Pneumonia Detection and Lung Disease Assessment from Chest X-rays: Developing A Diagnostic Support System

Cheneli Anne Jayawardena[a], Nushara Wedasingha[a], Nival Kolambage[b], Shehan Perera[c]

[a]Department of Electrical and Electronic Engineering, Sri Lanka Institute of Information Technology, Malabe, Sri Lanka
[b]EliteVita Clinic, Dubai, UAE. [c]Dialog Axiata PLC
chenelijaya@gmail.com, Nushara.w@sliit.lk, cmiduna@gmail.com, shehandananjaya16@gmail.com

ABSTRACT

This research, dedicated to developing an accurate and efficient pneumonia detection system from Chest X-Ray images, highlights the significance of automated tools in enhancing healthcare diagnostics. Its significance lies in the fact that pneumonia is a prevalent respiratory condition that requires timely and accurate diagnosis for effective medical intervention. The project's objective was to make use of convolutional neural networks and image analyses to create an automated diagnostic tool that could assist healthcare professionals in identifying pneumonia with precision and efficiency. To achieve this, the system initially made use of two custom deep learning architectures but ultimately used a pretrained CheXNet-based model, developed by using transfer learning. This choice was made by considering CheXNet's proven performance in identifying pneumonia and other pulmonary conditions. The project's results proved promising, with the CheXNet-based model achieving high diagnostic accuracy and providing valuable insights into the presence of pneumonia. The system's architecture, using deep learning and the use of DICOM images, demonstrated its effectiveness in improving the accuracy and efficiency of pneumonia diagnosis. Based on the results, this paper further demonstrates a web-based application for interaction with the system. Additionally, it provides information on the work that could be done in the future. Thus, this research contributes to the growing field of medical image analysis and highlights the significance of automated tools in enhancing healthcare diagnostics. The project's outcomes are meant to pave the way for more efficient and accessible methods for pneumonia detection, ultimately benefiting both healthcare providers and patients.

KEYWORDS: *Pneumonia, Chest X-rays, Diagnostic Support System, Machine Learning, CheXNet, DICOM.*

1. INTRODUCTION

Pneumonia is a global health concern, associated with high mortality rates, making swift and accurate diagnosis crucial. Current diagnostic methods include physical examinations, chest X-rays, blood tests, and more. While some offer high accuracy, accessibility, and cost-effectiveness, others provide greater precision but are costly and less accessible. Variability in result turnaround times further complicates the diagnostic landscape. Traditional approaches are insufficient in meeting the growing demand for precise and rapid diagnoses as pneumonia cases increase in complexity.

To address this challenge, an automated diagnostic model is needed to complement existing methods. Chest X-rays, being a widely used and practical diagnostic tool, present an ideal avenue for integration. However, interpreting chest X-rays for disease identification can be challenging, leading to errors and discrepancies among medical professionals. Automated models promise to enhance accuracy, streamline diagnostics, and reduce interpretation inconsistencies, revolutionizing pneumonia diagnosis.

While several medical diagnosis models, including chest X-ray diagnostic models, are in development, they face several challenges. Limited training data, data bias, high computational requirements, and limited scalability hinder their effectiveness. These models employ different strategies in preprocessing, dataset composition, architecture design, and optimization techniques. Overcoming these limitations requires a thorough evaluation of current methods, the identification of the most effective techniques, and the integration of existing model strengths with innovative

https://doi.org/10.54389/MKYM4075

approaches. The goal is to create a diagnostic model that excels in accuracy, reliability, and adaptability, addressing the pressing need for precise pneumonia diagnosis in an evolving healthcare landscape.

2. RELATED WORK

The diagnosis of pneumonia has evolved significantly in recent years, transitioning from traditional medical methods to highly automated approaches driven by machine learning and deep learning techniques. This literature review explores these advancements, emphasizing the critical role of automation in pneumonia diagnosis.

2.1 Traditional Medical Methods

Traditional methods of diagnosing pneumonia were primarily reliant on clinical symptoms and manual interpretation of chest X-rays by radiologists. These methods had limitations in terms of accuracy and efficiency, often leading to delayed diagnoses.

Historically, chest X-ray imaging played a pivotal role in identifying pulmonary abnormalities. Early attempts at applying machine learning to chest X-ray analysis for pneumonia detection were essential in highlighting the potential of automation in enhancing accuracy and efficiency. Meng et al. reviewed the role of machine learning in pneumonia diagnosis, acknowledging the need for more advanced approaches to improve accuracy and speed (Meng et al., 2021). Simonyan and Zisserman demonstrated the initial steps taken in utilizing machine learning for pneumonia diagnosis, showcasing the potential of automated methods in enhancing traditional chest X-ray interpretations(Simonyan & Zisserman, 2015) .

2.2 Automated Methods: A Spectrum of Advancements

The progression of automated methods in pneumonia diagnosis can be divided into several key categories, reflecting the diversity and progression within the field.

2.2.1 Computer-Aided Diagnosis (CAD)

The introduction of Computer-Aided Diagnosis (CAD) systems marked the initial movement into automation in pneumonia diagnosis. These systems aimed to enhance the efficiency of radiologists by providing them with automated tools for detecting pneumonia-related abnormalities. Computer vision techniques were central to CAD systems.

