances the ability to detect subtle patterns, in addition to the use of multimodal data (images, temperature, audio) (Gabhale et al., 2017).

Agard et al. (2025) proposed a model called PREDICT, which aims to predict early ventilator-associated pneumonia (VAP) in intensive care units (ICUs) by analysing only vital signs, without the need for images or complex laboratory tests. The model relies on deep learning techniques. A Long Short-Term Memory (LSTM) model was used. The MIMIC-IV dataset was used to train the model, which included real-world ICU records from 2008 to 2019. There were 904 patients, and 452 were confirmed VAP cases. The model was compared with machine learning models, including Random Forest, XGBoost, and Logistic Regression. The PREDICT model achieved a specificity of 99.5%, outperforming all conventional models. Among the advantages of this study is that this model is the first deep model dedicated to the early prediction of VAP using only vital signs. Among the disadvantages of this study is that the model has not been tested in real-time clinical trials(Li et al., 2022).

Table 1 illustrates previous research on pneumonia detection. The comparative study identifies a research gap in real-time deployment capabilities, which our suggested methodology aims to remedy. Related research shows some limitations and challenges, most notably limited comparisons across methods, as many researchers focus on a few algorithms or architectures. Interpretability of results is a challenge, especially with deep learning techniques such as CNNs, despite their ability to automatically extract features. The types of data used also vary across studies, making it difficult to understand the results of unsupervised learning. Other challenges include limited data volume, the complexity of combining multiple techniques, and the use of pneumonia from chest x-ray images, which may not always be available.

ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 120-152
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 15, 2025
Peer reviewed: December 20, 2025
Revised: February 26, 2026
Published: April 2026

Table 1: Related works comparison

Authors/year	Objective	Methods/ Algorithms	Datasets	Results	Strengths	Limitations
Jakhar & Hooda (2018)	Classify pneumonia vs. normal X-rays using deep	DCNN, RF, SVM, AdaBoost, Logistic Regression, Decision Tree, Naive Bayes .	5,863X-ray images	DCNN: Accuracy=84%, F1=77%, TP=92%, FP=11%. RF: 82% accuracy	Large dataset, K-Fold CV, multiple metrics	Imbalanced data, no computational time/memory details
Sheu et al. (2023)	Predict pneumonia recovery (discharge ≤7 days) using multimodal	Dense-BN, SVM, RF, XGBoost, Decision Tree	3,972 patients (CXR + vital signs + lab data) .	Dense-BN: Accuracy=0.77, F1=0.76, Precision=0.79, Recall=0.57 .	Integrated clinical + image data	No K-Fold CV, no computational details
Pruthi et al. (2023)	Compare ML/DL for pneumonia classification	RF, VGG-16, Inception V3	Not specified (80-10-10 split)	VGG-16: Accuracy=91.99%; RF: 75%.	Data augmentation, Early Stopping, Dropout .	No K-Fold, limited metrics (only accuracy), no resource analysis
Vetrithangam et al . (2023)	Early pneumonia detection using optimized ResNet	ResNet152V2 (optimized)	1,485X-rays (RSNA Kaggle) .	Accuracy=99.77%, Recall=99.86%, F1=99.86%. Training time reduced.	Pretrained model + normalization.	No K-Fold CV
Puspita et al. (2024)	Pneumonia prediction from X-rays using VGG16/DenseNet	VGG16, DenseNet121	5,216 train, 624 test, 16 validation images (Kaggle)	VGG16: Accuracy=90%; DenseNet121: 88%	Real-world Kaggle data, powerful models	Tiny validation set, no K-Fold or computational analysis

ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 120-152
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 15, 2025
Peer reviewed: December 20, 2025
Revised: February 26, 2026
Published: April 2026

Baik et al. (2024)	Predict pneumonia mortality using lab data	Ensemble (XGBoost, CatBoost, RF), SVM, KNN, MLP	1,065 patients (80,940 lab records)	Ensemble: AUROC=0.9006, Accuracy=90%, F1=81%. XGBoost: AUROC=0.8989.	SHAP interpretability, Ensemble boost	No imaging data integration
Byeon (2024)	Hybrid model for pneumonia detection (segmentation + classification	U-Net + DenseNet	Not specified (256×256 X-rays) .	Accuracy=96.4%, F1=94.8%, Recall=95.2%, Precision=94.3% .	Combines segmentation + classification .	No dataset/K-Fold details, no resource comparison
Rajawat et al (2025)	Multimodal pneumonia diagnosis (X-rays + symptoms) .	CNN-LSTM + Attention, RF, KNN, SVM, Deep Belief Networks .	224×224 X-rays + thermal/audio data.	CNN-LSTM: Accuracy=98.02%, F1=94.9%, AUC=96.1%. RF: 65.8%.	Multimodal data, attention mechanism.	No K-Fold or computational analysis

METHODOLOGY

In this research, several convolutional neural network models were used in addition to ensemble learning techniques, and the resulting models were applied to the dataset after it was prepared.

Dataset

This study utilised a dataset of chest X-ray images (pneumonia vs normal) sourced from Kaggle.com. Earlier research utilised comparable datasets that vary in various dimensions, including image size, clarity, and resolution. The dataset we got has 5,878 chest x-ray images, split into two groups: 4,284 of them are from pneumonia patients and 1,594 are from healthy patients. We made many changes to the images in this dataset, including ensuring they were all 224 × 224 × 3 pixels. This made it easier and more accurate for the models to tell which patients had pneumonia. Figure 1 shows some examples from the dataset.

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 120-152
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 15, 2025
Peer reviewed: December 20, 2025
Revised: February 26, 2026
Published: April 2026

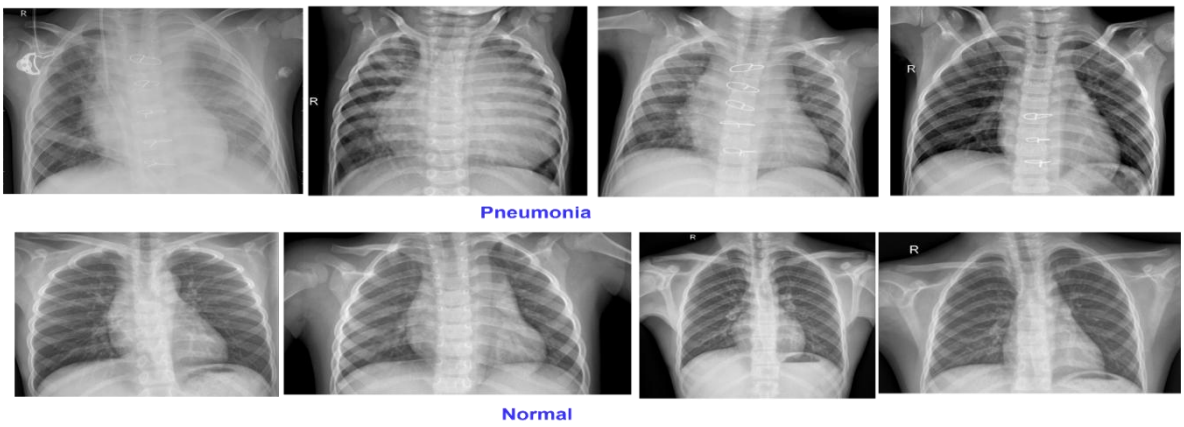


Figure 1: Sample chest x-ray picture for the dataset.

Data Preparation

Data preparation is a crucial element in deep learning projects, as it involves organising the dataset for use by the model. The primary objective of data preparation is to enhance data quality and minimise potential errors during training. In Figure (2), the data preparation operation typically encompasses the following phase:

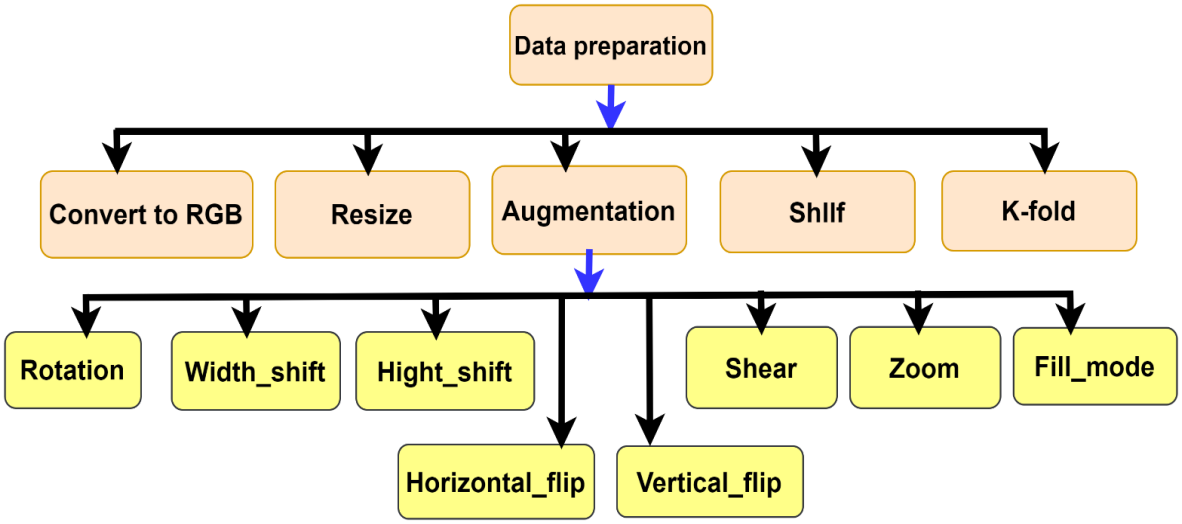


Figure 2: Data preparation phase

Figure 2 illustrates that the data preparation phase generally encompasses multiple steps, including conversion to RGB, resizing, applying data augmentation techniques (such as

ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 120-152
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 15, 2025
Peer reviewed: December 20, 2025
Revised: February 26, 2026
Published: April 2026

rotation, width adjustment, height adjustment, horizontal and vertical reflections, shear transformation, zoom, and fill mode), shuffling, and performing K-fold cross-validation.

