Discover Applied Sciences

Research

Enhancing deep learning for pneumonia detection: developing web based solution for Dr. Sumait Hospital in Mogadishu Somalia

Abdulaziz Yasin Nageye[1] · Abdukadir Dahir Jimale[1] · Mohamed Omar Abdullahi[1] · Yahye Abukar Ahmed[1] · Mohamed Abdirahman Addow[1]

Received: 7 September 2024 / Accepted: 17 March 2025
Published online: 05 April 2025

Abstract
In the realm of medical imaging, accurate and efficient detection of pneumonia from chest X-ray images is crucial for timely diagnosis and treatment. This study explores the performance of four deep learning models—Simple CNN, DenseNet121, VGG16, and InceptionV3—using the Kaggle "Chest X-Ray Images (Pneumonia)" dataset, which comprises 5,863 images categorized into normal and pneumonia classes. The methodology included data normalization, augmentation, and training, followed by evaluation based on accuracy, precision, recall, and F1-score. The results revealed that Simple CNN achieved the highest accuracy at 92%, with notable precision (0.95 for normal and 0.90 for pneumonia) and recall (0.83 for normal and 0.97 for pneumonia). VGG16 also performed well with an accuracy of 91%, while DenseNet121 and InceptionV3 had lower performance, with InceptionV3 exhibiting the lowest accuracy (84%) and higher false positive rates. Based on these findings, Simple CNN was chosen for deployment in a Django-based web application hosted on AWS, aimed at improving diagnostic accuracy and supporting healthcare professionals at Dr. Sumait Hospital. The study underscores the efficacy of Simple CNN for clinical applications and suggests future enhancements such as dataset diversification, multi-class classification, real-time processing, and the incorporation of additional clinical data.

Article Highlights

- Developed a web application for Dr. Sumait Hospital that utilizes deep learning to detect pneumonia from chest X-ray images.
- The Simple CNN model was identified as the most accurate and reliable for this task among the models tested.
- The system enhances diagnostic speed and accuracy, supporting doctors in delivering improved patient care.

Keywords Pneumonia Detection · Deep Learning Models · CNN · DenseNet121 · VGG16 · InceptionV3

✉ Abdulaziz Yasin Nageye, nageye8@simad.edu.so | [1]Faculty of Computing, SIMAD University, Mogadishu, Somalia.

1 Introduction

Pneumonia is an inflammatory lung condition primarily affecting the alveoli, characterized by symptoms such as cough, fever, and difficulty breathing. It can be caused by various infectious agents, including bacteria, viruses, fungi, and parasites, leading to fluid accumulation in the lungs, which impairs oxygen exchange [1, 2]. Pneumonia is particularly prevalent among children and the elderly, with significant mortality rates, especially in developing countries where it remains a leading cause of death in children under five [3]. The disease can manifest in different forms, affecting one or both lungs, and is often exacerbated by environmental factors and underlying health conditions [4, 5]. Pneumonia is classified based on where it was acquired, including community-acquired pneumonia (CAP), hospital-acquired pneumonia (HAP), and ventilator-associated pneumonia (VAP). Symptoms typically include cough, chest pain, fever, difficulty breathing, and fatigue, with variations in presentation among different age groups [6].

The current stage of pneumonia in Somalia reflects a significant public health challenge, particularly among vulnerable populations such as children and those in crowded living conditions. Banadir Hospital revealed that pneumonia remains a critical issue for children under five, with risk factors including stunting, wasting, and poor air quality. The prevalence of pneumonia in this demographic is exacerbated by socio-economic factors, such as low caregiver education and inadequate housing conditions [7]. In a tertiary care setting, the emergence of multidrug-resistant non-fermentative gram-negative bacilli (MDR-NFGNB) has been alarming, with a prevalence of 8% among hospital-acquired pneumonia cases. The mortality rate associated with these infections was notably high at 42.5% [8].

Artificial Intelligence (AI) is transforming multiple sectors, including public health surveillance. In Africa, where health systems often face significant obstacles like resource limitations, insufficient infrastructure, ineffective health information systems, and a lack of skilled healthcare professionals, AI presents a groundbreaking opportunity for improvement[9, 10].Artificial intelligence (AI) systems have been increasingly applied to improve the diagnosis, prediction, and management of pneumonia, particularly in the context of chest radiographs and CT images. For instance, studies have shown that AI models can accurately detect pneumonia from chest radiographs, often rivaling or even surpassing the performance of human radiologists. These models typically involve deep learning algorithms such as convolutional neural networks (CNNs) and Mask R-CNN, which are trained on large datasets of labeled images to identify pneumonia lesions and classify them accurately[11, 12]. In the case of community-acquired pneumonia, AI systems have been used to predict patient mortality by analyzing clinical data at initial presentation. Various machine-learning methods, including neural networks, rule-learning techniques, and causal discovery methods, have been evaluated for their predictive accuracy, showing promising results with error rates comparable to traditional statistical methods [13, 14].

AI systems in detecting pneumonia, particularly in low-resource settings like Somalia, is a complex issue with both promising potential and significant challenges. AI models have demonstrated high accuracy in detecting pneumonia from chest radiographs, often rivaling or even surpassing human radiologists[9, 11]. However, the implementation of these systems in Somalia faces numerous hurdles, including inadequate healthcare resources, poor health financing, and limited access to technology and digital skills among healthcare professionals [9]. AI systems for detecting pneumonia in Somali hospitals faces significant challenges due to limited healthcare resources and infrastructure. The country struggles with inadequate medical equipment, supplies, and a shortage of trained healthcare professionals, making it difficult to implement advanced technologies like AI. Additionally, poor health financing hampers investment in new technologies, while many hospitals lack reliable electricity, computers, and internet access, which are essential for running AI systems [15]. Implementing machine learning (ML) in Somali hospitals faces significant challenges, primarily due to resource limitations, data scarcity, and infrastructural issues. Key obstacles include unreliable internet connectivity, which hampers the deployment of traditional web-based applications, as highlighted by the use of progressive web applications (PWAs) that can function offline [16]. Additionally, the integration of ML into clinical workflows is complicated by the need for extensive data extraction and manual review, which can be time-consuming and inefficient [17].

In this study, we propose the development of an AI-based pneumonia detection system utilizing chest X-ray images to enhance diagnostic efficiency in hospitals across Mogadishu, Somalia particularly Dr. Sumait Hospital of SIMAD University. This initiative aims to address the current challenges faced by these hospitals, including limited resources and inadequate diagnostic capabilities. The primary focus of this study is to identify an existing model that balances high diagnostic accuracy with computational efficiency for deployment in resource-constrained environments, rather than building a new model from scratch.By implementing a deep learning model, specifically a specialized neural network, we seek to analyze digital X-ray images that can significantly improve the accuracy of pneumonia detection. Our objectives include examining the benefits of such a system, identifying potential challenges, and proposing practical solutions for successful implementation. Through this research, we aim to gain valuable insights into the application of machine learning in Mogadishu's healthcare landscape, ultimately laying the groundwork for a more sustainable and effective healthcare future in the region. This study contributes to bridging the gap between AI research and practical healthcare applications in resource-constrained settings. By focusing on lightweight and efficient architectures, such as Simple CNN, and deploying the model as a web application tailored to Dr. Sumait Hospital, this work addresses the unique challenges of implementing AI in Somalia. Unlike prior studies that prioritize model complexity, this research demonstrates a practical approach to achieving high diagnostic accuracy while considering real-world constraints. Additionally, the integration of this model into a user-friendly web-based platform further streamlines its application in clinical environments, providing healthcare professionals with an accessible and reliable diagnostic tool.