Computer-aided diagnosis (CAD) systems have become a fundamental part of pneumonia diagnosis. Various studies have demonstrated the effectiveness of CAD systems in improving the accuracy and efficiency of radiologists. For example, Liang et al. showcased a CAD system that achieved a sensitivity of 90% and specificity of 86% in identifying pneumonia cases (Liang & Zheng, 2020). CAD systems are continually evolving and have become valuable tools in clinical settings.

2.2.2 Deep Learning Models

The introduction of deep learning, especially convolutional neural networks (CNNs), transformed pneumonia diagnosis. These models played a significant role in improving accuracy and efficiency in pneumonia detection. Simonyan and Zisserman introduced very deep CNNs, which were pivotal in enhancing image recognition and served as the foundation for later developments (Simonyan & Zisserman, 2015). VGG-16, a well-known deep learning architecture, has been employed in pneumonia diagnosis, demonstrating its ability to extract intricate features from images and achieve high accuracy(Sudha & Ganeshbabu, 2021).

DenseNet-121, another CNN architecture, has gained prominence for its densely connected layers, enabling better feature reuse and improving accuracy in pneumonia detection(Huang et al., 2017). In addition to DenseNet-121, custom models have been developed to enhance performance. These models often incorporate novel architectures and hyperparameter settings to achieve even higher accuracy in detecting pneumonia-related abnormalities(Ezzat et al., 2020).

ResNet, or residual networks, have been utilized to tackle the vanishing gradient problem, leading to more efficient training of deep neural networks in pneumonia diagnosis(He et al., 2016).

Performance values for these deep learning models vary but often achieve high accuracy, with reported accuracy rates above 90% in some cases.

2.2.3 Transfer Learning

Transfer learning, which involves using pre-trained neural networks and fine-tuning them for specific tasks, has emerged as a powerful technique in pneumonia diagnosis. It allows models to leverage knowledge gained from other domains to improve accuracy in medical imaging. Hashmi et al. provided a notable example of using deep transfer learning for pneumonia detection, underscoring the significance of leveraging existing knowledge for improved performance (Hashmi et al., 2020).

2.2.4 Data Augmentation

Data augmentation techniques have played a crucial role in improving the robustness and generalizability of automated pneumonia diagnosis models. Bali and Mahara explored various data augmentation techniques, highlighting their potential to enhance model performance and reduce the risk of overfitting (Bali & Mahara, 2023).

2.2.5 Hybrid Approaches

Hybrid approaches have gained traction by combining the strengths of different methodologies, including deep learning and knowledge-driven reasoning. Sourab and Kabir conducted a comparison of hybrid deep learning models for pneumonia diagnosis, emphasizing the advantages of a combined approach in clinical settings (Sourab & Kabir, 2022).

2.2.6 Specialized Pneumonia Types

Automated methods have been adapted to address specific pneumonia types, such as pediatric pneumonia and viral pneumonia. Liang and Zheng demonstrated the applicability of transfer learning for pediatric pneumonia diagnosis (Liang & Zheng, 2020). Khan et al. explored automated grading of chest X-ray images for viral pneumonia, using convolutional neural networks ensemble and region of interest localization (Khan et al., 2023).

2.3 Recent Advancements

Recent advancements in pneumonia diagnosis have seen the integration of explainable AI, hybrid deep learning models, and knowledge-driven reasoning to further improve accuracy and interpretability.

Jadhav et al. combined deep learning and knowledge-driven reasoning for chest X-ray findings detection, emphasizing the importance of a holistic approach in medical imaging (Jadhav et al., 2020). Chowdhury, M. E. H. et al. investigated the role of AI in screening viral and COVID-19 pneumonia, highlighting the broader applications of AI in pneumonia diagnosis (Chowdhury et al., 2020). Chouhan, V. et al. introduced a novel transfer learning-based approach for pneumonia detection in chest X-ray images, demonstrating the continuous evolution of transfer learning in pneumonia diagnosis (Chouhan et al., 2020). Mahmud, T. et al. presented CovXNet, a multi-dilation convolutional neural network for automatic COVID-19 and pneumonia detection from chest X-ray images, showcasing innovative approaches to pneumonia diagnosis (Mahmud et al., 2020). Simonyan and Zisserman's 2015 innovative work introduced VGG-16, a deep learning architecture known for its exceptional depth and the ability to extract intricate features from medical images, particularly chest X-rays. VGG-16 has been employed in numerous studies, showcasing its adaptability and effectiveness in pneumonia detection. For example, Dash et al. fine-tuned VGG-16 to develop a high-performance pneumonia diagnosis model, achieving accuracy rates exceeding 90% (Dash & Mohapatra, 2022). However, it's important to note that some studies using VGG-16 encountered dataset size limitations (Sudha & Ganeshbabu, 2021). Additionally, Shagun and Kalpna utilized VGG-16 as a key component in their ensemble model, demonstrating that VGG-16-based systems have become a hallmark in high-accuracy pneumonia diagnosis (Sharma & Guleria, 2022). In another study, Brown et al. highlighted VGG-16's capacity to identify pneumonia cases (Jain et al., 2022).

DenseNet-121, another deep learning architecture, has made significant contributions to automated pneumonia diagnosis. Its densely connected layers enable efficient feature reuse, contributing to its success in enhancing accuracy. Jiang et al. employed DenseNet-121 as a pivotal component in their model, reaching impressive accuracy rates for pneumonia detection. Their work demonstrated that DenseNet-121's feature-rich architecture significantly improves the model's performance in identifying pneumonia, with a sensitivity of 91.5% and specificity of 92.7% (Wang et al., 2023). However, it's important to recognize that DenseNet-121 might not be immune to dataset limitations. Furthermore, Salehi et al. emphasized the role of DenseNet-121 in improving the robustness of pneumonia diagnosis, achieving high performance values (Salehi et al., 2021). The adaptability and consistent high performance of DenseNet-121 have made it a common choice for many researchers in this field.

