American Journal of Theoretical and Applied Statistics
2023; 12(6): 161-173
http://www.sciencepublishinggroup.com/j/ajtas
doi: 10.11648/j.ajtas.20231206.12
ISSN: 2326-8999 (Print); ISSN: 2326-9006 (Online)

Early Diagnosis of Pneumonia from Chest X-Rays Using a Capsule Network Model: Enhancing Accuracy and Efficiency in Automated Image Classification

Mbae Karwitha Maureen*, Thomas Mageto, Anthony Wanjoya

Department of Statistics and Actuarial Science, Jomo Kenyatta University of Agriculture and Technology (JKUAT), Nairobi, Kenya

Email address:

maureenmbae20@gmail.com (Mbae Karwitha Maureen), tttmageto@gmail.com (Thomas Mageto), awanjoya@gmail.com (Anthony Wanjoya)

*Corresponding author

To cite this article:

Mbae Karwitha Maureen, Thomas Mageto, Anthony Wanjoya. Early Diagnosis of Pneumonia from Chest X-Rays Using a Capsule Network Model: Enhancing Accuracy and Efficiency in Automated Image Classification. *American Journal of Theoretical and Applied Statistics*. Vol. 12, No. 6, 2023, pp. 161-173. doi: 10.11648/j.ajtas.20231206.12

Received: October 2, 2023; **Accepted**: October 20, 2023; **Published**: November 9, 2023

Abstract: Pneumonia is a significant public health concern worldwide, causing substantial morbidity and mortality. Early, accurate diagnosis is vital in ensuring timely treatment and improving patient outcomes. Chest X-ray analysis is the standard procedure used most frequently to diagnose pneumonia, but the accurate and timely interpretation of these images can be complex and time consuming. This research aimed to develop a capsule network (CapsNet) model for image classification, based on the Capsule network model introduced by Sabour and his colleagues in 2017 enabling automated chest X-ray analysis for early detection of pneumonia. Pneumonia impacts diverse populations, with vulnerable groups such as the elderly, young children and immunocompromised individuals at heightened risk. Delayed or missed diagnoses can lead to severe complications and increased healthcare costs. The reliance on human expertise for chest X-ray interpretation introduces the potential or errors, therefore there is a dire need to develop automated and precise diagnostic models and tools which are crucial for facilitating timely interventions. In this study secondary data obtained from Mendeley data was comprehensively pre-processed thoroughly by applying image resizing, standardization and normalization for consistent image quality, followed by a gaussian blur for noise reduction, and histogram equalization for contrast enhancement. The enhanced dataset enabled the main features of the pneumonia-infected images to be captured effectively during model training. The dataset was split into sets for training, testing and validation in an 80%, 10% and 10% ratio. The training set was used to train the CapsNet model which demonstrated a commendable performance with a 96% accuracy, a precision of 96.97% and a recall of 97.42%. The Capsule Network model shows a significant promise as a tool for improving the efficiency and accuracy of pneumonia diagnosis, thus befitting patients and healthcare providers.

Keywords: Pneumonia, Capsnet, X-Ray, Occlusions, Topography, Deep Learning, Artificial Intelligence

1. Introduction

The World Health Organization (WHO) describes pneumonia as a severe respiratory infection that mainly affects the lungs. Pneumonia causes the alveoli in the lungs which fill up with air when someone breaths to become packed with pus and fluid, (WHO, 2022). Patients with pneumonia have severe limitations in their ability to take in oxygen and have difficulty breathing. Pneumonia is caused by infectious agents and the most common causes of pneumonia are viruses and bacteria which are mainly spread through the air we breathe. The Centers for Disease Control and Prevention (CDC) classify pneumonia as a lower respiratory tract infection with bronchitis and tuberculosis.

Lower respiratory tract diseases were listed as the fourth leading cause of mortality worldwide in the WHO 2020 fact sheet. Over 450 million individuals suffer from pneumonia yearly Ruuskanen et al. [23]. More than four million people die from pneumonia each year, which accounts for 7% of the world's population. According to WHO (2021), 740,180

children under the age of five succumbed in 2019 due to pneumonia related complications. The deaths accounted for 14% of total child mortality in children under the age of five and approximately 22% of all fatalities in children between the ages of one and five. Pneumonia is the most common infectious cause of death in children globally. These numbers show that pneumonia is a worldwide pandemic that requires response and that it impacts kids and families everywhere. WHO establishes that pneumonia is a major worldwide health concern since it causes loss of loads of lives each year.