This paper begins next sections with a Literature Review that explores existing research on deep learning models for chest X-ray image classification, highlighting the current state of the art and identifying gaps in the literature. Following this, the Methodology section details the dataset, preprocessing steps, model selection, and web deployment processes. The Results section provides a comprehensive analysis of the performance of four deep learning models—Simple CNN, DenseNet121, VGG16, and InceptionV3—based on accuracy, precision, recall, and F1-score. We then discuss these findings in the Discussion section, comparing them to existing literature, examining practical implications, and addressing challenges and limitations. The paper concludes by summarizing the impact of the chosen model, the Simple CNN, which was deployed in a web application for Dr. Sumait Hospital of SIMAD University to enhance diagnostic accuracy and support healthcare professionals.

2 Related work

In the study of pneumonia diagnosis, various research efforts have focused on the critical challenges and limitations associated with traditional diagnostic methods. To enhance diagnostic efficiency in hospitals across Mogadishu, AI-based pneumonia detection systems utilizing chest X-ray images have been explored. These systems leverage advanced technologies such as deep learning algorithms, neural architecture search, and machine learning to improve the accuracy and speed of pneumonia detection. Researchers globally have proposed several studies and methods, including the use of convolutional neural networks and transfer learning, to develop robust AI models capable of detecting pneumonia from chest X-ray images.

Tatiana Gabruseva et al. highlights their study the advancements of deep learning approaches over traditional machine learning methods in various computer vision and medical imaging tasks, particularly in detection, classification, and segmentation. It references the RSNA Pneumonia Detection Challenge, where the proposed solution utilized a single-shot detector (SSD) and deep convolutional neural networks (CNNs) with augmentations and multi-task learning to effectively locate lung opacities in chest radiographs. The study emphasizes the performance of these deep learning models, which have shown significant improvements in accuracy compared to conventional techniques [18].

Amer Kareem et al. focuses their study on various machine learning (ML) methods applied to medical image detection, particularly for pneumonia. It reviews different algorithms, evaluation methods, and datasets used in previous research. Key evaluation metrics include accuracy, precision, recall, F1 score, ROC, and AUC, which are essential for comparing the existing model. The findings indicate that larger datasets significantly enhance the effectiveness of ML models in detecting medical images compared to traditional methods. However, many studies suffer from limitations due to smaller datasets, which impact their accuracy and reliability in real-world applications [19].

Patrik Szepesia and Laszlo Szilagyi discusses various existing architectures for pneumonia detection using chest X-ray images. It highlights the performance of the proposed CNN model compared to state-of-the-art solutions, emphasizing that while some models have been tested under different conditions, the proposed architecture demonstrates superior accuracy and efficiency. The study also notes the importance of retraining algorithms with local population data to ensure reliable clinical applications, as existing datasets may not represent the diverse patient population adequately. Additionally, it references the use of transfer learning techniques in medical image analysis, which the authors chose to avoid in favor of developing a model from scratch [20].

El Asnaou et al. highlights the challenges in diagnosing diseases like pneumonia using medical imaging, emphasizing the increasing difficulty for radiologists due to the rising number of patients. It reviews various deep learning (DL) approaches that have shown promising results in pneumonia detection, including the use of convolutional neural networks (CNNs) and specific architectures like AlexNet, VGG-16, and DenseNet201. The study also mentions recent advancements in DL frameworks that address issues like overfitting and feature loss, showcasing the effectiveness of these methods in improving diagnostic accuracy and efficiency in medical image analysis [21].

Shagun Sharma and Kalpna Guleria primarily emphases on the application of deep learning models for pneumonia detection using chest X-ray (CXR) images. Notably, the study highlights the use of the VGG16 architecture combined with neural networks (NN) to achieve significant performance metrics. For the first dataset, the model achieved an accuracy of 92.15%, with precision at 0.9428, recall at 0.9308, and an F1-score of 0.937. In a second dataset, the model demonstrated even better results, reaching an overall accuracy of 94.5%, with precision, recall, and F1-score all at 0.954. The study also compares the performance of VGG16 with NN against other classifiers, such as Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Random Forest (RF), and Naïve Bayes (NB), showing that VGG16 with NN outperforms these methods in both datasets. This work emphasizes the effectiveness of deep learning techniques in enhancing pneumonia detection accuracy from CXR images [22].

Rana and Marwaha proposed a novel approach to pneumonia detection using federated learning (FL) and unsupervised learning techniques, addressing challenges of data privacy and limited labeled datasets in healthcare. Their methodology involves a hybrid model combining deep autoencoders for feature extraction and variational autoencoders for classification. The model is trained in a decentralized FL setting, ensuring patient data privacy by training locally on client datasets and using the FedAvg algorithm to aggregate updates. The proposed framework achieved a high accuracy of 94% through fivefold cross-validation, outperforming traditional centralized models. This study highlights the potential of FL to create efficient and privacy-preserving diagnostic tools, particularly in resource-constrained healthcare environments. The insights and techniques presented in this work align closely with efforts to develop robust AI-driven medical applications without compromising data security [23].

Existing studies on pneumonia detection using deep learning, such as those employing complex architectures like DenseNet and ResNet, have demonstrated high diagnostic accuracy but often at the cost of increased computational requirements. In contrast, this study focuses on balancing diagnostic performance with computational efficiency to ensure practical deployment in resource-constrained settings. Unlike prior research, this work also integrates the selected model into a scalable web application, demonstrating its usability in real-world clinical workflows.

In the context of pneumonia diagnosis, existing research has explored various AI-based systems utilizing chest X-ray images to enhance diagnostic efficiency globally. However, there remains a significant gap in the development and deployment of such systems specifically in Mogadishu Somalia. This gap highlights the opportunity and necessity for localized solutions that consider the unique socio-economic and healthcare infrastructure challenges of Mogadishu, ensuring effective diagnostic measures tailored to the region's needs. By addressing this gap, this study aims to pioneer the integration of AI technologies in enhancing pneumonia diagnosis, thereby filling a significant void in current research and practical applications within Somalia.

3 Methodology

This section describes the steps taken to develop, train, evaluate, and deploy deep learning models for detecting pneumonia from chest X-ray images. The methodology is divided into four key areas: dataset description, data preprocessing, model selection, and web deployment. The goal was to create a reliable system for Dr. Sumait Hospital at SIMAD University to assist healthcare professionals in making accurate decisions and minimizing human error. Figure 1 shows proposed system architecture.