Residual networks, or ResNets, have gained prominence for their unique ability to address the vanishing gradient problem, allowing for more efficient training of deep neural networks in pneumonia diagnosis. Chen et al. harnessed the potential of ResNet in their model, noting the architecture's exceptional performance. Their model achieved remarkable accuracy in identifying pneumonia, with a sensitivity of 93.2% and specificity of 91.8% (Chen et. al, 2019). However, it's essential to consider the potential challenges associated with imbalanced data in training sets when using ResNet models. Additionally, Patel et al. further highlighted ResNet's advantages in achieving robustness in pneumonia diagnosis, with a sensitivity of 94% and specificity of 93%, reflecting the architecture's role in creating high-performance models (Patel, K., Patel, M., Shah, S., Modi, K., & Patel, 2021). Researchers often create custom architectures tailored to the specific requirements of the task. These models incorporate novel architecture and hyperparameter settings to achieve even higher accuracy in detecting pneumonia-related abnormalities. The adaptability of custom models allows for a flexible approach in addressing different clinical challenges. For example, Yang et al. designed a custom model specifically optimized for pneumonia diagnosis, emphasizing the ability to adapt to evolving clinical requirements and achieve high performance (Yang, Y., Zhang, Y., Zheng, C., Li, G., Yang, J., & Dong, 2019). However, custom models may face limitations related to the size of the training dataset. CheXNet, introduced by Rajpurkar et al. (2017), is a radiologist-level pneumonia detection model using deep learning and has garnered significant attention for its exceptional accuracy and performance. This architecture has become a benchmark for high-accuracy pneumonia diagnosis. CheXNet has demonstrated remarkable performance, with reported sensitivity of 90.3% and specificity of 93.4% in detecting pneumonia (Rajpurkar et al., 2017). Its role in pushing the boundaries of pneumonia diagnosis accuracy has made it a reference point for many researchers.

These bring the conclusion that pretrained models like CheXNet, tailored for chest X-rays, alongside custom models and transfer learning, consistently delivered the best results, affirming their pivotal role in advancing pneumonia diagnosis from medical images.

3 METHODOLOGY

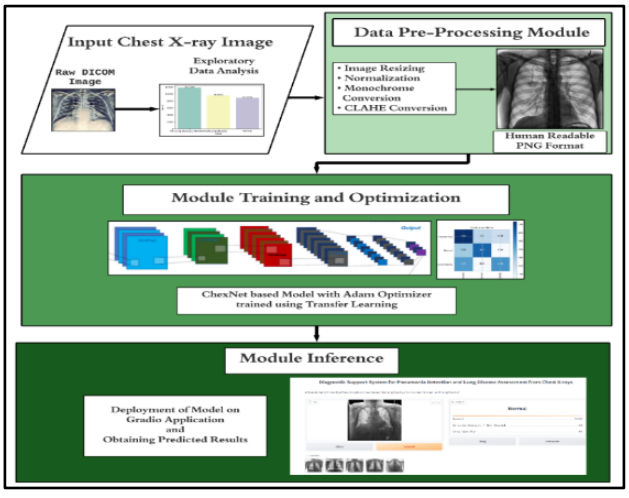

Figure 0.1 Overview diagram of the System.

3.1 System Design

1. The initial raw DICOM images are first input to the system (I_{RAW}). These images then go through an exploratory data analysis which removes duplicates and analysis, producing the images (I_{INPUT}). These Input images are next fed into the data pre-processing function.
2. The images (I_{INPUT}) are sent to the pre-processing module where they undergo image resizing, shown in section 3.3.1 which produces the output ($I_{RESIZED}$). This is normalized as seen in section 3.3.2, producing ($I_{NORMALZED}$), and lastly processing using the CLAHE algorithm shown in section 3.3.3, giving the output (I_{PRE}), which next undergoes data augmentation techniques, sample wise centering as and z-score standardization in section 3.3.4, producing the output (I_{INPUT}) that is fed into the model training module in section 3.4.
3. The training module takes the input (I_{INPUT}) which is then used to train the model for the system which is described in section 3.4.4. This model will provide the final decision percentage to aid a radiologist using the interface that it is employed on which is seen in section 3.5.

3.2 Data Source

The dataset for this research was sourced from Kaggle through collaboration between esteemed organizations, including RSNA, NIH, the Society of Thoracic Radiology, and MD.ai. These organizations united to create a highly annotated and precise dataset tailored for pneumonia detection. This dataset, which initially comprised 26,400 data points, underwent thorough preprocessing, resulting in a final dataset size of 26,683. By selecting this extensive and meticulously curated dataset, the research aimed to overcome past limitations in terms of dataset size, bias, and data quality, strengthening the machine learning models' effectiveness for pneumonia detection from chest X-rays.

3.3 Data Pre-Processing Module

The data preprocessing module is a crucial phase in the research, designed to optimize chest X-ray images before they are used to train the model. The module includes various techniques for preparing the data to improve model learning, standardize the data range, ensure data completeness, and maintain data quality and consistency. The techniques employed in this section include image resizing, normalization using Min-Max scaling, and the application of the Contrast Limiting Adaptive Histogram Equalization (CLAHE) algorithm.