Method Of Chest X-Ray Image Classification

Early detection of pneumonia is of utmost importance; however, relying solely on traditional procedures that rely on the physician's expertise in examining chest x-ray images is insufficient and requires the adoption of contemporary alternatives that utilize artificial intelligence techniques. This section describes the proposed workflow for a deep learning approach, DeepCOVNet, which includes data pre-processing, followed by a classification model derived from a CNN model(Thirunavukkarasu et al., 2024) .

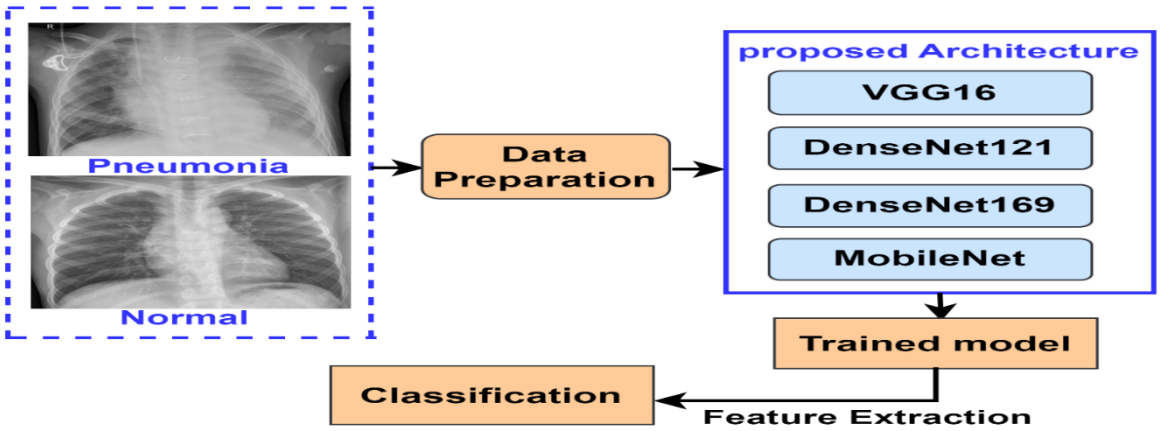


Figure 3: Proposed approach for chest X-ray picture classification.

This project involves training, evaluating, and testing four pre-trained deep learning frameworks for pneumonia detection on chest X-ray images. The VGG16, DenseNet121, DenseNet169, and MobileNet models serve as the principal detection models, augmented by ensemble learning methodologies. Figure 3 illustrates the comprehensive methodology of this study, outlining the suggested deep learning approach based on an efficient pipeline: data augmentation followed by the pre-processing of chest X-ray images. The detection model, trained via transfer learning, is then trained to extract relevant information, enabling the system to independently differentiate pneumonia cases from normal cases.

ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 120-152
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 15, 2025
Peer reviewed: December 20, 2025
Revised: February 26, 2026
Published: April 2026

Pre-Processing Data

Algorithm 1: Pre-processing of Chest X-ray picture for Pneumonia Detection

- **Input :** Unprocessed chest X-ray pictures dataset
- **Output:** Processed chest X-ray picture

1. Obtain a varied and standardized dataset from chest X-ray picture
2. Implement noise reduction through digital filtering methodologies.
3. Modify brightness and contrast to improve image clarity.
4. Improve fine details by sharpening the image.
5. Execute picture segmentation to separate the chest region from the background.
6. Standardize pixel intensity values to a uniform scale.
7. Provide the preprocessed images.

As shown in Algorithm 1, the preprocessing step includes several steps intended to improve chest X-ray images and prepare them for the detection task. The steps include removing noise, adjusting brightness and contrast, sharpening the image, breaking it into parts, and normalising it.

Training Models

The suggested model adheres to a standardised deep learning pipeline, as seen in the

ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 120-152
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 15, 2025
Peer reviewed: December 20, 2025
Revised: February 26, 2026
Published: April 2026

Algorithm 2.Models training

- **Input:** Processed chest X-ray images.
- **Output:** trained detection models

1-Obtain an RGB picture or resize it to 224×224 pixels.
Implement pretreatment transformations, including normalization and optional data augmentation.
2 -Convolutional Feature Extraction- :
The fundamental structure comprises many convolutional layers from feature extraction.
Employ activation functions (ReLU or ReLU6) subsequent into each convolution.
Internal MaxPooling layers to systematically diminish spatial dimensions.
3 -Flattening: Transform the final feature mappings into a one-dimensional vector.
4-Fully Connected Layer
Implement a dense layer for 512 neurons.
Implement ReLU or ReLU6 activation functions.
Implement a Dropout layer to mitigate overfitting.
5 -Fully Connected Layer 2:
Implement a dense layer of 256 neurons.
Implement ReLU or ReLU6 activation functions.
Implement a Dropout layer to improve generalization.
6-Output Layer:
Implement a dense layer together the number of units corresponding into the number of classes.
Employ Softmax activation during inference to generate class probabilities.
7-Training Configuration:
Utilize a learning rate of 0.00001 for training.
Employ an optimizer like Adam to adjust weights.
Iterate the procedure for a predetermined number of epochs until convergence is achieved.

Convolutional Neural Networks

Convolutional Neural Networks have been shown to be very good at classifying, recognising, and detecting objects in images, making them a good choice for diagnosing pneumonia using chest x-ray images. A standard CNN building with several layers of input, convolution, pooling, a completely connected layer, and output information can easily deal with shifts and scales. Common architectures, such as VGG16, DenseNet121, DenseNet169, and MobileNet, have made training faster and less error-prone by allowing parameter sharing and modular designs, even when trained on a small dataset. But their overall performance depends a lot on the quality of the dataset, and high-resolution images are necessary for correct classification (Alaidany, 2024). Also, it is still hard to make sure that models can be used on different datasets and make them easier to understand. Future research should focus on enhancing data quality and creating explainable AI methodologies to increase the reliability of these models in medical diagnostics(Sheu et al., 2022; Çınar et al., 2021).

ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 120-152
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 15, 2025
Peer reviewed: December 20, 2025
Revised: February 26, 2026
Published: April 2026

Figure 4 displays a conceptual CNN-based system for sorting chest X-ray pictures. In this system, the chest x-ray images pass through several convolutional layers with ReLU activation to extract features ranging from simple to complex. Then a pooling layer reduces the size of the data and stabilises it. Then, the topographical maps are reduced to a two-dimensional form and sent as vectors through fully connected layers to learn how features relate to one another. Finally, a softmax function outputs a class probability to choose the most likely class (Guefrechi et al., 2021). The system uses a CNN trained on the ImageNet dataset to improve classification accuracy and better handle previously unseen data. By using transfer learning and freezing the pre-trained layers, the training process is sped up, and overfitting is less likely to happen (Alaidany, 2025). Keeping an eye on the difference between training and validation accuracies is still very important to make sure that pneumonia classification using chest x-ray images works well (Bhattacharjee et al., 2023).

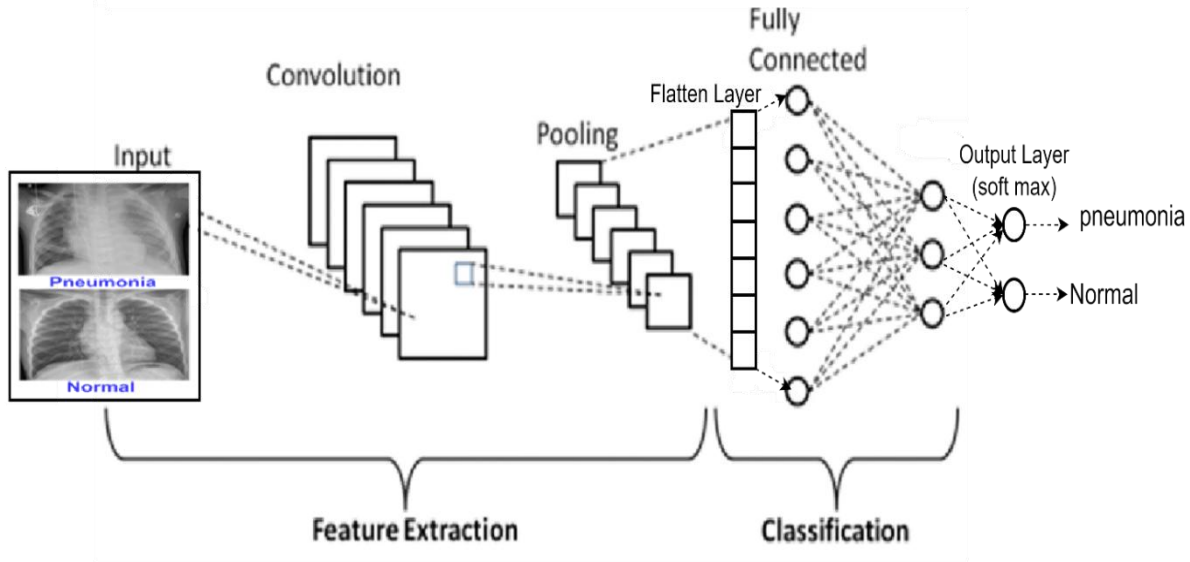


Figure 4: CNN Architecture for chest X-ray pictures classification

Numerous enhancements were implemented to improve the efficiency and effectiveness of all models. Layer normalisation was implemented following extensive experimentation with several layers denoting certain classes. The outputs were thereafter transmitted to a fully connected layer that decreased the dimensionality to 512 or 256 units. Each model possesses distinct value, enhancing stability throughout the training process. Employed K-fold cross-validation with K = 5 across more than 10 or 50 epochs, depending on the model. The model analyses images measuring 224 × 224 pixels, employs the Nadam optimiser with a learning rate of 0.0001 or 0.00001, utilises a ReLU or ReLU6 activation function for the hidden layers,

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 120-152
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 15, 2025
Peer reviewed: December 20, 2025
Revised: February 26, 2026
Published: April 2026

and operates with a batch size of 32 or 64. The dropout rates were varied across architectures to 0.4, 0.5, 0.6, 0.7, and 0.8. The adjustments produced a more resilient and efficient model.