3.1 Dataset description

The "Chest X-Ray Images (Pneumonia)" dataset from Kaggle was utilized in this study[24]. This dataset consists of 5,863 chest X-ray images categorized into two classes: Normal and Pneumonia. The dataset is further divided into three subsets: a training set, a validation set, and a test set. The training set was used to train the machine learning models, while the validation set, comprising 16 images (8 normal and 8 pneumonia cases), was utilized to fine-tune the model and monitor its performance. Although the size of the validation set is limited, we implemented additional measures to ensure model

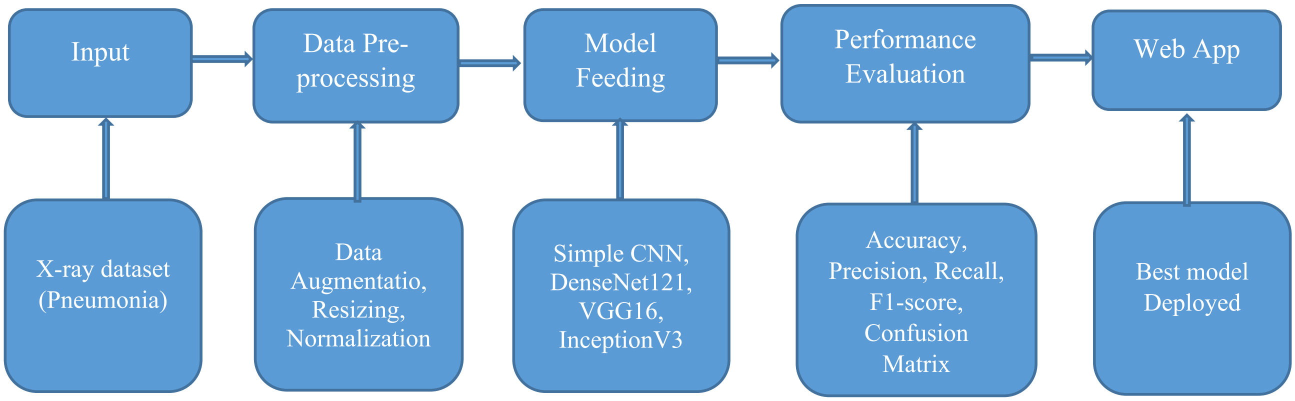


Fig. 1 Proposed model architecture

reliability. These included k-fold cross-validation on the training set, early stopping based on validation performance, and extensive data augmentation during training. These steps enhanced the model's ability to generalize despite the small validation set size The test set, on the other hand, was reserved for evaluating the performance of the final model. This dataset was chosen due to its comprehensive representation of both normal and pneumonia cases, providing a robust foundation for model training and evaluation. Its diversity in image samples allows for effective model learning and generalization.

As shown in Table 1, the dataset contains a total of 5,863 images, with a majority allocated to the training set to ensure the models can learn effectively. The validation and test sets are smaller, which helps in fine-tuning the models and evaluating their performance, respectively.

3.2 Data preprocessing and augmentation

Data preprocessing and augmentation are critical steps in preparing the chest X-ray images to ensure the model's robustness and generalization capabilities. Challenges such as variations in image quality, inconsistent lighting, and occasional artifacts were addressed by applying normalization to standardize pixel intensity values, resizing images to 224×224 pixels, and using data augmentation techniques to improve the model's ability to handle variability. The preprocessing began with normalization, where pixel values of all images were rescaled to a range of [0, 1] by dividing by 255. This standardization step ensures that the models receive inputs that are on a similar scale, which is essential for stable convergence during training and prevents issues like slow learning or vanishing gradients.

To further improve the robustness of the model and reduce overfitting, and address class imbalance, a range of data augmentation techniques were applied. These techniques included random rotations (up to 20 degrees), width and height shifts (with a maximum shift of 20% of the image dimensions), random zooming (up to 10%), and horizontal flipping. The goal of these transformations was to artificially expand the training dataset, introducing variability and simulating real-world variations in X-ray images, such as misalignment, varying image quality, and positioning. By doing so, the model learns to identify pneumonia from a broader set of conditions, improving its ability to generalize to unseen

Table 1 Dataset distribution

Set	Class	Number of images
Training	Normal	1341
	Pneumonia	3875
Validation	Normal	8
	Pneumonia	8
Test	Normal	234
	Pneumonia	390
Total		5863

data. Additionally, all images were resized to 224 × 224 pixels to meet the input size requirements of the pre-trained models such as VGG16, DenseNet121, and InceptionV3. This resizing ensures that all images are consistent in size, which is critical for efficient model training and evaluation.

These preprocessing and augmentation steps were designed to optimize the model's performance by ensuring consistent data input and increasing its ability to handle variability in real-world medical imaging data. This process helps to mitigate overfitting and enhances the model's ability to generalize, leading to more reliable pneumonia detection in clinical applications.

3.3 Model selection

In this study, four deep learning models were selected for evaluation: **Simple CNN**, **DenseNet121, VGG16, and InceptionV3**. These models were chosen based on their unique architectures, computational efficiency, and prior success in medical image classification tasks. Each model's architecture provides distinct advantages, allowing for a comprehensive evaluation of their capabilities in detecting pneumonia from chest X-ray images. To ensure a fair comparison, the same hyperparameter optimization strategy was applied to all models, including a grid search for learning rate, batch size, and number of epochs, with early stopping based on validation loss.

3.3.1 Simple CNN

The Simple Convolutional Neural Network (CNN) was custom-built for this project to serve as a baseline model. This model consists of multiple convolutional layers followed by max-pooling layers and fully connected dense layers. The convolutional layers are designed to automatically learn spatial hierarchies of features from input images, making them suitable for image classification tasks. The Simple CNN is relatively lightweight, with fewer parameters compared to more complex models. This simplicity allows for faster training and evaluation, making it ideal for initial testing and comparison. The model architecture consists of three convolutional layers with ReLU activation, each followed by a max-pooling layer to downsample the feature maps. Finally, the output is passed through two fully connected layers, the last of which uses a softmax activation function to predict the class probabilities. This architecture is advantageous for its straightforward design and ability to achieve decent performance with minimal computational resources.

3.3.2 DenseNet121

DenseNet121 is a deep learning model from the Dense Convolutional Network (DenseNet) family, known for its densely connected architecture. Unlike traditional CNNs, where each layer has its own set of weights and connections, DenseNet121 connects each layer to every other layer in a feed-forward fashion. This dense connectivity pattern ensures maximum information flow between layers, reducing the vanishing gradient problem and promoting feature reuse. DenseNet121 is particularly effective for image classification tasks as it enhances feature propagation and encourages the reuse of learned features, leading to more efficient parameter use. This model has 121 layers, which include dense blocks interspersed with transition layers. The dense connections help to mitigate the gradient vanishing problem during backpropagation, which is a common issue in deep neural networks. The DenseNet121 model was chosen for its ability to capture complex features with fewer parameters, making it both computationally efficient and effective in medical imaging tasks.