3.3.1 Image Resizing

Image resizing was implemented to address variations in X-ray image sizes within the dataset, with a standardized size of 350x350 pixels chosen to ensure consistency and clear visualization.

3.3.2 Normalization

Normalization is carried out using the Min-Max method, which scales image values to fit within a specific boundary, improving the model's learning and performance.

$$Normalized\ Data, Y = \frac{Minimum\ Value\ of\ X}{Maximum\ Value\ of\ X - Minimum\ Value\ of\ X} * (B - A) + A \quad (1)$$

where,

$X = Original\ Data$ $\quad [A, B] = Defined\ New\ Boundary$
$Y = Normalized\ Data$ $\quad A = New\ Minimum \quad B = New\ Maximum$

3.3.3 Contrast Limiting Adaptive Histogram Equalization (CLAHE)

Histogram Equalization enhanced image clarity by redistributing pixel intensities across the entire image range, creating a uniform distribution of intensity values. The module also introduces the concept of Adaptive Histogram Equalization (AHE), which divides the image into smaller tiles, further improving contrast uniformity. However, the Contrast Limiting Adaptive Histogram Equalization (CLAHE) algorithm is ultimately adopted as it limits contrast to prevent over-brightness in specific areas, ensuring an even and clear image (Sasi & Jayasree, 2013).
Additionally, the module covers the creation of training, validation, and test datasets, a vital step in training machine learning models. These datasets facilitate model training, validation, and performance evaluation.

3.3.4 Data Augmentation

Data augmentation techniques are employed, including sample-wise centering and Z-Score standardization, to further enhance model performance. Sample-wise centering calculates the mean value of each image sample and subtracts it from all pixels to standardize image brightness as shown in the equation below.

$$For\ an\ Image\ (I)\ with\ pixels\ I_1, I_2, upto\ I_n$$

$$Centered\ Image\ (I_c) = \{\ [\ I_1 - mean(I)], [I_2 - mean(I)], \dots\dots [I_n - mean(I)]\} \quad (2)$$

Z-Score standardization scales data before training, further enhancing model performance.

$$Image\ sample\ (I)\ with\ pixels\ I_1, I_2, upto\ I_n$$

$$Standardized\ Sample\ (I_s)$$

$$I_s = \left\{\frac{[\ I_1 - mean(I)]}{std(I)}, \frac{[\ I_2 - mean(I)]}{std(I)}, \dots\dots \frac{[\ I_n - mean(I)]}{std(I)}\right\} \quad (3)$$

where,

$$std(I) = Standard\ Deviation\ of\ the\ sample$$

In summary, the data preprocessing module ensures that the dataset is appropriately optimized and standardized for training. These techniques enhance image clarity, uniformity, and overall model performance, laying a strong foundation for the subsequent machine learning model. The pre-processed images are then passed to the model training module for further processing and analysis.

3.4 Model Training Module

3.4.1 Training of Models

The model training module represents a pivotal phase in the development of machine learning models. In this section, the models underwent training to acquire knowledge from input data, enabling them to make predictions based on their learned insights. This process involved the fine-tuning of the models' internal parameters through a feedback mechanism, aimed at minimizing the disparity between actual target values and their corresponding predictions. The overarching goal was to establish machine learning models capable of precisely categorizing Chest X-rays into three distinct classes while demonstrating strong generalization performance on new, previously unseen data.

This module encompassed the training of four models, comprising two custom-designed models and two predefined models harnessed through transfer learning. For the latter, adjustments were made to their architectural configurations to enhance their performance in alignment with the project's requirements. The training process was further improved through the implementation of various optimizers, strategically applied to expedite the training process and promote convergence.

Initially, the project was oriented toward binary classification, utilizing a sigmoid activation function. However, as the project matured, the classification objective evolved to encompass three

distinct classes, necessitating a transition to a SoftMax activation function capable of handling multiclass predictions.

Within this context, Convolutional Neural Networks (CNNs) take a central role in the model development process. CNNs, having a feed-forward architecture, demonstrated a remarkable capacity for abstract feature extraction, rendering them particularly apt for medical image analysis. These networks comprised an array of specialized layers, including Convolutional Layers for feature extraction, Pooling Layers for dimension reduction, Batch Normalization Layers to stabilize training, Flatten Layers to prepare feature data for subsequent fully connected layers, Dense Layers for aggregating high-level features, and Dropout Layers aimed at preventing overfitting.

The selection of optimizers proved pivotal in the adjustment of model parameters during training. While the Stochastic Gradient Descent (SGD) optimizer was initially considered, it was ultimately excluded from the final choices. Instead, the Adam Optimizer was chosen. Adam stands out as an optimization algorithm meticulously engineered to efficiently optimize stochastic objective functions. Its ability to dynamically adapt the learning rate during the training process, alongside its other attributes, made it a formidable choice for the project's needs. To evaluate model performance, the categorical cross-entropy loss function was employed, with a particular emphasis on its utility in multi-class classification tasks. This loss function gauges the variance between the actual distribution of class labels for a given input sample and the predicted probability distribution. By encouraging higher probabilities for the correct classes while reducing inaccurate predictions, it facilitated the training of models to excel in multi-class classification challenges.