The CDC established that lower respiratory diseases are the second leading root source of mortality in low-income countries, Kenya included. According to the Kenyan Ministry of Health (MoH), Kericho, Kisumu, and West Pokot counties share the most burden of pediatric pneumonia deaths. According to Opuba and Onyango [20], pneumonia is the second leading cause of mortality for children under five years in Kenya and is responsible for 16% of all child fatalities. According to Hashmi et al. [10], poor health-seeking by caregivers and limited access to healthcare are to blame for the burden of pneumonia. Pneumonia can be deterred, treated with inexpensive medicine, low tech care and prevented with basic treatments but this is only possible if the sickness is detected in a timely manner, Hashmi et al. [10].

1.1. Background

Pneumonia is a significant health threat, ranking as the second leading cause of death in Kenya, with alarming global statistics. Hospital Clínic de Barcelona's research highlights that pneumonia accounted for 2.5 million deaths worldwide in 2019, including 600,000 child fatalities under five years of age. UNICEF reports that pneumonia prerogatives more fledgling lives than any other communicable disease, causing over 700,000 child deaths annually, including newborns tragically, these deaths are mostly preventable.

Globally, pneumonia affects more under five children than malaria and diarrhea combined, with over 870,000 children succumbing yearly. This disease represents approximately 15% of all child deaths annually. Alarming statistics from 2019 reveal that 2.5 million people died of pneumonia, with nearly a third being children under five, Dadonaite et al. [5]. As a result, pneumonia is the main cause of death in this age bracket. Compared to prior years, the death rate for older individuals decreased slightly in 2017, although the number of deaths among those aged 70 and above rose.

American Thoracic Society states that pneumonia is the world's principal cause of mortality, attributing to 16% of fatalities globally. In 2015, 2400 pneumonia deaths were reported daily, and 120 million episodes were reported yearly, amongst which 10% progressed to severe. The United Nations Children's Fund (UNICEF) reported 880,000 deaths, among which most were infants below two years in 2016. In the United States, pneumonia is less fatal for children but still a problem in the national hospitals. As half of pneumonia deaths are between 18-64, young people are also affected by pneumonia. This is because pneumonia is the key root cause of sepsis and septic shock. UNICEF also stated that every 43 seconds, a child dies of pneumonia. The latest WHO data from 2020 show that 22,571 people died in Kenya from influenza and pneumonia, accounting for 8.57% of all fatalities. Kenya is the 38th leading country in the world, with an age-adjusted death rate of 85.68 per 100,000 people. Pneumonia caused 15% of child deaths in 2018 and was the leading cause of mortality for children under five in the former year. More than one child per hour, or about 9,000 children under five, died from pneumonia in 2018. Between 2000 and 2018, pneumonia mortality decreased at an average annual rate of 6%. Kenya is anticipated to meet the 2025 GAPPD target with a 4-year gap in 2029 if progress continues at the same rate.

Pneumonia patients, once infected, the presenting features differ from trifling to severe reliant on the kind of infection, age of the patient, and general health. In some instances, patients might be infected but present no features, especially children under five. WHO (2021) states that most pneumonia symptoms are mistaken for flu or common cold, and failure to diagnose the disease at its early stages has proven fatal. Medical practitioners have brought forward procedures for diagnosing pneumonia in patients presenting disease symptoms.

The primary emphasis of this study was the utilization of chest X-ray images as a reliable method for identifying pneumonia in patients. Metal appears fully white in an X-ray image, bone appears virtually white, fat, muscle, and fluids appear in various hues of grey, and air and gas seem completely black. To provide an appropriate diagnosis, medical professionals need to carefully examine the photos. In the diagnosis of pneumonia, the white sick tissues contrast sharply with the dark air in the lungs, blocking more of the X-ray.

The necessity for computerizing the diagnosis is important because there are some areas of the world with limited access to experienced medical personnel, radiologists, and imaging specialists, whose forecast of such diseases matters substantially. This study, therefore, developed a model that can perform binary classification to detect the presence and absence of pneumonia from patients' X-rays, especially in the early stages of infection. Early and accurate diagnosis is crucial for initiating appropriate treatment and reducing the morbidity and mortality associated with pneumonia.