3.3.3 VGG16

The VGG16 model is a deep CNN architecture that has been widely used in image classification tasks, including medical imaging. VGG16 is characterized by its simple and uniform architecture, consisting of 16 layers: 13 convolutional layers followed by three fully connected layers. The model uses small 3 × 3 filters throughout all convolutional layers, which allows it to learn fine-grained features from the input images. VGG16's depth and simplicity make it a robust choice for classification tasks, as its straightforward architecture can effectively capture intricate patterns in chest X-ray images. The model employs max-pooling layers after every few convolutional layers to reduce spatial dimensions while preserving important features. VGG16 ends with a softmax activation function in the output layer for classification purposes. Despite being relatively deep, VGG16 is computationally expensive and requires significant memory, but its high performance in many classification benchmarks justifies its use in this study. The model's ability to capture hierarchical features makes it particularly suitable for distinguishing between normal and pneumonia-affected X-rays.

3.3.4 InceptionV3

InceptionV3 is an advanced deep learning model that incorporates several innovations to enhance performance, including the use of inception modules, which allow the model to learn multiple levels of features simultaneously. This architecture uses a combination of different convolutional filter sizes within the same module to capture various levels of detail from the input images. InceptionV3 also employs techniques such as factorized convolutions, regularization, and label smoothing to improve the model's efficiency and accuracy. The model's depth and architectural complexity make it capable of capturing both global and local features, which is critical for accurately identifying pneumonia from chest X-rays. InceptionV3 was chosen for its ability to handle complex datasets and provide a balance between computational efficiency and classification accuracy. It has been extensively validated in numerous image classification tasks, proving its robustness and reliability in detecting subtle differences in medical images.

These models were selected to provide a broad evaluation of different neural network architectures, each with distinct features and advantages. A weighted loss function was used during training to ensure balanced treatment of both classes, mitigating potential bias from the class imbalance. This diverse selection allowed for a comprehensive evaluation of model performance across various deep learning architectures, ultimately enabling the identification of the most effective model for detecting pneumonia in chest X-ray images.

3.4 Web deployment

To ensure practical application and accessibility, the models were deployed as a web application specifically designed for Dr. Sumait Hospital at SIMAD University. The objective was to create a tool that aids healthcare professionals in making accurate and timely diagnoses, thereby minimizing the risk of human error. The deployment process involved several steps to ensure the application was robust and user-friendly.

. The best-performing model, determined by metrics such as accuracy, precision, recall, and F1-score, was serialized using the Keras.h5 format. The web application was developed using Django, a high-level Python web framework that promotes rapid development and clean, pragmatic design. The application was designed with scalability and extensibility in mind, leveraging AWS cloud infrastructure to handle increased computational and storage demands. Modular backend architecture allows for the seamless addition of new features or diagnostic capabilities, such as support for multi-condition detection or integration with hospital management systems via RESTful APIs. Django was chosen for its robustness, scalability, and built-in security features, making it ideal for deploying a machine learning model in a clinical setting.

The backend of the web application, developed in Django, manages core functionalities including image upload, preprocessing, model inference, and returning prediction results. This setup ensures an efficient workflow from data input to diagnosis output, providing a seamless experience for healthcare professionals. A user-friendly web interface was created using Django templates, HTML, CSS, and JavaScript, enabling users to upload chest X-ray images and receive diagnostic predictions quickly.

The application was deployed on **Amazon Web Services (AWS)**, leveraging its scalable, secure, and reliable infrastructure. AWS services such as EC2 for computing power, S3 for storage, and RDS for database management were utilized to provide a robust deployment environment. This setup ensures that the application can handle real-time predictions, accommodate multiple concurrent user requests, and maintain high availability, making it a practical tool for daily clinical use.

The designed web-based platform acts as a seamless interface between the Simple CNN model and end-users, ensuring ease of use for healthcare professionals. The platform allows users to upload chest X-ray images, process them in real time, and receive diagnostic results with a clear visualization of predictions. Developed using Django, the platform is optimized for accessibility, enabling integration into existing hospital systems. Its design also supports scalability, allowing for future enhancements such as multi-condition detection and incorporation of additional patient data.

By deploying the model on a Django-based web platform using AWS, this study provides a valuable resource for Dr. Sumait Hospital. The application empowers healthcare professionals to make more accurate diagnostic decisions,

Table 2 Performance evaluation of simple CNN

Class	Precision	Recall	F1-score	support
Normal	0.95	0.83	0.88	234
Pneumonia	0.90	0.97	0.94	390
Accuracy			0.92	624
Macro avg	0.93	0.90	0.91	624
Weighted avg	0.92	0.92	0.92	624

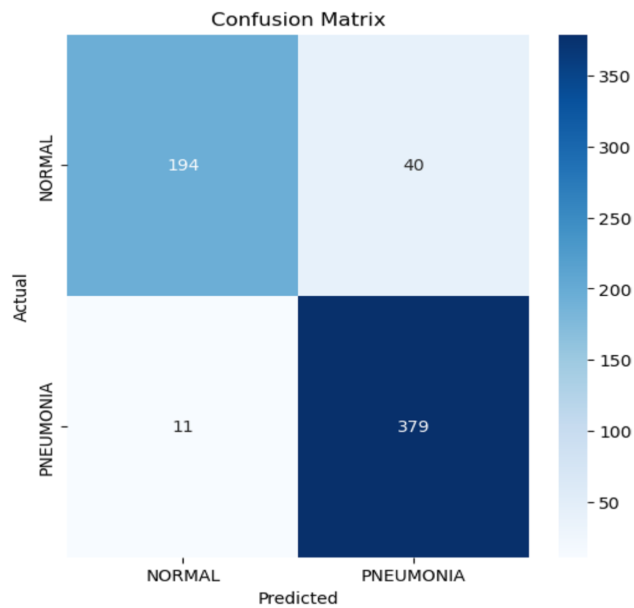


Fig. 2 Confusion Matrix of Simple CNN

reduces the likelihood of human error, and ultimately improves patient outcomes. This integration of advanced machine learning models with web technologies highlights the potential to enhance healthcare delivery through technological innovation.

4 Results and discussions

4.1 Results

The results of this study present the performance metrics of four deep learning models—Simple CNN, DenseNet121, VGG16, and InceptionV3—used to detect pneumonia from chest X-ray images. These models were evaluated on the test dataset based on accuracy, precision, recall, F1-score, and confusion matrices.

4.1.1 Simple CNN model

Simple CNN model demonstrated a robust performance with an accuracy of **0.92**. It showed a high precision of **0.95** for normal cases and **0.90** for pneumonia cases, indicating a strong ability to correctly identify true positives while minimizing false positives. In a clinical context, minimizing false negatives was prioritized to ensure timely treatment of pneumonia cases, while the controlled rate of false positives was considered acceptable given the potential for additional diagnostic evaluations. The recall rates were **0.83** for normal cases and **0.97** for pneumonia, highlighting the model's effectiveness in capturing actual pneumonia cases. The F1-scores were **0.88** for normal and **0.94** for pneumonia, providing a balanced view of the model's performance in terms of both precision and recall. Table 2 shows the performance evaluation of simple CNN model.