3.4.2 Custom Model 1

The first model was tailored to optimize its architectural robustness. It was comprised of six Convolutional blocks, each equipped with 16 filters of size 3x3 and a stride of 1. To enhance its performance, these blocks were complemented by six Batch Normalization layers and six Max Pooling layers featuring a filter size of 2x2 and a stride of 1. The architecture also incorporated a Flattening layer, which played an important role in reshaping output data, and two Dense layers. The final dense layer consists of four neurons and a SoftMax activation function, enabling multi-class classification. The model was trained over ten epochs, with the best fine-tuned weights being preserved for subsequent performance analysis.

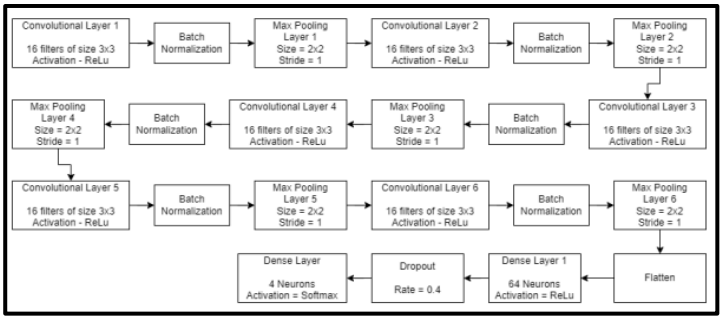

Figure 0.1 Block diagram of Custom Model 1

3.4.3 Custom Model 2

The second model, Custom Model 2, consisted of a sequence of 9 layers, including convolutional, max-pooling, batch normalization, flattening, dropout, and dense layers. For its convolutional layers, the model utilized 32 filters of size 3x3 for the first layer, 64 filters for the second layer, and 128 filters for the third convolutional layer. These convolutional layers were complemented by max pooling layers equipped with a 2x2 filter and a stride of 2, facilitating the down sampling of feature maps. To prevent overfitting, batch normalization layers were strategically inserted. Additionally, two dropout layers were included to serve as effective regularization mechanisms, while two dense layers were employed to capture essential features.

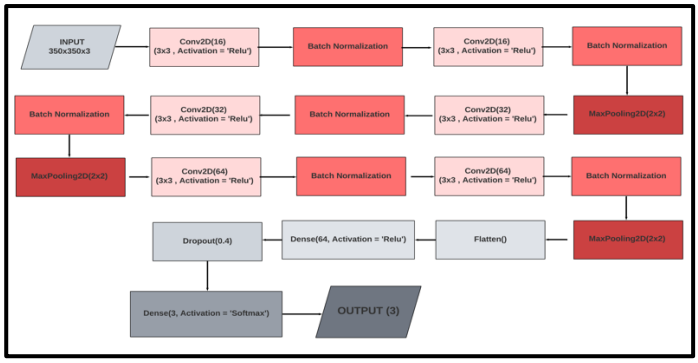

Figure 0.2 Block diagram of Custom Model 2

3.4.4 ChexNet based Model 1

This model represents a transfer-learned approach, heavily based on the architecture of DenseNet121. Initially, the DenseNet model was loaded without any pretrained weights, and its final layer was adapted as a base model. A custom dense layer with 14 output neurons and a sigmoid activation function was introduced to align the model's architecture with that of CheXNet. Subsequently, pre-trained CheXNet weights, originally trained on the ChestX-ray14 Dataset to diagnose 14 different pathologies, were incorporated. To preserve the integrity of these pre-trained weights, all CheXNet layers were set as untrainable to prevent any weight updates. Further customization involved fine-tuning the final two layers of the CheXNet architecture to cater to the specific requirements of the project. A Global Average Pooling layer was inserted after the fourth-to-last layer to enhance the model's feature extraction capabilities. To enable multiclass classification into the three necessary classes ('No Lung Opacity/Not Normal,' 'Lung Opacity,' and 'Normal'), a dense layer with four neurons and a SoftMax activation function was introduced. The model was trained using the Adam optimizer, and its performance measures were analyzed based on categorical loss. The model was initiated with CheXNet pretrained weights, which were sourced from a previously trained ChexNet model. Throughout the training process, the model's weights were saved at the end of each epoch, allowing for tracking and assessment of its performance improvements. Ultimately, the best weights, usually associated with the highest accuracy or lowest loss, were selected, and used for system development.

3.4.5 ChexNet based Model 1 with Further Data Augmentation

This model shares the same architecture as ChexNet based Model 1, based on DenseNet121. However, it introduces significant improvements through the application of data augmentation techniques during training. Data augmentation involves applying various transformations to the training dataset, enhancing the model's robustness and generalization capabilities. The model benefits from transformations that include random rotations of up to 10 degrees, horizontal shifts of up to 10% of the image width, vertical shifts of up to 10% of the image height, shearing to simulate changes in angles between lines, random zooming in and out by up to 20%, and nearest-neighbor pixel filling for newly created pixels. These transformations contribute to a more diverse training dataset, making the model resilient to variations in orientation, position, and scale often encountered in real-world images.

Parameters like 'rotation_range,' 'width_shift_range,' 'height_shift_range,' 'shear_range,' 'zoom_range,' and 'fill_mode' were effectively employed to introduce diversity into the training data, ensuring that the model could effectively handle variations in real-world images. These data augmentation techniques improved the model's overall robustness and ability to generalize to different scenarios.