Chest X-rays have long been a valuable tool in the diagnostic process, allowing healthcare professionals to visualize the lungs and identify characteristic signs of pneumonia, such as infiltrates and consolidations. However, accurate interpretation of chest X-rays for pneumonia diagnosis requires high level expertise and experience. Interpretation challenges arise due to the variability in image presentation, overlapping features with other respiratory conditions, and the subjective nature of visual interpretation. These factors can lead to diagnostic errors, delays in treatment initiation, and subsequent negative impacts on patient outcomes.

Artificial intelligence and deep learning models are becoming

quite popular in medicine. They are used as problem solvers in diagnoses to provide solutions beyond medical practitioners' knowledge at minimum costs. They act as supplements in making clinical decisions, saving cost and time Convolutional neural networks were suggested by Milletari et al. [18] as a way for representing the prostrate in MRI volumes. In 2017, Curigliano et al. [4] established a deep learning model for skin cancer classification at the level of a dermatologist, proving the ability of artificial intelligence to classify skin cancer at a close proficiency analogous to that of dermatologists. These models have also been used by researchers to identify disorders via X-rays. Two deep three-dimensional (3D) modified mixed link network (CMixNet) architectures were employed by Nasrullah et al. [19] for the identification and categorization of lung nodules.

Pneumonia has been detected and diagnosed using artificial intelligence models by Aledhari et al. [1], Yee and Raymond [26], and Erdem and AYDN [6]. These studies used the convolution neural network model for diagnosis. Hashmi et al. [10] used deep transfer learning with several models, and Ayan et al. [2] used an ensemble model, to mention a few to diagnose pneumonia from chest X-rays. Recent advancements in deep learning and neural network models have shown great potential in automating medical image analysis tasks. CapsNet, a novel neural network architecture proposed by Sabour et al. [24], offers a unique approach by introducing capsules, which capture spatial relationships and hierarchical representations within images.

Manual interpretation of chest X-rays for pneumonia diagnosis has inherent limitations. Radiologists may experience difficulties in distinguishing between pneumonia and other lung conditions with similar radiographic findings. Moreover, the process can be time consuming, leading to diagnosis and treatment initiation delays. The growing demand for accurate and efficient pneumonia diagnosis has driven researchers to explore automated image classification techniques, particularly leveraging the supremacy of deep learning.

Deep learning, a subsection of artificial intelligence, has revolutionized various fields, including computer vision and image examination. Automated image classification models built on deep learning techniques offer several potential advantages for pneumonia diagnosis. They can quickly process large volumes of data, detect subtle abnormalities, and provide consistent and objective assessments. Furthermore, these models can potentially improve the efficiency of pneumonia diagnosis, allowing healthcare professionals to prioritize and expedite treatment plans for affected individuals. By leveraging the power of deep learning and automated image classification, we can potentially enhance the accuracy, efficiency, and accessibility of pneumonia diagnosis from chest X-ray images. This can significantly impact patient outcomes, clinical decision-making, and public health strategies.

1.2. Literature Review

Deep learning has emerged as a powerful subset of machine learning that has revolutionized various fields, including computer vision and image analysis. It has shown remarkable success in solving complex problems by automatically learning hierarchical representations of data, particularly in the form of artificial neural networks. Deep learning models have become the go-to approach for image classification tasks; for instance, CNNs are designed to mimic the human visual system's hierarchical organization, allowing them to learn and extract meaningful features from images effectively. The essential advantage of deep learning lies in its ability to automatically learn intricate patterns and representations from large amounts of data. Unlike traditional machine learning methods that rely on manual feature engineering, deep learning models can autonomously study the topographies directly from the raw data. This makes them highly effective in capturing complex and abstract patterns that may be difficult to define explicitly.

Deep learning techniques have demonstrated great promise in the context of pneumonia detection from chest X-ray images. By training CNN models on large-scale datasets, such as the ChestX-ray14 dataset, these models can learn to distinguish regular chest X-rays from those exhibiting pneumonia-related abnormalities. They can automatically learn the subtle patterns and characteristics indicative of pneumonia, allowing for accurate classification. Image classification is a fundamental task within deep learning, where the goal is to assign a label or class to an input image based on its content. In the case of pneumonia diagnosis, image classification models can learn to differentiate between healthy and infected chest X-rays, enabling the detection and identification of pneumonia-related features.