The network design was altered by adjusting the number of dense layers and constructing a sequence of dense layers with regularised dropout, followed by a convolutional process. This pattern is frequently employed in neural network architectures to facilitate dense layers in acquiring progressively intricate features, mitigate overfitting, and amalgamate features via convolution into the model by utilising information from the outputs of the initial and secondary dense layers.

Ensemble learning

The combination of recall and accuracy is an important way to measure performance. This metric evaluates the chest x-ray picture classification model's performance across all areas, including false positives and false negatives. The F1-Score is particularly helpful for unbalanced datasets because it provides a clear picture of how well a model performs. Accuracy might not count certain kinds of mistakes, so the F1-Score is a better way to assess how well a model is performing. This is because it looks at both kinds of mistakes. The integration of many classifiers or base learners significantly improves accuracy relative to a single classifier or base learner (Mathew & Sathyalakshmi, 2024).

Voting classification is one of the most powerful ensemble techniques. It combines the predictions from doubled underlying models to reach a final decision on a classification problem. It is a common technique until deep learning, where the results of several individual classification models are combined to arrive at a more accurate final decision. There are two types of voting classifiers: hard voting and soft voting.

Majority Voting

This method works by summing the "votes" of each classifier for a given category, then selecting the category with the highest vote count as the final result. We can conclude that Fusion-Based Voting (FBMV) is a classification model because it assigns data to categories, taking input data (the results of individual classifiers) and outputting a classification or category for that data, and because it uses a set of quantifiable characteristics (the number of votes) received by each individual classifier. Figure 5 shows the overall voting (Taye, 2023).

ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 120-152
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 15, 2025
Peer reviewed: December 20, 2025
Revised: February 26, 2026
Published: April 2026

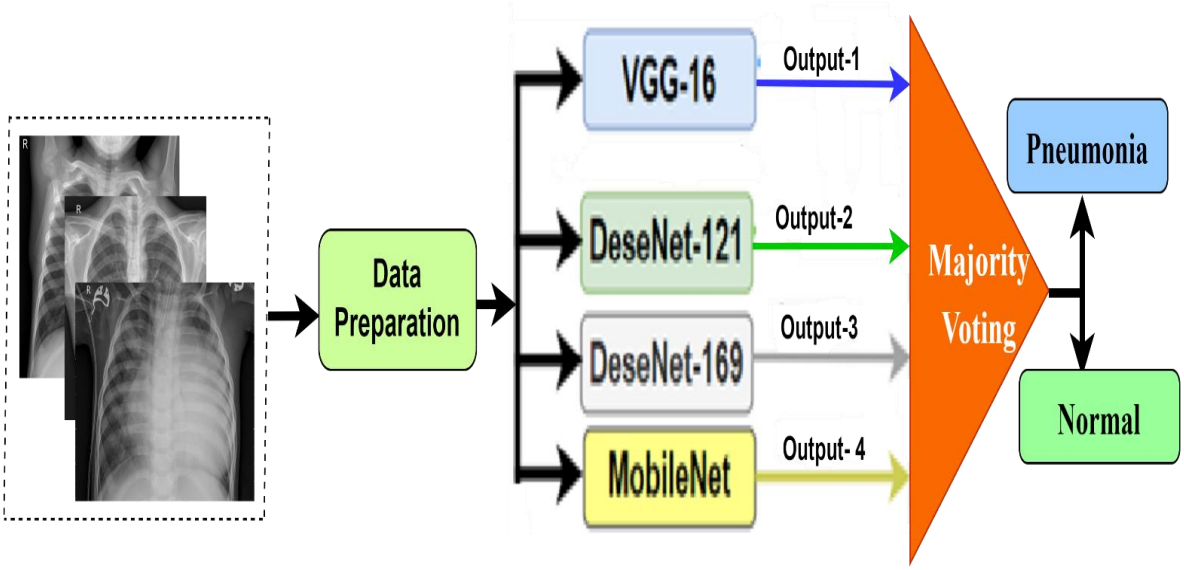


Figure 5: majority voting for classification

AVERAGING (SOFT VOTING)

Soft voting uses the model predictions for each data sample. The arithmetic mean of the model's predictions is found for each sample. Then, the class with the highest cumulative chance is picked. One big problem is that it could lead to overfitting. If the models used for soft voting are too effective, the results may not be perfect. Soft voting is demonstrated in Figure 6 (Jiang & Li, 2023).

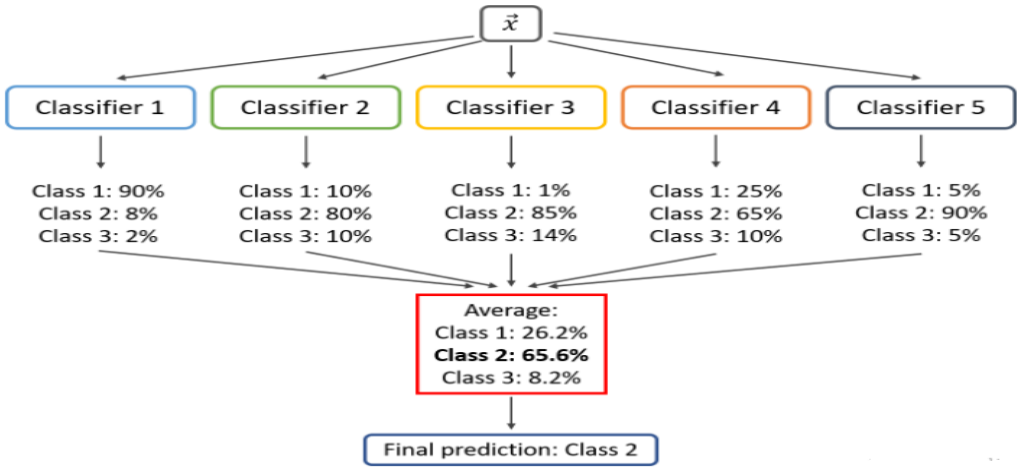


Figure 6: Soft Voting for detection

ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 120-152
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 15, 2025
Peer reviewed: December 20, 2025
Revised: February 26, 2026
Published: April 2026

Evaluation Metrics

In chest x-ray image classification, conventional evaluation tools and metrics define the performance assessment of a practical application model. Using several evaluation criteria and calibration approaches, one can assess the correctness, resilience, and efficiency of these models. The key metrics and techniques applied in the evaluation of chest x-ray image classification algorithms are discussed in this part (Li et al., 2021).

1. Accuracy: is a fundamental evaluative metric used in several fields, including chest x-ray images classification applications. It signifies the proportion of correct predictions to the total number of samples, formally defined as shown in Equation (1) (Haitham et al., 2025a):

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \qquad (1)$$

2. Precision: is quantifying the ratio of correct positive identifications to the total positive matches found, formally represented as shown in Equation (2) (Haitham et al., 2025b).

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

3. Recall: include the measurement of the ratio of correctly identified positive matches to all positive matches found, as formally stated by Equation (3) (Liu, 2021).

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

4. F1-Score: The F1-Score, derived from recall and accuracy, is a crucial evaluation metric. It assesses all facets of the chest X-ray image classification model's performance, including false positives and false negatives. Particularly beneficial for imbalanced datasets, the F1-Score offers a thorough evaluation of model efficacy. It serves as a more dependable measure of performance than accuracy, which may overlook specific errors, as it accounts for both error types. This calculation is presented in the equation. (4) (Hussien & Nemer, 2025).

$$\text{F1} - \text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \qquad (4)$$

5. Confusion Matrix encapsulates the model's predictions in relation to the actual labels, including accurate, erroneous, and false predictions. It can be utilized into pinpoint areas of concern and establish various evaluation criteria. True Positives (TP) refer to accurately identified positives; True Negatives (TN) denote correctly identified negatives; False Positives (FP) indicate inaccurately identified positives (Type I mistake); and False Negatives (FN) signify inaccurately identified negatives (Type II error) (Hussien & Nemer, 2025).

ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 120-152
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 15, 2025
Peer reviewed: December 20, 2025
Revised: February 26, 2026
Published: April 2026

RESULTS

In this section, convolutional neural network models were used into classify pneumonia patients based on chest x-ray images using Python via Google Colab. A dataset is used in this research, taken from previous research, and performance measures are employed. Techniques, such as majority voting and soft voting, were used to improve the performance of convolutional neural network models. The results are discussed.