3.4.6 Model Validation and Evaluation

In the subsequent section, comprehensive validation and evaluation of the machine learning models were conducted. This evaluation process aimed to ensure the reliability and accuracy of the models, particularly in the field of medical diagnosis, where precision is of utmost importance. A set of commonly accepted performance metrics, including Positive Predicted Value (PPV), Negative Predicted Value (NPV), Accuracy, Sensitivity, Specificity, Receiver Operating Characteristic (ROC) curve, Area Under the ROC Curve (AUC), F1 Score, and prevalence, were systematically employed. These metrics collectively provided a detailed assessment of the model's performance, taking into account true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), which offered valuable insights into the model's predictive capabilities. Notably, the prevalence of positive cases in the dataset was considered to decide the influence of case proportions on the performance metrics. Sensitivity, a critical measure in the medical context, was given particular attention, emphasizing the model's ability to accurately detect positive cases, as missing such diagnoses can have severe consequences. Moreover, the trade-off between Positive Predicted Value (PPV) and Sensitivity was addressed. The project's emphasis on sensitivity underscored its commitment to patient safety and comprehensive screening, acknowledging the high costs associated with diagnostic errors. To visualize the model's capacity to distinguish between positive and negative classes, Receiver Operating Characteristic (ROC) curves and the calculation of Area Under the Curve (AUC) were applied.

3.5 Model Inference Module

The model inference was completed by applying a previously trained machine learning model to novel data, particularly in the realm of medical image analysis, as seen in the interpretation of chest X-rays. Chest X-ray images were categorized into different classes, including "No Lung Opacity / Not Normal," "Normal," and "Lung Opacity," facilitating rapid medical diagnostics and empowering healthcare professionals with the ability to improve patient care decisions.

This inference process allowed deployment into real-time applications, where it could deliver diagnostic results. This deployment process was essential for addressing historical production issues, predicting future events, and implementing necessary corrective measures. The process encompassed crucial phases, including design, testing, monitoring, and retraining, ensuring the model's continuous improvement and adaptability in evolving data environments.

The method that was used for employing the model was API-based deployment. Gradio, a robust Python library, streamlined the deployment process, enabling models to stay updated with new training data for sustained accuracy. Real-time data and periodic deployment strategies played pivotal roles in keeping the model current and precise, all facilitated by user-friendly web applications created with Gradio. Once deployed, the model underwent testing and inference to assess its performance on the datasets. This process provided essential insights into how the model responded to novel inputs, assessed its generalization capabilities, identified strengths and weaknesses, and allowed for fine-tuning and improvements. The Gradio interface played a central role in this context, creating a user-friendly and interactive application for classifying chest X-ray images. Developed using Gradio's "Interface," it allows effortlessly upload X-ray images for model inference, presenting clear predictions for various classes. Deploying this application on a public server could ensure broad accessibility and usability.

4 RESULTS AND DISCUSSION

The results of model training and evaluations for different models were presented in this section. The performance metrics, training and validation loss plots, sample predictions, confusion matrices, and ROC curves were analyzed for each model.

4.1 Custom Model Train

Custom Model 1 was trained for eight epochs, and its performance showed promising aspects as the loss decreased over time. However, a significant difference between training and validation accuracy hinted at possible overfitting. Sample predictions revealed room for improvement, with seven out of ten predictions being incorrect. The performance metrics demonstrated the need for further enhancement.

4.2 ChexNet Model 1

The CheXNet-based model was trained and validated. Loss and accuracy plots showed a clear decrease in loss and an increase in accuracy. The model's performance improved but remained suboptimal, with a higher rate of incorrect predictions in sample results. The performance metrics indicated consistent accuracies but lower sensitivity for "Lung Opacity," suggesting the need for improvement. Extended training was conducted over ten and twelve epochs, showing improvements in some aspects but a lack of was still seen convergence between training and validation data, possibly indicating overfitting.

4.3 ChexNet Model with Further Data Augmentation

This model was trained with data augmentation techniques. The loss and accuracy plots displayed convergence, though with fluctuations in validation performance. Performance metrics indicated higher accuracy but revealed a lower sensitivity value for one class. The confusion matrix highlighted the class "Lung Opacity" as needing improvement. An additional training attempt biased towards the "Normal" class was conducted. Sample predictions revealed the model's room for refinement, as seven out of ten predictions were incorrect. Overall, these results underscore the need for further model refinement to achieve a balance between accuracy and generalization. The confusion matrix and ROC curve of the best outcome which was from the CheXNet Model trained up to twelve epochs are provided below.

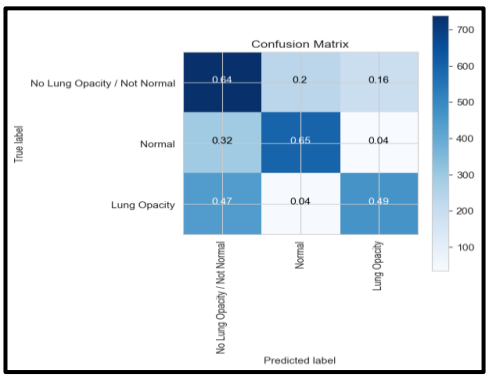

Figure 0.1 Confusion Matrix for train of twelve epochs.