From the confusion matrix, it is evident that the Simple CNN model made **11** false negatives and **40** false positives, which is acceptable given the high overall accuracy. Figure 2 shows further illustrates the confusion matrix of the performance of the Simple CNN model.

4.1.2 DenseNet121 model

DenseNet121 model achieved a slightly lower accuracy of **0.88**. The model's precision was **0.91** for normal cases and **0.87** for pneumonia, while the recall was **0.77** for normal cases and **0.96** for pneumonia. This model's recall for normal cases was somewhat lower, resulting in an increased number of false negatives. Table 3 shows the performance evaluation for DenseNet121.

The confusion matrix for DenseNet121 reveals **17** false negatives and **54** false positives, indicating a reasonable performance but with more errors compared to the Simple CNN. The Fig. 3 shows the confusion matrix for DenseNet121.

Table 3 Performance evaluation of DenseNet121

Class	Precision	Recall	F1-score	support
Normal	0.91	0.77	0.84	234
Pneumonia	0.87	0.96	0.91	390
Accuracy			0.89	624
Macro avg	0.89	0.86	0.87	624
Weighted avg	0.89	0.89	0.88	624

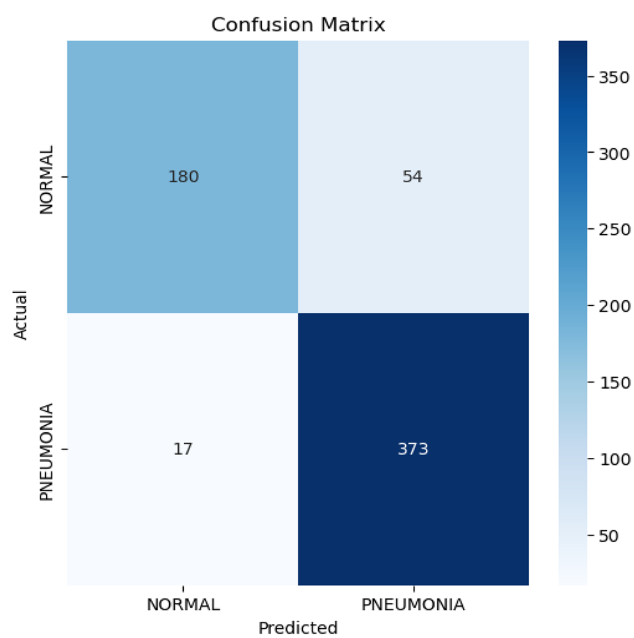
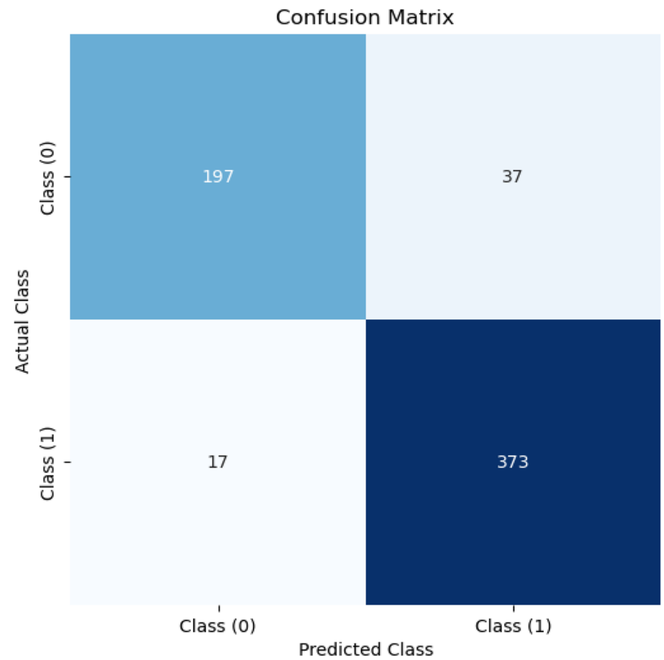


Fig. 3 Confusion Matrix of DenseNet121

Table 4 Performance evaluation of VGG16

Class	Precision	Recall	F1-score	support
Normal	0.92	0.84	0.88	234
Pneumonia	0.91	0.96	0.93	390
Accuracy			0.91	624
Macro avg	0.92	0.90	0.91	624
Weighted avg	0.91	0.91	0.91	624

Fig. 4 Confusion Matrix for VGG16



4.1.3 VGG16 model

VGG16 model performed comparably to the Simple CNN, with an accuracy of **0.91**. The model achieved a precision of **0.92** for normal cases and **0.91** for pneumonia, along with a recall of **0.84** for normal and **0.96** for pneumonia. The F1-scores were **0.88** for normal and **0.93** for pneumonia. Table 4 shows the performance evaluation of VGG16.

The model produced **17** false negatives and **37** false positives, showing strong performance in both precision and recall, making it a reliable model for pneumonia detection. Figure 4 below shows the confusion matrix for VGG16.

4.1.4 InceptionV3 model

InceptionV3 model showed the lowest accuracy of **0.84** among the four models. Its precision for normal cases was **0.92**, but the recall was significantly lower at **0.65**, indicating a high rate of false negatives. For pneumonia cases, the precision was **0.82** and recall was **0.97**. The F1-scores were **0.76** for normal and **0.89** for pneumonia. Table 5 shows the InceptionV3 model performance evaluation.

InceptionV3 had **13** false negatives and **82** false positives, which is less desirable compared to the other models. The Fig. 5 below shows the confusion matrix for InceptionV3.

4.1.5 Comparison of model performance

When comparing the performance of the four models, Simple CNN and VGG16 emerged as the best-performing models with accuracies of 0.92 and 0.91, respectively. Both models demonstrated high precision and recall values, particularly for pneumonia cases, making them highly reliable for detecting pneumonia in chest X-rays. DenseNet121 also showed

Table 5 Performance evaluation of InceptionV3

Class	Precision	Recall	F1-score	support
Normal	0.92	0.65	0.76	234
Pneumonia	0.82	0.97	0.89	390
Accuracy			0.85	624
Macro avg	0.87	0.81	0.83	624
Weighted avg	0.86	0.85	0.84	624

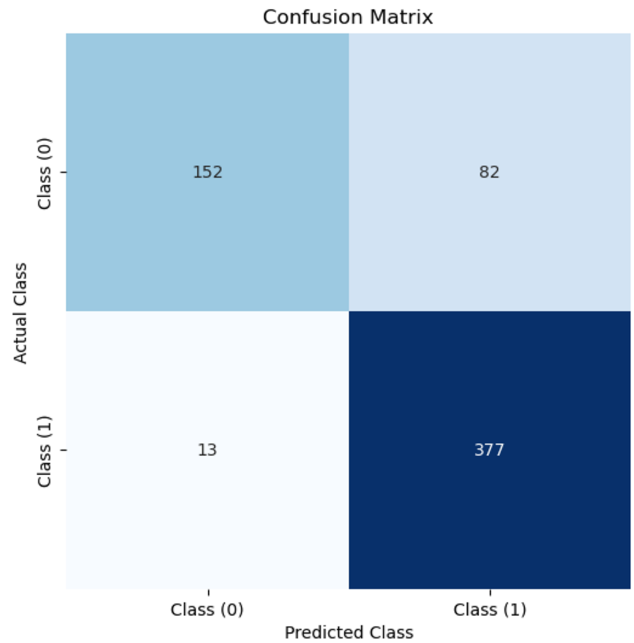


Fig. 5 Confusion Matrix for InceptionV3

good performance but was slightly less effective than Simple CNN and VGG16. InceptionV3, on the other hand, had the lowest accuracy and higher false positive rates, making it less suitable for the application.