A dataset from chest X-ray picture consisting from two classes (pneumonia and normal) was used to train the VGG16, DenseNet-121, DenseNet-169, and MobileNet deep learning models, as well as the Fusion-Based Majority Voting (FBMV) and Fusion-Based Smooth Voting (FBSV) models. The total dataset used was 5,878 chest X-ray images, from which 10% (588 images) were used for testing and the remaining 5,290 images were used for training and validation. Using k-fold cross-validation techniques, the data was split into a training dataset (4,232 pictures) and a validation dataset (1,058 images). These models demonstrated exceptional classification performance during the task.

During the training process for each of these models, hyperparameters were modified to improve model performance, achieve better accuracy, reduce losses, and minimise computational resource consumption. Sparse cross-entropy was used for all models, an image size of (224, 224, 3) was used for all images, and data augmentation was used to reduce overfitting. The following paragraphs describe the parameter values used in all models, as well as the results obtained.

VGG16 Model

The VGG16 classification model was optimised to achieve the highest possible accuracy and the lowest possible loss, with no overtraining. This was achieved using the Nadam optimisation algorithm, a learning rate of 0.00001, the ReLU6 activation function, Dropout (0.8), and batch size=64 and epochs=50. The test accuracy was achieved at 97.79%, and the test loss was 0.0752. The model took 61.0252 minutes to execute and consumed 19.1510 GB of memory. Figure 7 displays the training and validation accuracy and loss curves. Figure 8 shows the confusion matrix results for this model.

ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 120-152
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 15, 2025
Peer reviewed: December 20, 2025
Revised: February 26, 2026
Published: April 2026

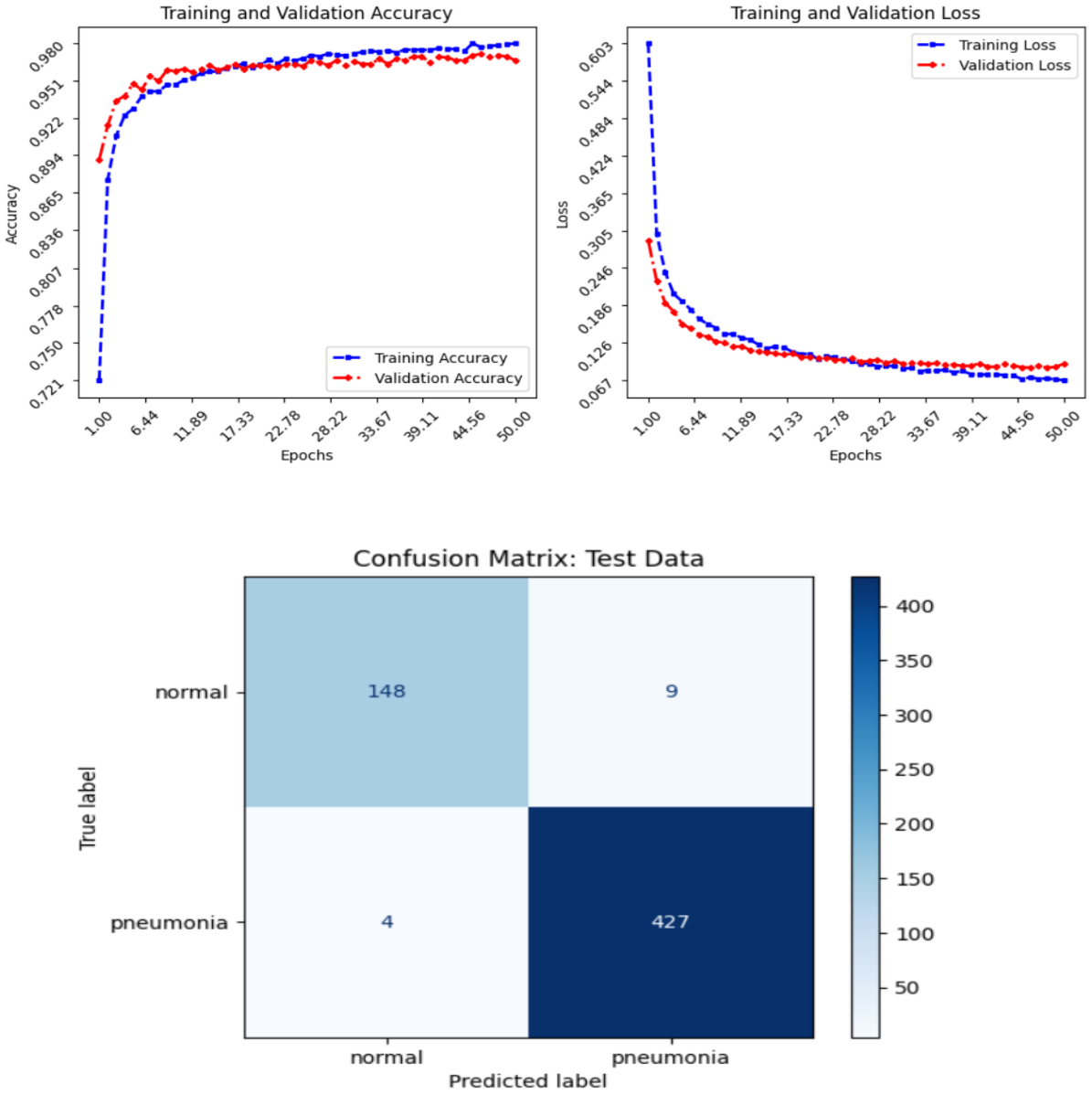


Figure 8: confusion matrix results

Densenet-121 Model

The DenseNet121 classification model settings were modified to achieve the highest possible accuracy and lowest loss with no overtraining by using the Nadam optimisation algorithm, learning rate 0.00001, ReLU6 activation function, Dropout (0.8), batch size=64, and epochs=50. The test accuracy was achieved at 97.11%, and the test loss was 0.0671. The model

ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 120-152
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 15, 2025
Peer reviewed: December 20, 2025
Revised: February 26, 2026
Published: April 2026

took 40.9354 minutes to execute and consumed 19.2864 GB of memory. Figure 9 shows the train and validation accuracy and loss curves. Figure 10 shows the confusion matrix results from this model.

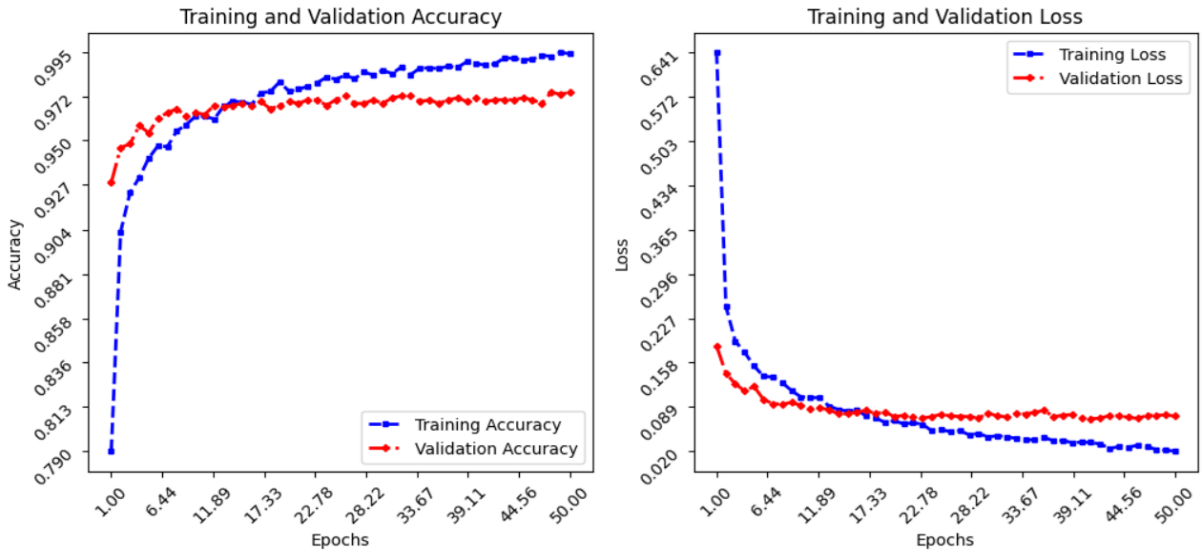


Figure 9: Training and validation accuracy and loss

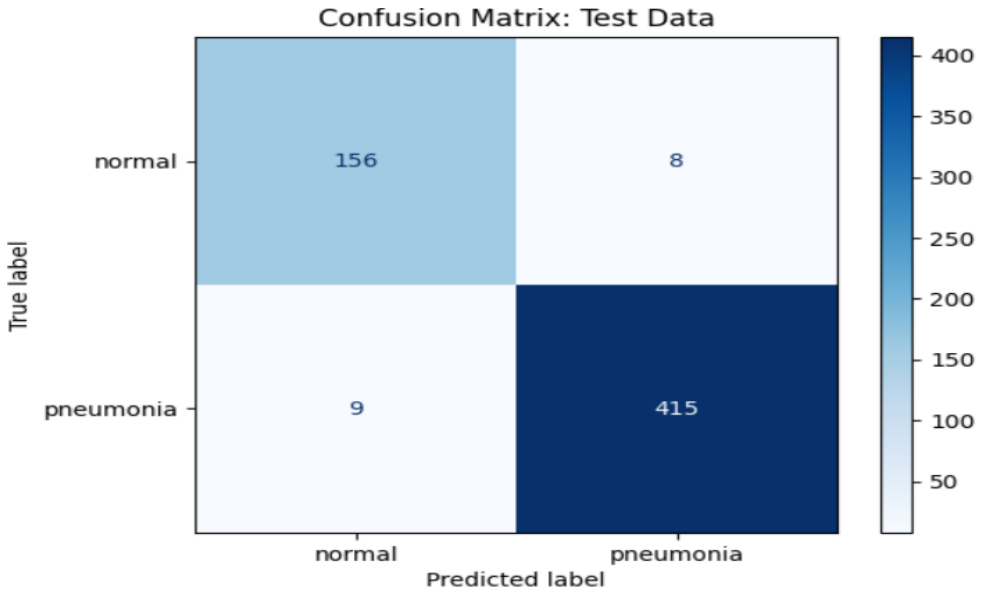


Figure 10: Confusion matrix results.

ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 120-152
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 15, 2025
Peer reviewed: December 20, 2025
Revised: February 26, 2026
Published: April 2026

Densenet-169 Model

The DenseNet169 classification model settings were modified to achieve the highest accuracy possible and lowest loss with no overtraining by using the Nadam optimisation algorithm, learning rate 0.00001, ReLU activation function, Dropout (0.7), batch size=64, and epochs=50. The test accuracy was achieved at 97.62% and Test Loss: 0.0610. The model took 34.6889 minutes to execute and consumed 19.3489 GB of memory. Figure 11 shows the training and validation accuracy and loss curves. Figure 12 shows the confusion matrix results for this model.

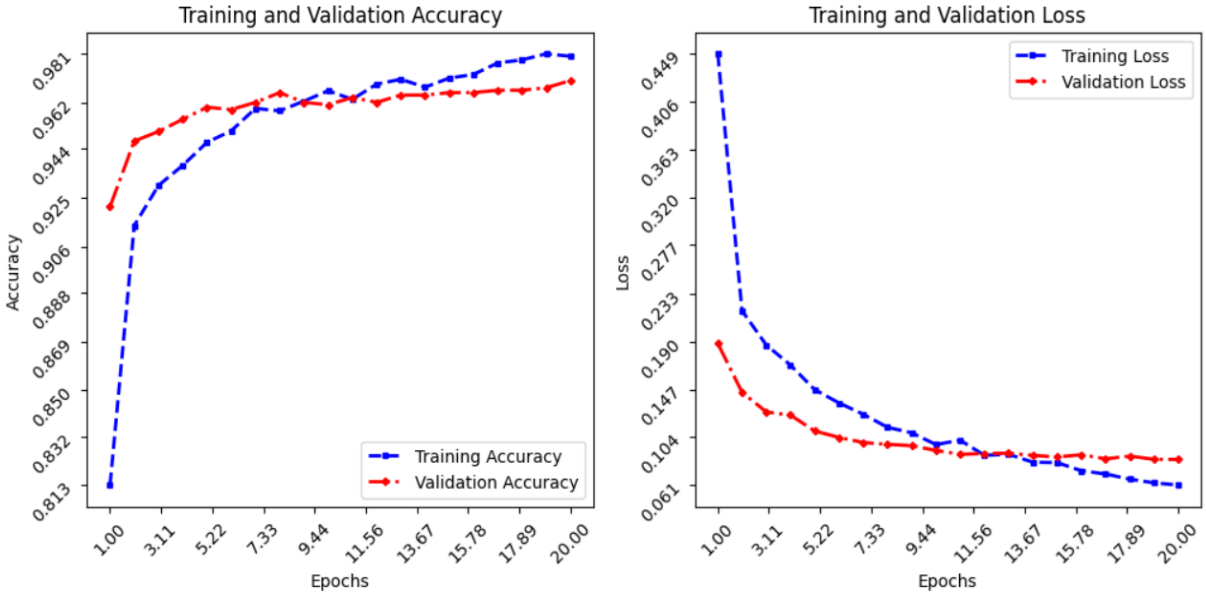


Figure 11: Training and validation accuracy and loss

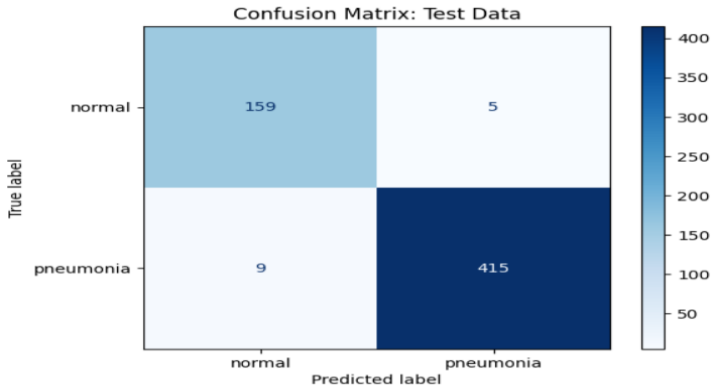


Figure 12: Confusion matrix results

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 120-152
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 15, 2025
Peer reviewed: December 20, 2025
Revised: February 26, 2026
Published: April 2026

Mobile-Net Model

The Mobil Net model classification model settings were modified to obtain the highest possible accuracy and loss, with the lowest possible loss, with no overtraining, using the Nadam optimisation algorithm with a learning rate of 0.00001, the ReLU6 activation function, Dropout (0.6), batch size=64, and epochs=50. The test accuracy was achieved at 97.96%, and Test Loss: 0.0653. The model execution took 17.9079 minutes and consumed 19.1655 GB of memory. Figure 13 displays the training and validation accuracy and loss curves. Figure 14 shows the confusion matrix results for this model.

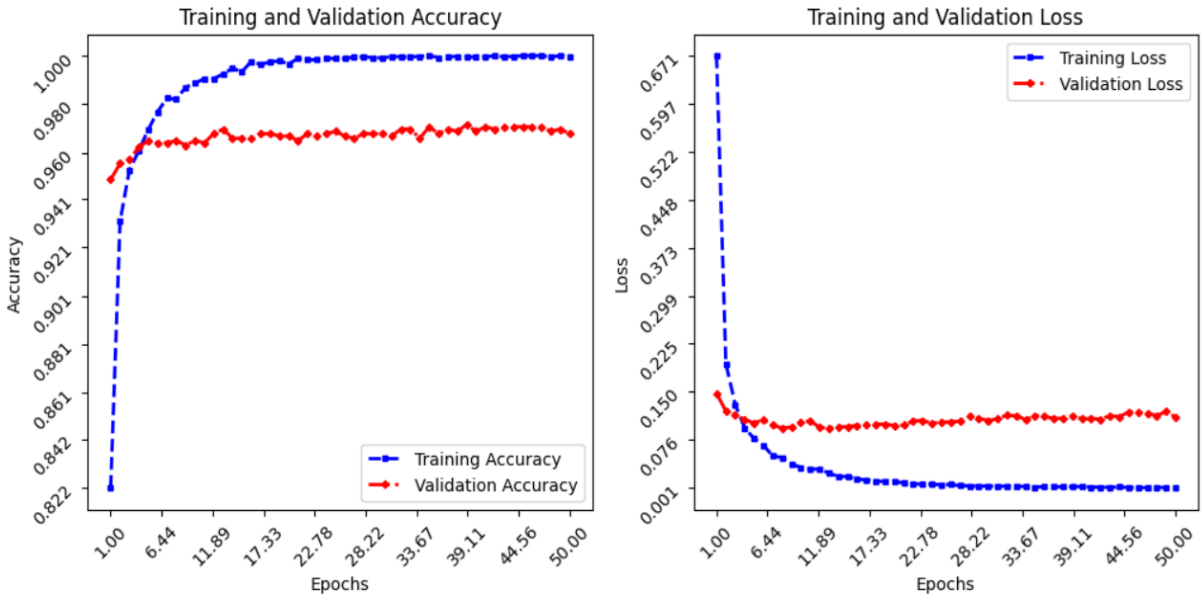


Figure 13: Training and validation accuracy and loss

ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 120-152
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 15, 2025
Peer reviewed: December 20, 2025
Revised: February 26, 2026
Published: April 2026

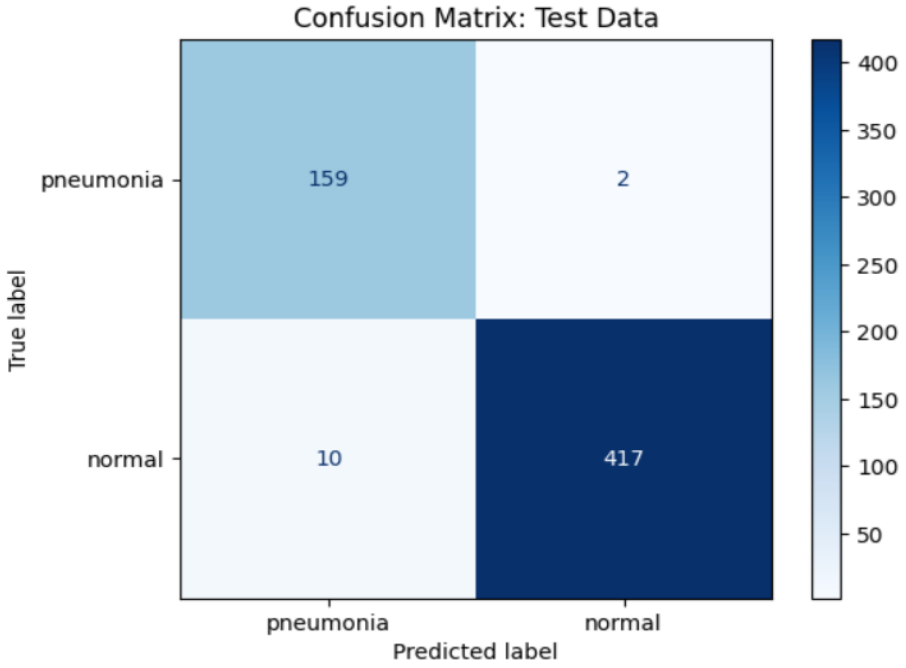


Figure 14: Confusion matrix results.