Deep learning models, particularly CNNs, have performed remarkably in various image classification tasks, including pneumonia detection from chest X-ray images. They have accomplished high accuracy rates, often comparable to or surpassing human radiologists' performance. This suggests their potential as valuable tools in assisting healthcare professionals in diagnosing pneumonia and improving efficiency, accuracy, and patient outcomes. In recent years, many researchers have highlighted the importance of the prompt detection and diagnosis of pneumonia and other diseases using deep learning techniques. A deep-learning method for detecting diabetic retinopathy in retinal fundus pictures was developed and validated by Gulshan et al. [8]. Although their work was not specifically about diagnosing pneumonia, it did validate the efficiency of deep learning in this field. Convolutional neural networks were used by Lakhani and Sundaram [14] for the automated categorization of pulmonary TB on chest radiographs. Although their study focused on tuberculosis specifically, the use of deep learning in the diagnosis of pneumonia has similarities in terms of image processing and categorization.

To better detect pulmonary nodules on chest radiographs, Lee et al. [15] created a deep learning-based system, showcasing the potential of deep learning to increase the accuracy of pneumonia diagnosis. In addition, Chassagnon et al. [3] gave information on current developments in the automated detection of pulmonary nodules on chest CT scans,

which can guide the creation of sophisticated diagnostic tools for pneumonia. According to a report by Kareem et al. [11], multiple machine learning methods, such as convolution neural network (CNN), k-nearest neighbor (KNN), RESNET, CheXNet, and DECNET, can be utilized to diagnose pneumonia. They thoroughly examined the literature to determine how we may pool hospitals and medical facilities to train machine learning models from their datasets, allowing the ML algorithms to detect diseases more accurately and effectively.

Convolution Neural network, in particular, has been widely used in computer vision and problems involving vast complex data sets that do not promise any precise statistical distributions. Gupta [9] developed a CNN model for detecting pneumonia from chest X-ray images and accomplished a validation accuracy of 93% and a training accuracy of about 99%. GM et al. [7] performed a study on the detection of pneumonia using 15 different CNN models to find a model that is easy to train and computationally less expensive. They concluded that there is a need for extensive studies to obtain simple models that can maintain high-level accuracy and performance in diagnosing pneumonia. It is therefore evident that CNN models are limited because they require very high computation power and can take much time during training, hence time inefficient. CNN also requires large amounts of datasets to process and train the model.

Ensemble learning has also been investigated to enhance the robustness and accuracy of pneumonia diagnosis models. By combining predictions from multiple deep learning models, ensemble models can capture a diverse range of features and improve the overall classification performance. Ensemble methods have shown promising results, further increasing the accuracy and reliability of pneumonia diagnosis. Sirazitdinov et al. [25] suggested an ensemble model for pneumonia identification and localization using a significant chest X-ray database based on two CNN models, RetinaNet and mask R-CNN. They showed that focus loss and object detection were superior in terms of categorization metrics, although accuracy levels suggested that the study needed to be expanded.

Liz et al. [17], used ensembles of CNN models to diagnose pediatric pneumonia to maximize the Area Under the curve (AUC) and the true positivity rate TPR. Their study was mainly focused on addressing the challenges in CNN models with highly unexplainable outputs that make the classification process rigorous, as well as the fact that they require vast datasets for training to achieve a high degree of accuracy. The ensemble model developed attained robust outcomes compared to the Kermany et al. [13] study, which used deep transfer learning to distinguish viral from bacterial pneumonia from chest x-rays.

An ensemble of deep convolutional neural networks was utilized by Ayan et al. [2] to identify pediatric pneumonia in chest X-ray images. They used seven CNN models that had already been trained on the ImageNet dataset. The ensemble was trained on the chest X-ray dataset using the appropriate transfer learning and fine-tuning techniques. They were successful in correctly classifying normal, viral, and bacterial pneumonia in chest X-ray images with an accuracy of 90.71. Ayan et al. [2] assert that using transfer learning and fine-tuning to find the hyperparameters of pre-trained CNN models exposes a time-consuming trial-and-error process. The important complimentary traits of ensemble CNN models are high variance and low bias. Due to their potentially poor interpretability, need for big datasets, and high computing demands, ensemble learning techniques have some limitations. The models are extremely intricate and their creation necessitates a high level of expertise. They also add learning time and memory restrictions to the issue because they are computationally expensive.