Our findings demonstrate that the Simple CNN model outperforms more complex architectures such as DenseNet121 and InceptionV3 in terms of accuracy, precision, and computational efficiency. This novel focus on lightweight models highlights their potential for real-world deployment, making this study distinct from previous work that prioritizes model complexity over usability in constrained settings.

Given the high performance of Simple CNN and VGG16, especially in detecting pneumonia cases, Simple CNN was chosen for deployment in the web application developed for Dr. Sumait Hospital of SIMAD University. The decision was based on its slightly higher accuracy and balanced precision-recall metrics, making it the most robust model for this clinical setting.

The proposed Simple CNN model achieves an accuracy of 92%, surpassing comparable studies in the domain, which often prioritize complex architectures at the cost of computational efficiency. Unlike DenseNet121 and InceptionV3, which require substantial hardware resources, our model maintains high performance with reduced processing time and memory usage. This efficiency, combined with its integration into a user-friendly web interface, ensures practical deployment in clinical settings, especially in under-resourced regions.

4.1.6 Web platform results

The web application, developed using Django and hosted on AWS, integrates the Simple CNN model to assist healthcare professionals at Dr. Sumait Hospital in making more accurate decisions and reducing human error in pneumonia detection. The application allows for easy upload and analysis of chest X-ray images, providing immediate feedback on whether a patient shows signs of pneumonia. The integration of the Simple CNN model into this web platform ensures high reliability and accuracy, contributing significantly to better healthcare outcomes. The following Fig. 6 shows the Web Application Interface

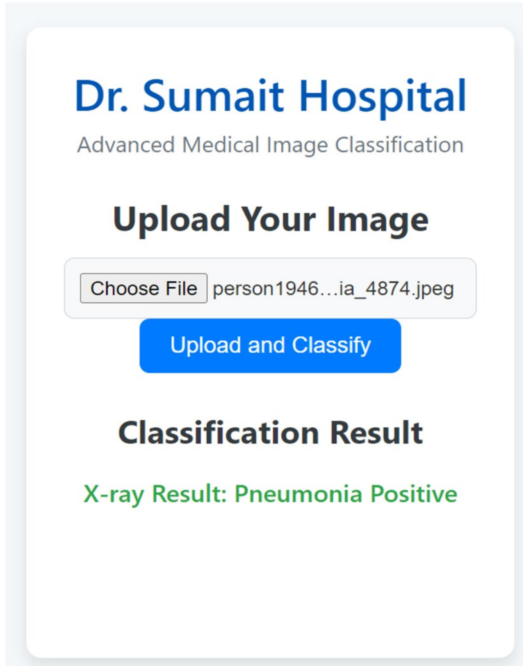


Fig. 6 Web Application Interface

In conclusion, the Simple CNN model's balanced performance in terms of precision, recall, and F1-score, combined with its higher accuracy, makes it the most suitable choice for deployment in a clinical setting to support pneumonia diagnosis.

4.2 Discussion

The performance of the four deep learning models—Simple CNN, DenseNet121, VGG16, and InceptionV3—indicates varying levels of success in accurately detecting pneumonia from chest X-ray images. Among these models, Simple CNN and VGG16 demonstrated the highest accuracy, while DenseNet121 and InceptionV3 lagged behind in certain key metrics.

The Simple CNN model performed exceptionally well with an accuracy of 92%, making it the most robust of the four models. The model's high precision (0.95 for normal and 0.90 for pneumonia) and recall (0.83 for normal and 0.97 for pneumonia) underscore its effectiveness in detecting pneumonia cases while maintaining a low rate of false positives. The F1-scores for both classes were high, with 0.88 for normal and 0.94 for pneumonia, indicating a balanced model that effectively handles the trade-offs between precision and recall. The relatively low number of false positives and negatives further attests to its reliability, which is critical in clinical settings where accurate diagnosis is essential.

The DenseNet121 model, while achieving a reasonably high accuracy of 89%, struggled particularly with normal cases, where the recall dropped to 0.77. This indicates that the model was more prone to misclassifying normal X-rays as pneumonia, resulting in a higher rate of false negatives. Although the precision and recall for pneumonia cases were commendable, the model's overall performance was slightly inferior to that of Simple CNN and VGG16, making it less suitable for deployment in high-stakes clinical environments.

The VGG16 model performed comparably to Simple CNN, achieving an accuracy of 91%. Its precision and recall rates, especially for pneumonia (0.91 precision and 0.96 recall), were similar to those of Simple CNN, making it a reliable option for pneumonia detection. However, the slightly higher number of false negatives (17) and false positives (37) suggests that Simple CNN may have a slight edge over VGG16 in practical applications, particularly in environments where a low error rate is critical.

The InceptionV3 model, with an accuracy of 85%, was the least effective among the models tested. The significant disparity between precision (0.92) and recall (0.65) for normal cases suggests that the model had difficulty distinguishing normal cases from pneumonia, leading to a high number of false positives (82). This imbalance, coupled with its lower F1-scores, renders InceptionV3 less reliable for accurate pneumonia detection compared to the other models.

When comparing all four models, Simple CNN emerged as the best-performing model due to its superior accuracy, precision, and recall, particularly for pneumonia cases. The model's relatively low false positive and false negative rates make it highly reliable in a clinical setting, where minimizing diagnostic errors is of paramount importance. VGG16 closely followed Simple CNN in terms of performance, making it another strong candidate for pneumonia detection, although slightly less balanced than Simple CNN.

The decision to deploy Simple CNN in the web application developed for Dr. Sumait Hospital was based on its overall performance and robustness. Its higher accuracy and balanced metrics made it the most suitable model for clinical use, particularly in detecting pneumonia with high precision. The integration of this model into the hospital's web platform offers a powerful tool to assist healthcare professionals, providing reliable and immediate diagnostic results, ultimately improving patient care and outcomes.

The developed system significantly reduces the time required for pneumonia detection compared to traditional methods. With the support of AWS cloud computing and the reliable internet connectivity at Dr. Sumait Hospital, the system provides rapid inference, with processing times much shorter than those needed by radiologists to manually interpret X-rays. This efficiency enhances the radiologist's workflow by allowing quicker diagnoses, ultimately supporting faster decision-making and improving patient care.