FBMV Model

The FBMV model classification model settings were modified to obtain the highest possible accuracy and lowest loss with no overtraining by using the Nadam optimisation algorithm, together with a learning rate of 0.00001, the ReLU6 activation function, Dropout (0.6), batch size of 32, and epochs of 10. The test accuracy was achieved at 96.60% and Test Loss: 0.034. The model execution took 112.0887 minutes and consumed 19.5437

GB of memory. Figure 15 displays the training and validation accuracy and loss curves. Figure 16 shows the confusion matrix results for this model.

ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 120-152
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 15, 2025
Peer reviewed: December 20, 2025
Revised: February 26, 2026
Published: April 2026

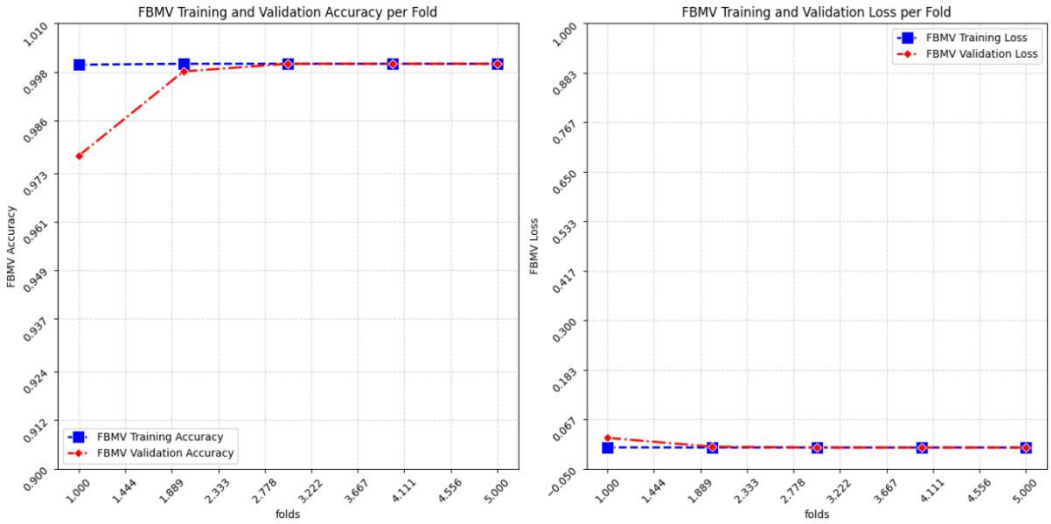


Figure 15: Training and validation accuracy and loss

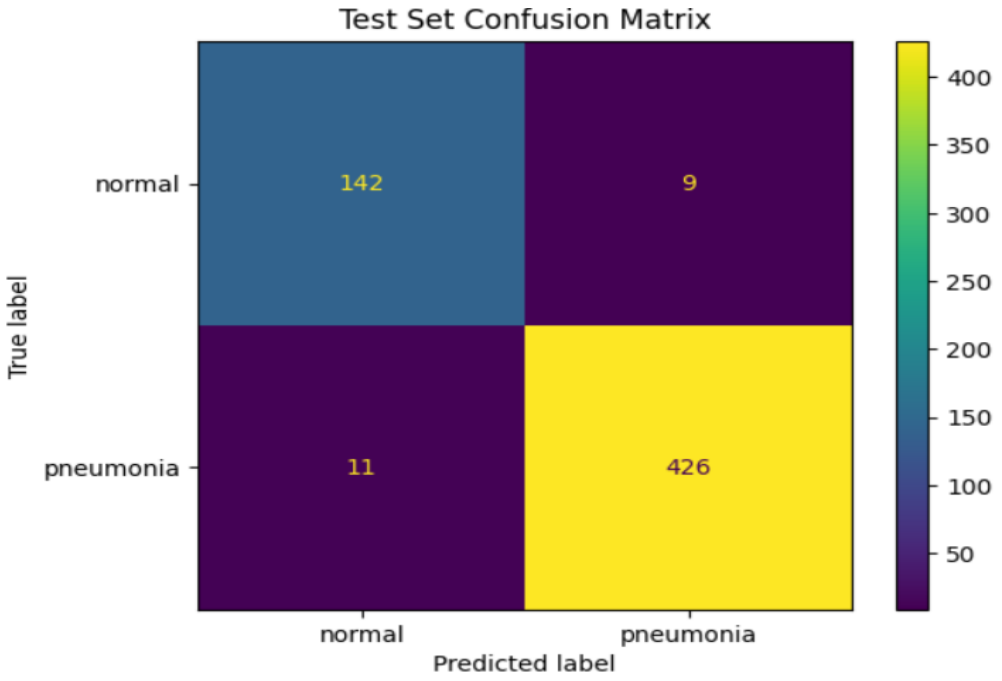


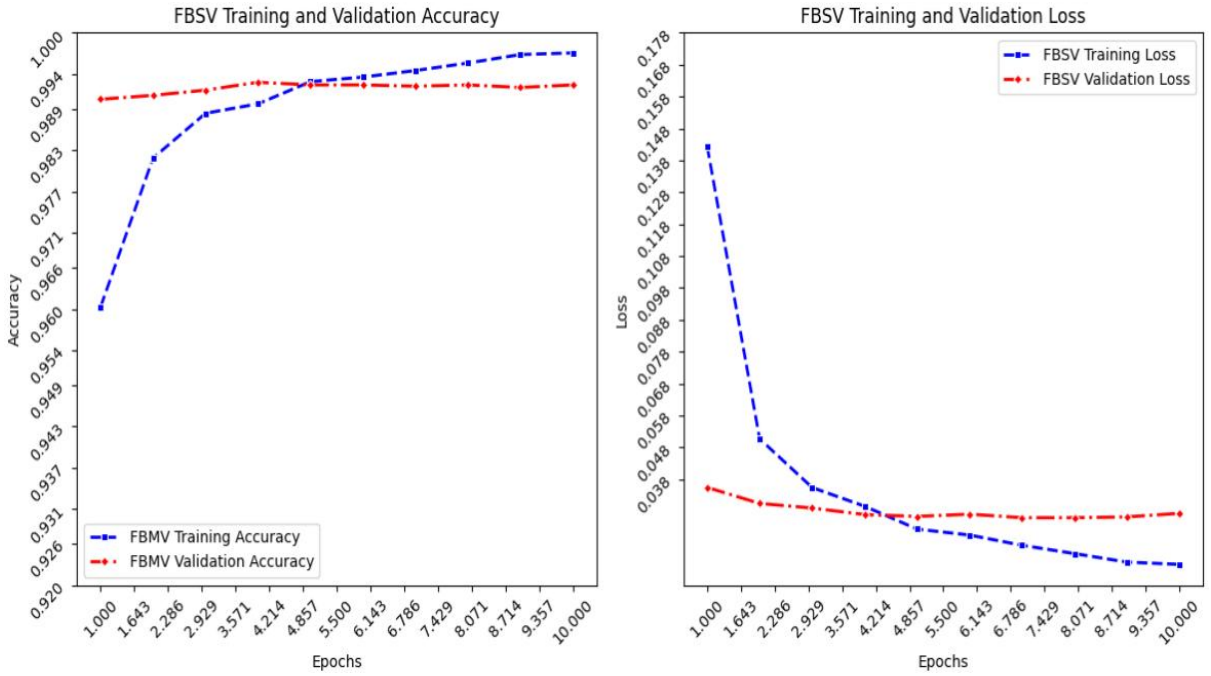
Figure 16: Confusion matrix results

FBSV Model

The FBSV model was optimised to attain the highest possible accuracy and lowest loss with no overtraining by using the Nadam optimisation algorithm, learning rate 0.00001, ReLU

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 120-152
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 15, 2025
Peer reviewed: December 20, 2025
Revised: February 26, 2026
Published: April 2026

scaling function, Dropout (0.6), batch size=64, and epochs=10. The test accuracy was achieved at 97.28%, and Test Loss: 0.0272. The model took 42.0746 minutes to run and consumed 19.2476 GB of memory. Figure 17 displays the training and validation accuracy and loss curves. Figure 18 displays the confusion matrix results for this model.



ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 120-152
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 15, 2025
Peer reviewed: December 20, 2025
Revised: February 26, 2026
Published: April 2026

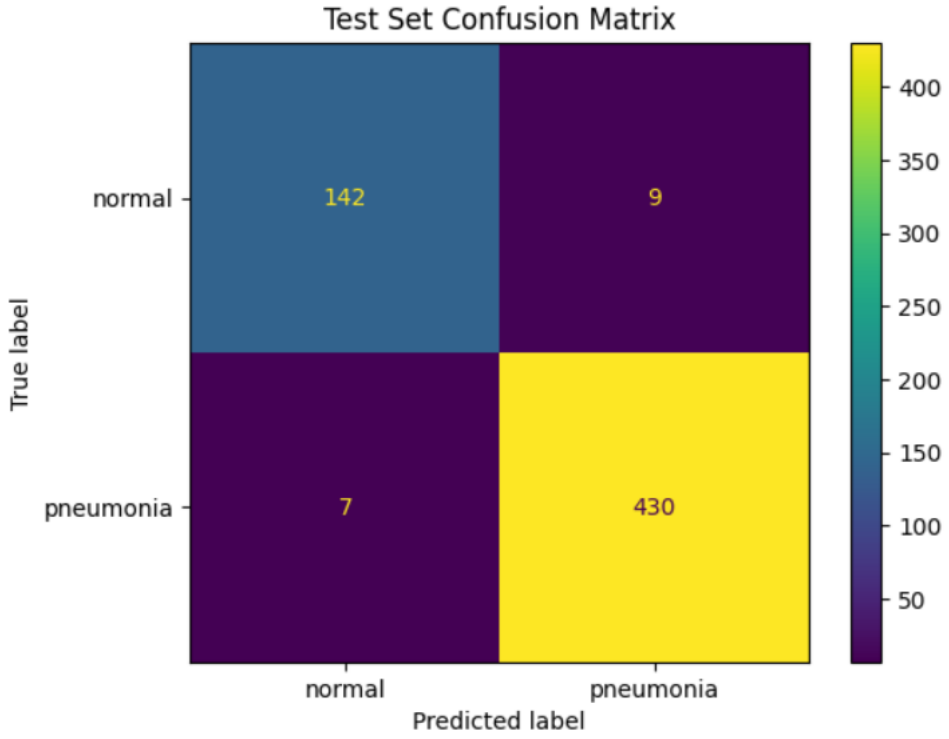


Figure 18: Confusion matrix results.