Researchers have also explored transfer learning, where pre-trained CNN models on large-scale image datasets, such as ImageNet, are fine-tuned on pneumonia-specific datasets. Transfer learning allows models to leverage knowledge learned from vast amounts of data, improving generalization and performance on pneumonia detection tasks. Kermany et al. [13] used transfer learning to diagnose pediatric pneumonia. The model used was highly accurate in diagnosing pneumonia as well as distinguishing between Viral and Bacterial pneumonia but was limited in that training a highly accurate model with a small data set will provide inferior performance to that of a model trained from a random initialization on a vast dataset which would take weeks to achieve accuracy. This indicates that transfer learning requires vast datasets for training to achieve high accuracy, which will require more time.

Liang and Zheng [16] suggested using transfer learning to set the model weight parameters and a residual structure and dilated convolution to classify pediatric pneumonia images. The main aim of the study was to find a solution to the issues of poor image resolution and overlying in the inflamed area of the chest X-ray. The model attained a f1-score of 92.7% and a classification accuracy of 96.7%. Transfer learning techniques were utilized to pre-train the model in a recent study by Rehman et al. [22] to create an effective diagnosis of COVID-19 disease by distinguishing it from healthy cases, viral pneumonia, and bacterial pneumonia. Despite achieving accuracy in the results, a large amount of CT and X-ray datasets was required. Transfer learning also is limited in that it requires large data sets to pre-train the model to achieve perfect weights and accuracy of the model. Transfer learning is inefficient as a model can take weeks to learn images if the data is very large. Additionally, transfer learning architectures are broad and many, and choosing a perfect architecture is vital to obtaining a perfect diagnosis in case of diseases.

These advances in deep learning for pneumonia diagnosis are promising, but several challenges and considerations remain. The lack of time and computationally efficient models to be used even when resources are limited and for small datasets needs to be addressed. Additionally, the interpretability and explain ability of deep learning models in pneumonia diagnosis are areas that require further research to build trust and facilitate integration into clinical practice.

Future research should address the remaining challenges to fully yoke the latent of deep learning in pneumonia diagnosis and enhance patient care.

2. Methodology

2.1. Dataset

This study used secondary data from Mendeley data. The dataset was created and uploaded by Kermany et al. [12] in 2018. They created a dataset that included hundreds of validated OCT and chest X-ray pictures, which were offered and scrutinized in "Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning." The dataset is readily available for research purposes. Python programming was used to perform all these analysis procedures.

Python seamlessly integrates with popular machine learning libraries like Scikitlearn and TensorFlow, enabling researchers to build and evaluate complex models for classification, regression, and other tasks. This is particularly valuable in image classification, where machine learning algorithms are commonly employed. Python promotes reproducibility in research by enabling the creation of clean, well-documented code that can be easily shared and replicated. Its ability to efficiently handle large datasets makes it suitable for scalable data analysis tasks, ensuring researchers can effectively tackle substantial volumes of data and derive meaningful insights. To achieve the goal of this study, the methodology entailed data pre-processing and model training.

2.2. Data Pre-Processing

Data pre-processing is a significant step in machine learning models as it enhances the model's accuracy. Pre-processing of data cleans the data, fills up the missing values, removes outliers, and identifies specific values in the data. The images in the dataset were provided as grayscale PNG images, and the following pre-processing procedures were carried out.

Resizing an image involves changing its dimensions while maintaining its aspect ratio. It involves altering the number of pixels in an image to make it larger or smaller. The chest X-ray images were resized to a standardized resolution. This step ensured uniformity and consistency in the input size of the images, which is crucial for model training.

Image normalization is a technique used to standardize the pixel values of an image by bringing the pixel values within a specific range or distribution, making them more suitable for further analysis or processing tasks. Normalizing an image can improve the performance of machine learning models, enhance image quality, and ensure fair comparisons between different images. This step helps reduce the influence of variations in image brightness and contrast, making the data more suitable for training the model. Mini max normalization was applied to the resized images.

Noise reduction aims to remove and reduce unwanted noise from images. Several variables, including sensor limits, transmission flaws, and ambient influences, might cause noise. Removing noise helps improve image quality, enhance details, and facilitate more accurate analysis and feature extraction. Gaussian blur was applied, a widely used image-processing technique for reducing noise and smoothing images. It is based on the mathematical concept of convolving an image with a Gaussian kernel.