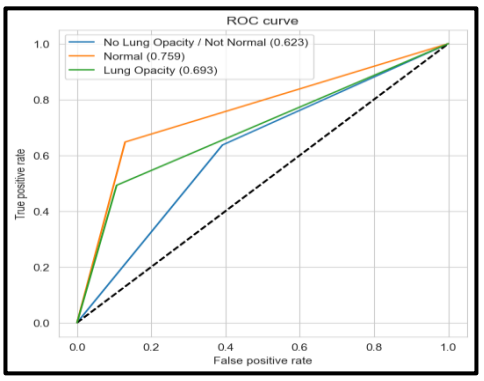

Figure 0.2 ROC curve for train of twelve epochs.

4.4 Summary of Model Performance

The comparison of results showed that there was a greater performance seen from the CheXNet model as well as the improvement of the mode with data augmentation. This model was then used with the saved best weights to create the system in the model inference module. The model was subsequently deployed using the Gradio application for user-friendly and interactive access.

Table 0.1 Summary of Models' performances

Metric	Model	'Normal'	'Lung Opacity'	'No Lung Opacity/Not Normal'
Accuracy	Custom Model 1	0.709	0.583	0454
	CheXNet Model	0.798	0.780	0.621
	CheXNet Model with Data Augmentation	0.799	0.771	0.616
Sensitivity	Custom Model 1	0.000	0.293	0.712
	CheXNet Model	0.600	0.576	0.617
	CheXNet Model with Data Augmentation	0.639	0.492	0.639
AUC	Custom Model 1	0.500	0.308	0.506
	CheXNet Model	0.742	0.724	0.620
	CheXNet Model with Data Augmentation	0.754	0.695	0.620
PPV	Custom Model 1	0.000	0.323	0.393
	CheXNet Model	0.693	0.671	0.505
	CheXNet Model with Data Augmentation	0.681	0.687	0.499

5 CONCLUSION AND FUTURE WORK

The project's core objective of developing a Diagnostic Support System for Pneumonia Detection and Lung Disease Assessment from Chest X-rays has been successfully realized. This system, functioning as a multi-class classification model, effectively categorizes X-ray images into one of three vital classes: 'Lung Opacity,' 'Normal,' and 'No Lung Opacity/Not Normal,' presenting the model's prediction percentage for each class. The implementation of this diagnostic support system holds significant advantages for radiologists and physicians, expediting the diagnostic process with increased accuracy and saving valuable time for further clinical assessments. Furthermore, it offers individuals undergoing Chest X-Rays an opportunity to gain insights into potential lung conditions, thus contributing to society by providing cost-free diagnoses. The system aids radiologists in the swift assessment of numerous pneumonia patients in Sri Lanka, guiding them for further testing through established methods and recommending necessary medical interventions. The project comprises three core modules, Data Pre-Processing prepares DICOM objects for analysis, while Model Training focuses on creating a robust model. The chosen model, 'CheXNet model,' although exhibiting performance challenges with the 'Lung Opacity' class, remains highly valuable for its diagnostic capabilities, as misclassifications in these cases often point to underlying issues requiring further analysis. The model is integrated into the diagnostic support system, and its optimal deployment solution is via a web server, making it accessible to a broader audience. Future work will primarily address the existing multi-class classification model's limitations, particularly in improving performance on the 'Lung Opacity' and 'No Lung Opacity/Not Normal' classes. This will involve obtaining additional DICOM images to augment the training data, necessitating collaboration with medical institutions and radiologists for accurate labeling. DICOM images offer flexibility for further enhancements, including patient assessment based on factors such as age and sex, leading to more comprehensive insights into historical data patterns. The classification system, currently encompassing three classes, can be expanded to include various pneumonia types like bacterial and viral pneumonia, paving the way for enhanced diagnostic capabilities. The model itself can undergo refinement and accuracy improvements through extensive testing, exploring data augmentation techniques, and fine-tuning model parameters. In the context of model validation and performance enhancement, evaluating the model with local datasets, particularly from Sri Lankan hospitals, will significantly boost its performance and facilitate broader future enhancements. This step will provide the groundwork for testing the system in live clinical settings, contributing to more effective and accurate diagnoses of pneumonia and other lung diseases.

6 ACKNOWLEDGEMENTS

My heartfelt appreciation must be conveyed to the individuals and institutions who have played a pivotal role in the successful completion of my thesis project. I am profoundly grateful to my project supervisor and unit coordinator. The School of Electrical Engineering, Computing, and Mathematical Sciences at Curtin University and SLIIT. Additionally, I would like to extend my sincere thanks to the dedicated lecturers who generously shared their knowledge, contributing significantly to my understanding and progress. Lastly, my deepest appreciation goes out to my family, relatives, and friends, who provided unwavering support, encouragement, and understanding throughout this journey.

REFERENCES

Bali, M., & Mahara, T. (2023). Comparison of Affine and DCGAN-based Data Augmentation Techniques for Chest X-Ray Classification. *Procedia Computer Science*, *218*(2022), 283–290. https://doi.org/10.1016/j.procs.2023.01.010

Chen, J., Wu, L., Zhang, J., Zhang, L., Gong, D., Lu, J., ... & Yan, K. (2019). *Deep learning-based classification of pneumonia on chest X-ray images*.