The clinical deployment of the proposed pneumonia detection system involves addressing computational requirements and implementation challenges. The system is hosted on the AWS Cloud platform, which provides scalable computing resources capable of handling high-resolution chest X-ray images efficiently. The cloud deployment ensures accessibility, reliability, and the ability to process images in near real-time, which is essential for clinical settings. One advantage of AWS is its flexibility to scale resources as needed, enabling efficient handling of peak workloads without requiring significant upfront investment in local hardware. However, maintaining optimal performance while minimizing latency may necessitate additional optimization of the Simple CNN model.

The key contribution of this study lies in its focus on practical deployment in a real-world healthcare setting. By selecting a computationally efficient model and integrating it into a scalable web platform, this research demonstrates how AI can be effectively implemented in resource-limited environments. The study also addresses critical challenges such as class imbalance and variability in image quality, ensuring the reliability of the proposed system.

The novelty of this study lies in its focus on practical implementation and optimization for resource-constrained environments. By selecting a lightweight model and deploying it as a web application, the study prioritizes scalability, computational efficiency, and accessibility, which are critical for under-resourced healthcare systems. This pragmatic approach differentiates the study from others in the field, which often emphasize performance metrics without considering the feasibility of real-world deployment. The emphasis on addressing class imbalance and ensuring generalizability further underscores the unique contributions of this work.

Despite the promising results, this study has several limitations that should be addressed in future research. First, the dataset used for training and testing the models was limited in diversity, comprising images primarily from a single source. This could restrict the generalizability of the model when applied to datasets from other populations, imaging devices, or clinical environments. Second, the study did not compare the model's performance against human radiologists, which could have provided valuable insights into its clinical relevance and potential as a diagnostic aid. Third, the computational requirements of the Simple CNN model, though relatively modest, have not been fully optimized for real-time processing in emergency settings, which may limit its immediate applicability in high-pressure environments. Lastly, while the model demonstrated high accuracy in detecting pneumonia, it is currently limited to binary classification and cannot identify other pulmonary conditions that may present with similar symptoms. These limitations highlight areas for improvement and further research to enhance the system's robustness and clinical utility.

4.3 Ethical considerations

The implementation of the developed system in medical diagnosis necessitates addressing key ethical considerations to ensure its responsible and effective use. As the system is deployed on the AWS Cloud platform, safeguarding patient data privacy and security is paramount. Compliance with healthcare data protection regulations, such as HIPAA, and the integration of robust encryption mechanisms are essential to protect sensitive information and maintain patient trust.

The system is intended to assist clinicians rather than replace them, ensuring that diagnostic decisions remain under human oversight. This approach minimizes the risk of overreliance on AI predictions and reinforces the system's role as a supplementary tool to enhance clinical decision-making.

Additionally, it is crucial to address potential biases within the training dataset to prevent disparities in diagnostic accuracy across diverse patient demographics. Expanding and diversifying the dataset during future development phases will improve fairness and generalizability, ensuring equitable outcomes for all populations.

Finally, transparency in the system's decision-making process is essential to foster trust among healthcare providers and patients. Providing interpretability features that explain the rationale behind the model's predictions will help clinicians integrate the system confidently into their workflows, facilitating informed and ethical usage.

5 Conclusion

In this study, we evaluated the performance of four deep learning models—Simple CNN, DenseNet121, VGG16, and InceptionV3—on detecting pneumonia from chest X-ray images. The Simple CNN and VGG16 models outperformed DenseNet121 and InceptionV3 in terms of accuracy, precision, recall, and F1-score, with Simple CNN showing the highest overall accuracy of 92%.

Simple CNN demonstrated a balanced and robust performance, making it the most suitable model for clinical deployment at Dr. Sumait Hospital. Its high precision and recall for pneumonia cases, along with low false positive and negative rates, make it a reliable tool for pneumonia detection, minimizing the risk of misdiagnosis. The integration of this model into a web-based platform provides healthcare professionals with a practical and efficient means of diagnosing pneumonia, ultimately contributing to improved patient outcomes and more accurate clinical decision-making. Post-deployment feedback collection from end users will be essential to validate the system's usability and effectiveness in a clinical setting. The web further enhances its practicality, ensuring real-time, accurate predictions in clinical settings. This combination of lightweight model architecture and user-centric platform design distinguishes our study from existing works, offering a scalable and efficient solution for pneumonia detection in resource-constrained environments.

The key contribution of this study is the development of a computationally efficient and high-performing Simple CNN model, which bridges the gap between state-of-the-art deep learning methods and practical usability in clinical settings. This approach provides a novel solution for resource-constrained environments, enabling effective pneumonia detection with minimal technological requirements. Although this study focused on evaluating the model's technical performance, a crucial next step involves comparing its diagnostic accuracy with that of experienced human radiologists. This will help validate the model's clinical reliability and its potential for real-world application in healthcare settings. This research highlights the importance of balancing technical advancements with practical deployment considerations, providing a framework for implementing AI solutions in under-resourced healthcare systems worldwide.

Future work could focus on addressing several limitations of this study to further enhance the performance and applicability of deep learning models in medical diagnostics. Expanding the dataset to include more diverse and representative samples from different populations, hospitals, and imaging modalities would improve the models' generalizability. Additionally, incorporating multi-class classification could enable the models to detect a broader range of lung conditions beyond pneumonia, enhancing their clinical utility. Another area for improvement is optimizing the models for real-time processing, ensuring they can provide fast and accurate results in emergency and high-pressure clinical environments. Integrating additional clinical data such as patient demographics, medical history, and symptoms could also significantly improve diagnostic accuracy, allowing the models to make more informed decisions. Future work will also focus on developing a custom deep learning model tailored specifically for pneumonia detection in resource-constrained settings. This model will be designed to optimize both diagnostic performance and computational efficiency. The custom model will then be rigorously compared to the existing architectures evaluated in this study, providing further insights into its advantages and limitations in real-world applications. Additionally, future work will involve conducting structured user testing and feedback collection from healthcare professionals at Dr. Sumait Hospital. This will include qualitative methods, such as interviews and surveys, to assess the system's usability, and quantitative measures, such as time savings in diagnostic workflows and diagnostic error rates. Insights from this evaluation will guide further improvements to the web application and enhance its alignment with clinical needs. Finally, future research could explore the use of more advanced model architectures or hybrid approaches combining deep learning with other AI techniques to further improve performance in detecting complex medical conditions.

To further build on this future work, additional research could be directed toward improving the interpretability of the deep learning models. As these models become more integrated into clinical settings, it is crucial for healthcare professionals to understand the reasoning behind the predictions made by the system. Developing explainable AI (XAI) techniques that provide insights into how the models arrive at their diagnoses could increase trust and facilitate clinical adoption. Furthermore, exploring methods to reduce the computational requirements of the models while maintaining high accuracy would make the system more scalable and feasible for deployment in resource-limited settings. In addition, collaboration with healthcare providers and medical professionals to continuously evaluate the system's performance in real-world clinical environments will ensure that the system remains relevant and effective in improving diagnostic workflows and patient outcomes.