Contrast enhancement is a technique used to improve images' visual quality and distinguishability by increasing the contrast between different pixel intensities. It aims to enhance the differences in brightness or color between image regions, making them more prominent and improving the perception of details. Histogram equalization was applied to the images; it is a widely used image enhancement technique that helps improve an image's contrast and visibility. It redistributes the pixel intensities in an image to cover a broader range of values, enhancing its overall appearance. Histogram equalization was applied to the images to help improve the visibility of details to capture the presence of any pneumonia infiltrations.

The pre-processed images were then split into 80%, 10% and 10% proportions for training, testing and validation, respectively. Data partitioning allows for rigorous evaluation of models and helps ensure they are robust and reliable when deployed in practical applications. It enables the selection of appropriate models and hyperparameters, leading to improved performance and reliable predictions of unseen data.

2.3. The Capsule Network Model

The Capsule Network model, also known as CapsNet, is a type of neural network architecture introduced by Geoffrey Hinton and his colleagues in 2017. It was designed as an alternative to traditional convolutional neural networks (CNNs) and aimed to address some of their limitations, particularly in tasks involving object recognition and understanding spatial relationships. The Capsule Network model is based on the concept of "capsules," which are groups of neurons that encode the instantiation parameters of an entity and the presence of that entity in an image. These capsules capture the entity's presence and properties, making them more potent than individual neurons in CNNs. The proposed capsule model is given as follows:

The key idea behind CapsNet is introducing a dynamic routing mechanism called "routing-by-agreement." This mechanism allows capsules at one layer to communicate with capsules at the next layer, enabling them to reach a consensus on the existence and properties of entities in the input. This process considers the pact between lower-level and higher-level capsules to determine the weights of connections, which facilitates the hierarchical representation of complex spatial relationships. The CapsNet model has shown promising results in various tasks, including digit recognition, object detection, and image segmentation. It has demonstrated better resistance to occlusion and viewpoint changes than traditional CNNs. However, due to its relative

novelty, CapsNet is still an active research area, and further advancements and improvements are being explored.