Models Comparison

After obtaining the results of the convolutional neural network models from classifying chest x-ray images and knowing the values from their performance evaluation and the time taken by means of each model in the classification process, into addition until the size of the memory space used to implement the model, we can determine the best model among the models used in the classification based on the metrics values. Through the data of the models' results for Dataset, which are shown in Table 2

Table 2: Results of all models for the dataset.

Metrics	VGG16	DenseNet-121	DenseNet-169	Mobile Net	FBMV	FBSV	Best model
Training accuracy	99.9	99.28	98.27	99.99	100	98.91	FBMV
Validation accuracy	96.69	96.79	96.86	97.22	99.51	99.2	FBMV
Test accuracy	97.79	97.11	97.62	97.96	96.6	97.28	Mobile Net

ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 120-152
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 15, 2025
Peer reviewed: December 20, 2025
Revised: February 26, 2026
Published: April 2026

Training Loss	0.0656	0.0231	0.0551	0.0017	0.00	0.0357	FBMV
Validation Loss	0.0886	0.0944	0.0907	0.1083	0.0049	0.0281	FBMV
Test Loss	0.0752	0.0671	0.061	0.0653	0.034	0.0272	FBSV
Precession	97	97	98	98	99.38	99.01	FBMV
Recall	98	98	98	98	99.38	99	FBMV
F1-score	98	98	98	98	99.38	99.01	FBMV
Execution Time	61.052	40.935	34.689	17.908	112.09	42.75	Mobile Net
Memory used	19.15	19.29	19.35	19.166	19.54	19.25	VGG16

Based on the results of the models used to classify chest X-ray pictures, the FBMV model performed best in terms of training accuracy (100%) and validation accuracy (99.51%), reflecting its high learning and generalisation capabilities. The training loss (0.00) and validation loss (0.0049) were the lowest compared to the other models, while the precision, recall, and F1-score values (99.38%) were the highest, making it the most reliable model into terms for statistical efficiency. The FBSV model achieved the lowest test loss (0.0272), along with good performance in validation and testing, making it a good balance between performance and accuracy. The MobileNet model excelled in execution time (17.91 seconds) and test performance (97.96%), making it the ideal choice in environments that require speed and time efficiency. All models consumed memory values ranging from 19.1 to 19.5 GB, with the VGG16 model being the least, making it preferable for use in an environment with limited resources. Figure 19 shows a comprehensive comparison of the models across all training, validation, and test accuracy results, F1 scores, time spent, and memory consumed. Figure 20 shows a comprehensive comparison between the models for all loss metrics.

ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 120-152
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 15, 2025
Peer reviewed: December 20, 2025
Revised: February 26, 2026
Published: April 2026

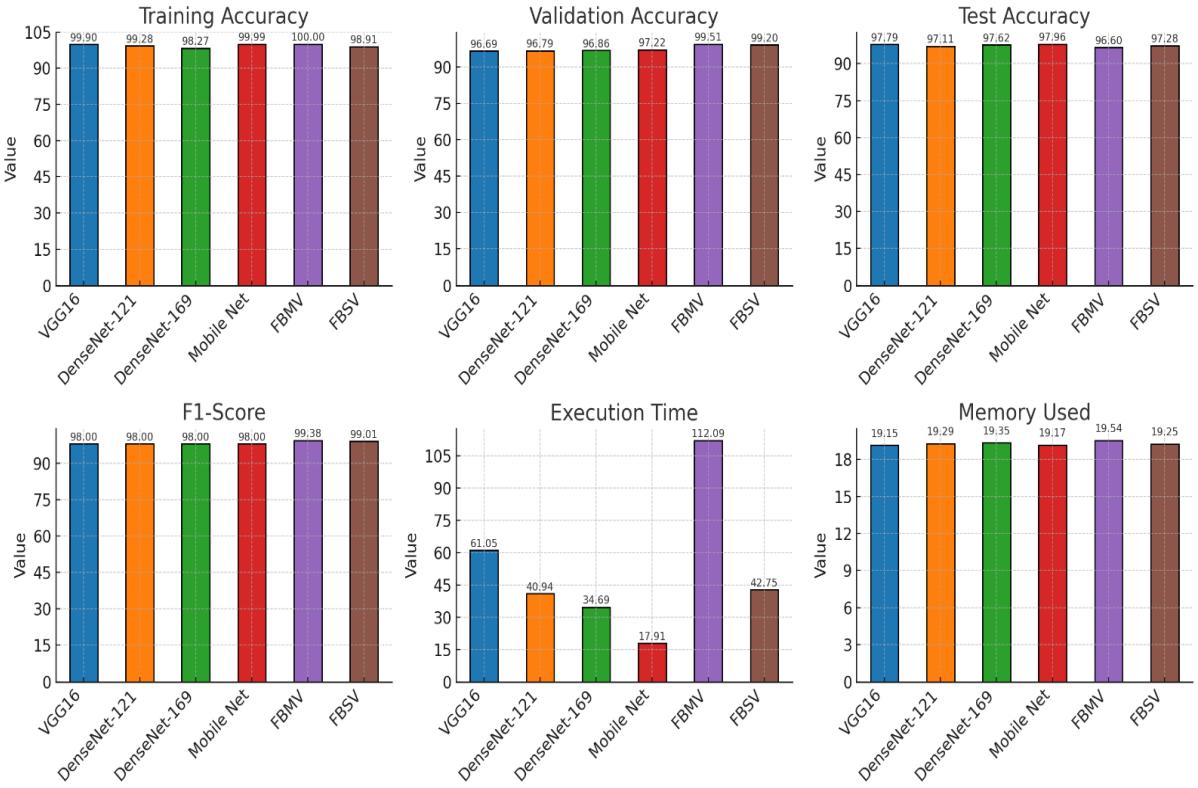


Figure 19: Comprehensive comparison between the results the models

Figure 20: Comparison between the models from all loss metrics

ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 120-152
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 15, 2025
Peer reviewed: December 20, 2025
Revised: February 26, 2026
Published: April 2026

Comparison With Related Works

The proposed models exhibit superiority over prior studies employing identical models, although utilising the same dataset as the previous research. Table (3) presents the detailed comparative results.

Table 3: Comparing propose model results with related work

Title	Methods of detection	Dataset images	Results Accuracy
Rani Puspita et al. (2024) [1]	VGG16 DenseNet121	5,216	VGG16=90 % DenseNet121=88%
Manan Pruthi et al. (2023) [2]	VGG16 Random Forest	---	VGG16= 91.99% Random Forest=75%
Our work	VGG16 DenseNet121 DenseNet169 Mobile Net FBMV FBSV	5878	VGG16 =97.79% DenseNet121=97.11% DenseNet169=97.62% Mobile Net =97.96% FBMV =96.6% FBSV =97.28%

DISCUSSION

The results of this study show that the proposed ensemble-based framework, particularly the FBMV model, achieved superior performance compared to individual deep learning models and several previously published studies. The high training accuracy (100%) and validation accuracy (99.51%) indicate that the model effectively learns complex patterns from chest X-ray images while maintaining strong generalisation. This finding is consistent with prior studies that emphasise the effectiveness of ensemble learning in improving classification performance, such as Seung Min Baik et al. (2024), in which ensemble models outperformed individual classifiers in predicting pneumonia outcomes.

Compared with earlier works, the proposed models in this study achieved higher accuracy than those reported by Rani Puspita et al. (2024), where VGG16 achieved 90% accuracy, and Manan Pruthi et al. (2023), where VGG16 achieved 91.99%. In contrast, the current study achieved 97.79% using the same architecture, which highlights the effectiveness of data preprocessing, augmentation techniques, and hyperparameter tuning applied in this work. Furthermore, the

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 120-152
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 15, 2025
Peer reviewed: December 20, 2025
Revised: February 26, 2026
Published: April 2026

results are comparable to those reported by Vetrithangam et al. (2023), who achieved very high accuracy (99.77%) using an optimised ResNet model, although their study lacked cross-validation, which was addressed in the current research.

Additionally, integrating multiple models via voting (hard and soft) significantly enhanced the robustness and stability of predictions. This aligns with findings from Rajawat et al. (2025), where combining CNN and LSTM with attention mechanisms improved detection performance. However, unlike multimodal approaches that require additional data types (e.g., thermal or audio data), the proposed method in this study achieves competitive performance using only chest X-ray images, making it extra practical for real-world deployment.