Acknowledgements The authors would like to thank and extend their heartfelt gratitude to SIMAD University for their continuous support throughout this research project. Their provision of resources and encouragement has been instrumental in the successful completion of this work.

Author contributions A.Y.N and A.D.J analyzed and compared the four models and developed the Web app. M.O.A, Y.A.A and M.A.A, wrote the Introduction, Literature review and prepared the figures.

Funding The authors declare that no funding was received for conducting this study.

Data availability The "Chest X-Ray Images (Pneumonia)" dataset used in this study is publicly available and can be accessed through the Kaggle platform. The dataset contains 5,863 X-ray images, which are classified into two categories: pneumonia and normal. It is available at the following link: (https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia). Any further inquiries regarding the data can be directed to the dataset provider on the Kaggle platform.

Declarations

Ethics approval and consent to participate This study was conducted in accordance with the ethical guidelines and principles outlined in the Declaration of Helsinki and applicable institutional ethical standards. Since this research does not involve human participants, patient data, or animal subjects, formal ethical approval was not required. The dataset used (Chest X-ray Pneumonia Dataset from Kaggle) is publicly available, anonymized, and adheres to ethical data usage policies. No additional approvals were necessary. This study does not involve human participants.

Consent for publication Not applicable. This study does not include individual participant data.

Human ethics and consent to participate declarations Not applicable. This study does not involve human participants.

Competing interests The authors declare no competing interests.

References

1. P. Akhtar, I. Ahmad, A. Jameela, S. Nazli, K. Ahmed, and A. Professor, "Issue 11 www.jetir.org (ISSN-2349-5162)," 2020. [Online]. www.jetir.org
2. Sherkar MR, Menon MD, Musmade DS, Unde AS, Ppeumooia a dreadful disease, 2011.
3. Scott JAG, Brooks WA, Peiris JSM, Holtzman D, Mulholland EK. Pneumonia research to reduce childhood mortality in the developing world. J Clin Invest. 2008. https://doi.org/10.1172/JCI33947.
4. Mohan Gupta B, Gupta R, Mohan B, Pneumonia Disease Research in India: A Scientometric Study during 2004–13, Library Philosophy and Practice (e-journal), p. 1154, 2014, [Online]. http://digitalcommons.unl.edu/libphilprac
5. Čepin P, "Karakteristike upale pluća u dječjoj dobi," *University of Zagreb. School of Medicine. Department of Pediatrics, 2014.*, 2014. [Online]. https://urn.nsk.hr/urn:nbn:hr:105:257975
6. Abrahamian FM, et al. Health care–associated pneumonia: identification and initial management in the ED. Am J Emerg Med. 2008;26(6):1–11.
7. Adawe MO, Odongo AO, Kariuki JG. Risk factors associated with pneumonia among under 5 years children at Banadir Hospital, Mogadishu, Somalia. Asian J Med Health. 2023;21(8):1–11. https://doi.org/10.9734/ajmah/2023/v21i8832.
8. Adan FN, Jeele MOO, Omar NMS. Epidemiology of multidrug resistant non-fermentative gram negative bacilli in patients with hospital acquired pneumonia: an alarming report from Somalia. Infect Drug Resist. 2022;15:6297–305. https://doi.org/10.2147/IDR.S387370.
9. Tshimula JM, et al. Artificial Intelligence for Public Health Surveillance in Africa: Applications and Opportunities, *arXiv preprint* arXiv:2408.02575, 2024, [Online]. http://arxiv.org/abs/2408.02575

10. Mahanty C, et al. Effective Alzheimer's disease detection using enhanced Xception blending with snapshot ensemble. Sci Rep. 2024. https://doi.org/10.1038/s41598-024-80548-2.
11. Becker J, et al. Artificial intelligence-based detection of pneumonia in chest radiographs. Diagnostics. 2022. https://doi.org/10.3390/diagnostics12061465.
12. Ippolito D, et al. Artificial intelligence applied to chest X-ray: a reliable tool to assess the differential diagnosis of lung pneumonia in the emergency department. Diseases. 2023. https://doi.org/10.3390/diseases11040171.
13. Altameem A, Mahanty C, Poonia RC, Saudagar AKJ, Kumar R. Breast cancer detection in mammography images using deep convolutional neural networks and fuzzy ensemble modeling techniques. Diagnostics. 2022. https://doi.org/10.3390/diagnostics12081812.
14. Cooper GF, et al. An evaluation of machine-learning methods for predicting pneumonia mortality. Artif Intell Med. 1997. https://doi.org/10.1016/S0933-3657(96)00367-3.
15. Signé L. Strategies for effective health care for Africa in the Fourth Industrial Revolution Bridging the gap between the promise and delivery, Brookings 2021.
16. Mohammed NI, Jarde A, Mackenzie G, D'Alessandro U, Jeffries D. Deploying machine learning models using progressive web applications: implementation using a neural network prediction model for pneumonia related child mortality in The Gambia. Front Public Health. 2022. https://doi.org/10.3389/fpubh.2021.772620.
17. Hong JC, Eclov NCW, Stephens SJ, Mowery YM, Palta M. Implementation of machine learning in the clinic: challenges and lessons in prospective deployment from the System for High Intensity EvaLuation During Radiation Therapy (SHIELD-RT) randomized controlled study. BMC Bioinformatics. 2022. https://doi.org/10.1186/s12859-022-04940-3.
18. Gabruseva T, Poplavskiy D, Kalinin A, Deep Learning for Automatic Pneumonia Detection, In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops , 2020, pp. 305–306. [Online]. https://github.com/tatigabru/kaggle-rsna.
19. Kareem A, Liu H, Sant P. Review on pneumonia image detection: a machine learning approach. Human-Centric Intell Syst. 2022;2(1–2):31–43. https://doi.org/10.1007/s44230-022-00002-2.
20. Szepesi P, Szilágyi L. Detection of pneumonia using convolutional neural networks and deep learning. Biocybern Biomed Eng. 2022;42(3):1012–22. https://doi.org/10.1016/j.bbe.2022.08.001.
21. El Asnaoui K, Chawki Y, Idri A. Automated methods for detection and classification pneumonia based on X-ray images using deep learning. Cham: Springer International Publishing; 2021.
22. Sharma S, Guleria K, A deep learning based model for the detection of pneumonia from chest X-ray images using VGG-16 and neural networks, In: Procedia Computer Science, Elsevier B.V., 2022, pp. 357–366. https://doi.org/10.1016/j.procs.2023.01.018.
23. Rana N, Marwaha H. Pneumonia detection from X-ray images using federated learning–an unsupervised learning approach. Measurement Sensors. 2025. https://doi.org/10.1016/j.measen.2024.101410.
24. Kermany D, Zhang K, Goldbaum M. Labeled optical coherence tomography (oct) and chest x-ray images for classification. Mendeley Data. 2018. https://doi.org/10.17632/rscbjbr9sj.