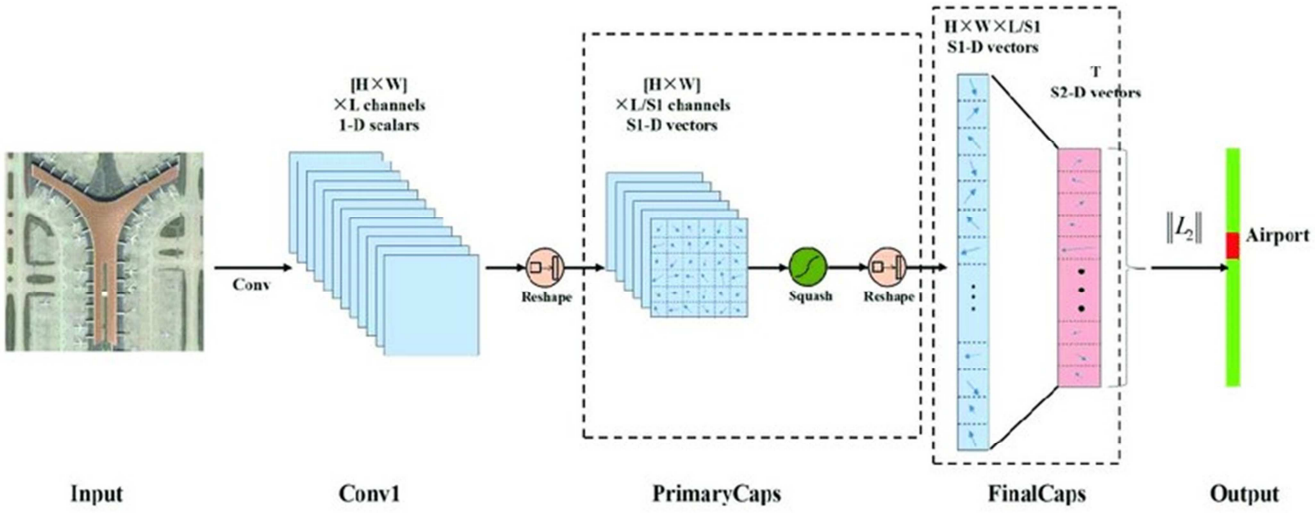

Figure 1. *The Capsule Network Model, Source (internet).*

2.3.1. The CapsNet Architecture

The CapsNet architecture consists of several layers of capsules that capture entities' hierarchical relationships and properties within input data. The input layer receives the raw data, usually images or feature vectors. The data is passed to the primary capsule layer: The primary capsule layer extracts low-level features on the input data and sends the data to the dense capsule layer made up of convolutional capsules that represent a higher-level feature and are connected to primary capsules through weighted connections. The primary capsules' outputs are transformed into vectors, representing the instantiation parameters, and are then weighted and combined to form the output of the convolutional capsules. The higher levels are passed through the routing mechanism. It involves iterative routing between lower-level and higher-level capsules to establish agreement on the existence and properties of entities. The routing process determines the weights of the connections between capsules based on the agreement, allowing for the representation of hierarchical relationships and spatial configurations. The extracted features are passed to the final layer that produces the model's classifications. It often consists of capsules that represent specific classes or categories.

In the context of the CapsNet architecture, let us now dive into the mathematical equations that govern its statistical framework. These equations represent the transformations and computations within the primary capsule layer, digit capsule layer, and routing algorithm. By understanding these equations, we can learn how the CapsNet architecture learns hierarchical representations of visual entities.

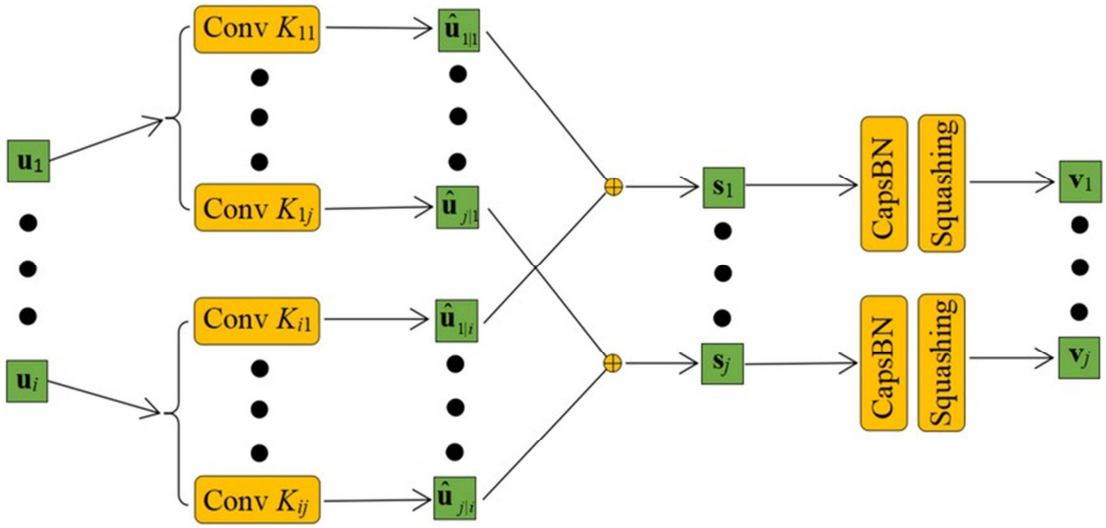

Figure 2. *The CapsNet model architecture, Source (Sabour et. al.).*

The input into CapsNet is the actual image supplied to the neural net. The input image is three-dimensional, with pixel values representing the image height and image width, and the image color channels. The chest X-ray images were

Grayscale with 224 by 224 pixels. This was captured in the input and stored in a 224x224x1 matrix [224, 224,1]. Let X represent the input image tensor received by the input layer and fed into the primary capsule layer. The primary capsule layer aims to generate primary capsules representing low-level features extracted from the input image. It extracts features from the image vector by convolution. In image processing and computer vision, convolution applies a kernel to an input image to extract specific features. Consider an image I, which can be represented as an n_1 by n_2 array and defined by a function as follows;

$$I: \{1, \ldots, n_1\} * \{1, \ldots, n_2\} \to W \subseteq \mathbb{R}(I, j) \mapsto I_{i,j}$$

$$K \in \mathbb{R}^{2h_1=1*2h_2+1}$$

$$= \begin{Bmatrix} K_{-h_1,-h_2} & \cdots & K_{-h_1,h_2} \\ \cdots & K_{0,0} & \cdots \\ K_{h_1,-h_2} & \cdots & K_{h_1,h_2} \end{Bmatrix}$$