Despite these promising results, some limitations remain. As in previous studies, such as Haewon Byeon (2024), reliance on a single dataset may limit the model's generalizability across different clinical environments. Moreover, although the FBMV model achieves the preferred overall performance, it requires more computational time than lightweight models such as MobileNet, which demonstrated faster execution and is more suitable for real-time applications. In summary, the findings confirm that ensemble deep learning approaches provide a reliable and effective solution of pneumonia detection from chest X-ray images, outperforming many traditional and single-model approaches. Future work should focus on improving generalisation across diverse datasets and on enhancing explainability to support clinical decision-making.

CONCLUSION

This study assessed the efficacy of various deep learning models, including VGG16, DenseNet-121, DenseNet-169, MobileNet, FBSV, and FBMV, in identifying pneumonia in chest X-ray images. The test results show the FBMV model had the highest accuracy and the lowest loss. This makes it the best choice for diagnostic applications where accuracy is critical. MobileNet isn't quite as accurate, but it's a great choice for places with limited resources because it can make decisions quickly.

Moreover, ensemble learning models, especially FBMV, consistently outperformed individual models, demonstrating the advantages of integrating various architectures to improve predictive stability and robustness. The study introduces an incoming hybrid framework that combines unsupervised clustering (K-means) with deep learning and ensemble voting (FBMV), enhancing accuracy and robustness in pneumonia detection. It also advances knowledge by comparing multiple CNN architectures and demonstrating the effectiveness of integrating voting mechanisms for more reliable medical image classification

ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 120-152
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 15, 2025
Peer reviewed: December 20, 2025
Revised: February 26, 2026
Published: April 2026

Model interpretability techniques, including Grad-CAM and SHAP, validated that the decisions of the top-performing models aligned with radiologically pertinent areas, hence enhancing their reliability in medical applications.

Ultimately, investigations conducted on balanced and augmented datasets confirmed the models' capacity to generalise effectively to novel data, hence endorsing their applicability for real-world implementation in pneumonia screening systems. The study contributes to society by enabling early and accurate pneumonia detection, which can reduce mortality rates and improve healthcare outcomes, especially for vulnerable populations. For the industry, it provides a practical AI-based diagnostic solution that supports real-time decision-making, reduces workload on clinicians, and can be deployed in resource-limited healthcare environments

REFERENCES

Agard, G., Roman, C., Guervilly, C., Forel, J. M., Orléans, V., Barrau, D., Auquier, P., Ouladsine, M., Boyer, L., & Hraiech, S. (2025). An Innovative Deep Learning Approach for Ventilator-Associated Pneumonia (VAP) Prediction in Intensive Care Units—Pneumonia Risk Evaluation and Diagnostic Intelligence via Computational Technology (PREDICT). *Journal of Clinical Medicine*, *14*(10), 0–18. https://doi.org/10.3390/jcm14103380

Alaidany, A. A. (2024). A Review of IoT-Based Wearable Sensor Systems for Healthcare Monitoring. *AMERICAN Journal of Engineering, Mechanics and Architecture*, *2*(5), 132–159. https://doi.org/10.13140/RG.2.2.18587.27684

Alaidany, A. A. (2025). Improving the Accuracy of Cancer Driver Gene Identification based on Dimensionality Reduction Using Deep AutoEncoders. *International Journal of Intelligent Engineering and Systems*, *18*(9). https://doi.org/10.22266/ijies2025.1031.40

Baik, S. M., Hong, K. S., Lee, J. M., & Park, D. J. (2024). Integrating ensemble and machine learning models for early prediction of pneumonia mortality using laboratory tests. *Heliyon*, *10*(14), e34525. https://doi.org/10.1016/j.heliyon.2024.e34525

Bhattacharjee, V., Priya, A., Kumari, N., & Anwar, S. (2023). DeepCOVNet Model for COVID-19 Detection Using Chest X-Ray Images. *Wireless Personal Communications*, *130*(2), 1399–1416. https://doi.org/10.1007/s11277-023-10336-0

Byeon, H. (2024). Development of a Hybrid Deep Learning Model Combining U-Net and DenseNet for Enhanced Pneumonia Prediction Using Chest X-Ray Images. *Nanotechnology Perceptions*, *20*(5). https://doi.org/10.62441/nano-ntp.v20i5.73

Chakraborty, S., Paul, S., & Hasan, K. M. A. (2022). A Transfer Learning-Based Approach with Deep CNN for COVID-19- and Pneumonia-Affected Chest X-ray Image Classification. *SN Computer Science*, *3*(1), 1–10. https://doi.org/10.1007/s42979-021-00881-5

Çınar, A., Yıldırım, M., & Eroğlu, Y. (2021). Classification of pneumonia cell images using

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 120-152
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 15, 2025
Peer reviewed: December 20, 2025
Revised: February 26, 2026
Published: April 2026

improved ResNet50 model. *Traitement Du Signal*, *38*(1), 165–173. https://doi.org/10.18280/TS.380117

Gabhale, B., Shinde, M., Kamble, A., & Kulloli, M. (2017). Tongue Image Analysis with Color and Gist Features for Diabetes Diagnosis. *International Research Journal of Engineering and Technology (IRJET)*, *4*(4), 523–526. https://www.irjet.net/archives/V4/i4/IRJET-V4I4104.pdf

Guefrechi, S., Jabra, M. Ben, Ammar, A., & Koubaa, A. (2021). *Deep learning based detection of COVID-19 from chest X-ray images*. 31803–31820.

Haitham, A., Amir, A., & Nemer, Z. N. (2025a). Deep Learning-Based Siamese Neural Network for Masked Face Recognition. *Journal of Information Systems Engineering and Management*, *10*, 867–882. https://doi.org/DOI:10.52783/jisem.v10i50s.10403

Haitham, A., Amir, A., & Nemer, Z. N. (2025b). Inclusive Review on Advances in Masked Human Face Recognition Technologies. *Iraqi Journal of Intelligent Computing and Informatics (IJICI)*, *4*(June), 1–17. https://doi.org/10.52940/ijici.v4i1.71

Hussien, G. H., & Nemer, Z. N. (2025). A Comprehensive Review of Advances in Tongue Image Classification Techniques for Diabetes Identification. *Iraqi Journal of Intelligent Computing and Informatics (IJICI)*, *4*(1). https://doi.org/10.52940/ijici.v4i1.90

Jakhar, K., & Hooda, N. (2018). Big data deep learning framework using keras: A case study of pneumonia prediction. *2018 4th International Conference on Computing Communication and Automation, ICCCA 2018*, *August*. https://doi.org/10.1109/CCAA.2018.8777571

Jiang, Z., & Li, J. (2023). Multiple color representation and fusion for diabetes mellitus diagnosis based on back tongue images. *Computers in Biology and Medicine*, *155(1)*(February). https://doi.org/10.1016/j.compbiomed.2023.106652

Li, J., Chen, Q., Hu, X., Yuan, P., Cui, L., Tu, L., Cui, J., Jiang, T., Yao, X., Zhou, C., Lu, H., & Xu, J. (2021). Establishment of noninvasive diabetes risk prediction model based on tongue features and machine learning techniques International Journal of Medical Informatics. *International Journal of Medical Informatics*, *149(1)*(May), 1–5.

Li, J., Cui, L., Tu, L., Hu, X., Wang, S., Shi, Y., Liu, J., Zhou, C., Li, Y., Huang, J., & Xu, J. (2022). Research of the Distribution of Tongue Features of Diabetic Population Based on Unsupervised Learning Technology. *Evidence-Based Complementary and Alternative Medicine*, *2022(4)*, 1–14. https://doi.org/10.1155/2022/7684714

Liu, Z. (2021). Machine learning algorithms in classifying TCM tongue features in diabetes mellitus and symptoms of gastric disease. *European Journal of Integrative Medicine*, *43*, 1–3. https://doi.org/10.1016/j.eujim.2021.101288

Mathew, J. K., & Sathyalakshmi, S. (2024). Sine hunter prey optimization enabled deep residual network for diabetes mellitus detection using tongue image. *Journal of Associated Medical Sciences*, *57*(2), 76–85. https://doi.org/10.12982/JAMS.2024.029

Rana, S., & Gautam, A. K. (2023). Online and Biomedical Engineering. *International Journal of Online and Biomedical Engineering*, *19*(9), 122–130.

Sheu, R.-K., Chen, L.-C., Wu, C.-L., Pardeshi, M. S., Pai, K.-C., Huang, C.-C., Chen, C.-Y., & Chen, W.-C. (2022). Multi-Modal Data Analysis for Pneumonia Status Prediction Using

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 120-152
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 15, 2025
Peer reviewed: December 20, 2025
Revised: February 26, 2026
Published: April 2026

Deep Learning (MDA-PSP). *Diagnostics*, *12*(7). https://doi.org/https://doi.org/10.3390/diagnostics12071706

Taye, M. M. (2023). Understanding of Machine Learning with Deep Learning : *Computers MDPI*, *12*(91), 1–26. https://doi.org/doi.org/10.3390/computers12050091

Thirunavukkarasu, U., Umapathy, S., Ravi, V., & Alahmadi, T. J. (2024). Tongue image fusion and analysis of thermal and visible images in diabetes mellitus using machine learning techniques. *Scientific Reports*, *14(1)*, 1–17. https://doi.org/10.1038/s41598-024-64150-0

Vetrithangam, D., Satve, P. P., Kumar, J. R. R., Anitha, P., Vidhya, S., & Saini, A. K. (2023). Prediction of Pneumonia Disease From X-Ray Images Using a Modified Resnet152V2 Deep Learning Model. *Journal of Theoretical and Applied Information Technology*, *101*(17), 6929–6942.

ISSN: 2408-7920