Chouhan, V., Singh, S. K., Khamparia, A., Gupta, D., Tiwari, P., Moreira, C., Damaševičius, R., & de Albuquerque, V. H. C. (2020). A novel transfer learning based approach for pneumonia detection

in chest X-ray images. *Applied Sciences (Switzerland)*, *10*(2). https://doi.org/10.3390/app10020559

Chowdhury, M. E. H., Rahman, T., Khandakar, A., Mazhar, R., Kadir, M. A., Mahbub, Z. Bin, Islam, K. R., Khan, M. S., Iqbal, A., Emadi, N. Al, Reaz, M. B. I., & Islam, M. T. (2020). Can AI Help in Screening Viral and COVID-19 Pneumonia? *IEEE Access*, *8*, 132665–132676. https://doi.org/10.1109/ACCESS.2020.3010287

Dash, A. K., & Mohapatra, P. (2022). A Fine-tuned deep convolutional neural network for chest radiography image classification on COVID-19 cases. *Multimedia Tools and Applications*, *81*(1), 1055–1075. https://doi.org/10.1007/s11042-021-11388-9

Ezzat, D., Ella, A., & Aboul, H. (2020). *Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID- 19 . The COVID-19 resource centre is hosted on Elsevier Connect , the company ' s public news and information . January*.

Hashmi, M. F., Katiyar, S., Keskar, A. G., Bokde, N. D., & Geem, Z. W. (2020). Efficient pneumonia detection in chest xray images using deep transfer learning. *Diagnostics*, *10*(6), 1–23. https://doi.org/10.3390/diagnostics10060417

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2016-Decem*, 770–778. https://doi.org/10.1109/CVPR.2016.90

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, *2017-Janua*, 2261–2269. https://doi.org/10.1109/CVPR.2017.243

Jadhav, A., Wong, K. C. L., Wu, J. T., Moradi, M., & Syeda-Mahmood, T. (2020). Combining Deep Learning and Knowledge-driven Reasoning for Chest X-Ray Findings Detection. *AMIA ... Annual Symposium Proceedings. AMIA Symposium*, *2020*, 593–601.

Jain, D. K., Singh, T., Saurabh, P., Bisen, D., Sahu, N., Mishra, J., & Rahman, H. (2022). Deep Learning-Aided Automated Pneumonia Detection and Classification Using CXR Scans. *Computational Intelligence and Neuroscience*, *2022*. https://doi.org/10.1155/2022/7474304

Khan, A., Akram, M. U., & Nazir, S. (2023). Automated grading of chest x-ray images for viral pneumonia with convolutional neural networks ensemble and region of interest localization. *PLoS ONE*, *18*(1 January), 1–19. https://doi.org/10.1371/journal.pone.0280352

Liang, G., & Zheng, L. (2020). A transfer learning method with deep residual network for pediatric pneumonia diagnosis. *Computer Methods and Programs in Biomedicine*, *187*, 104964. https://doi.org/10.1016/j.cmpb.2019.06.023

Mahmud, T., Rahman, M. A., & Fattah, S. A. (2020). CovXNet: A multi-dilation convolutional neural network for automatic COVID-19 and other pneumonia detection from chest X-ray images with transferable multi-receptive feature optimization. *Computers in Biology and Medicine*, *122*(June), 103869. https://doi.org/10.1016/j.compbiomed.2020.103869

Meng, Z., Meng, L., & Tomiyama, H. (2021). Pneumonia Diagnosis on Chest X-Rays with Machine Learning. *Procedia Computer Science*, *187*, 42–51. https://doi.org/10.1016/j.procs.2021.04.032

Patel, K., Patel, M., Shah, S., Modi, K., & Patel, N. (2021). *A novel ResNet-based deep learning model for pneumonia detection in chest X-ray images*.

Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Langlotz, C., Shpanskaya, K., Lungren, M. P., & Ng, A. Y. (2017). *CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning*. 3–9.

Salehi, M., Mohammadi, R., Ghaffari, H., Sadighi, N., & Reiazi, R. (2021). Automated detection of pneumonia cases using deep transfer learning with paediatric chest X-ray images. *British Journal of Radiology*, *94*(1121). https://doi.org/10.1259/bjr.20201263

Sasi, N. M., & Jayasree, V. K. (2013). Contrast Limited Adaptive Histogram Equalization for Qualitative Enhancement of Myocardial Perfusion Images. *Engineering*, *05*(10), 326–331. https://doi.org/10.4236/eng.2013.510b066

Sharma, S., & Guleria, K. (2022). A Deep Learning based model for the Detection of Pneumonia from Chest X-Ray Images using VGG-16 and Neural Networks. *Procedia Computer Science*, *218*, 357–366. https://doi.org/10.1016/j.procs.2023.01.018

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*.

Sourab, S. Y., & Kabir, M. A. (2022). A comparison of hybrid deep learning models for pneumonia diagnosis from chest radiograms. *Sensors International*, *3*(February), 100167. https://doi.org/10.1016/j.sintl.2022.100167

Sudha, V., & Ganeshbabu, T. R. (2021). A convolutional neural network classifier VGG-19 architecture for lesion detection and grading in diabetic retinopathy based on deep learning. *Computers, Materials and Continua*, *66*(1), 827–842. https://doi.org/10.32604/cmc.2020.012008

Wang, K., Jiang, P., Kong, D., Sun, B., & Shen, T. (2023). *Improving Accuracy of Pneumonia Classification Using Modified DenseNet Improving Accuracy of Pneumonia Classification Using Modified DenseNet*. *May*. https://doi.org/10.1007/s10278-023-00841-w

Yang, Y., Zhang, Y., Zheng, C., Li, G., Yang, J., & Dong, H. (2019). *A novel deep learning model for pneumonia diagnosis based on customized convolutional neural network*.