The convolution algorithm on the image I given a kernel K is:

$$(I * K)_{r,s} := \sum_{u=-h_1}^{h_1} \sum_{v=-h_2}^{h_2} K_{u,v}$$

The rectified linear activation function (ReLU), a non-linear function that outputs 0 otherwise and the input directly if the input is negative, is always applied after the convolution layer. Its defined as;

$$f(x) = \max(0, x)$$

The output from convolution is then captured in the primary capsule, which also works on the basic idea of convolution but captures complex features from the stack of convolution outputs. It has a set number of primary capsules whose job it is to take the convolution's basic features and produce combinations of them. Following the completion of their calculations on their inputs, capsules "encapsulate" the outcomes into a condensed vector of highly detailed outputs. The output of the primary capsule can be represented as a combination of vectors u_i. The alignments of the vectors convert some internal state of the detected items, and these vectors convert probabilities that lower-level capsules recognized in the related objects. This is now delivered into the dense capsule layer.

The dense capsule layer learns the instantiation parameters of the capsules and models' higher-level properties using the input of the primary capsules. The following procedures take place in the dense layer;

1. The input vectors u_i are multiplied by their conforming weight matrices W_{ij}. These matrices encode imperative spatial relationships between lower-level features and higher-level features. This is done such that the vectors achieve the affine transformation. This multiplication gives the prediction vector $\hat{\mu}_{j|i}$ as follows;

$$\hat{\mu}_{j|i} = W_{ij} u_i$$

2. The weight matrices are adjusted by a coupling coefficient C_{ij}, which is introduced to multiply a capsule's output by the prediction vector. This enables capsules to decide on which higher level capsule to send their output. The decision the capsule makes relies on the dynamic routing algorithm. The prediction vector is multiplied by the conforming coupling coefficient C_{ij};

$$s = C_{ij} \hat{\mu}_{j|i}$$

3. The weighted input vectors are combined to give the output s_j:

$$s_j = \sum C_{ij} \hat{\mu}_{j|i}$$

4. The vector is passed through the novel "Squashing" activation function. The "Squashing" function is a nonlinear activation function that takes a vector and then "squashes" it to have a dimension not exceeding 1 without altering its direction. The squashing function is given by;

$$a \Leftarrow \frac{\|a\|^2}{\|a\|^2} \frac{a}{\|a\|}$$

It transforms vector s_j into vector v_j;

$$v_j = \frac{\|s_j\|^2}{\|s_j\|^2} \frac{s_j}{\|s_j\|}$$

2.3.2. Dynamic Routing

The dynamic routing algorithm's fundamental principle is that a lower-level capsule will convey its input to a higher-level capsule that "agrees" with it. The algorithm updates the dense capsule layer's coupling coefficient C_{ij}. The algorithm decides how to modify weights in a network, specifying the method for allocating weights to the connections between neurons. A capsule network modifies the weights so that nearby high-level capsules are closely connected with nearby low-level capsules. The affine transformation determines the proximity measure.

The input of the algorithm is the prediction vector $\hat{\mu}_{j|i}$, r the number of routing iterations and l the number of layers. The procedure as given by Sabour et. al. [24]:

1. Introduce b_{ij} a temporary value used to initialize the coupling coefficient C_{ij} and updated in the iteration process.
2. Iterate the for loop r times.
3. Apply the softmax function to b_{ij} to output a non-negative coupling coefficient C_{ij}, where all the outputs sum to 1 as follows:

$$C_{ij} = \frac{e^{b_{ij}}}{\sum_k e^{b_{ik}}}$$

4. Compute the he weighted sum for all capsule in the subsequent layer.
5. Squash the weighted sum for every capsule in the subsequent layer.
6. Check and update weights b_{ij}.

$$b_{ij} \leftarrow b_{ij} + \hat{\mu}_{j|i} * v_j$$

Where $\hat{\mu}_{j|i}$ is the input to the capsule from low-level capsule i, and $v_j$is the output of high-level capsule j.

7. Return and restarts the c_{ij} calculation and repeats r times.

Following r iterations, all higher-level capsule outputs were computed, and routing weights were determined. The forward pass may continue to the network's subsequent level. The dense layer will produce a set of capsule activations of equal dimensions. The activations represent the "explicit pose parameters" of the class each chest X-ray image is associated with. The output layer performs classification based on the presence and properties of the entities represented by the dense capsules. The probabilities of the different classes are obtained through a SoftMax activation function:

$$y = softmax(v_j)$$

$$y = \frac{e^{v_j}}{\sum_k e^{v_j}}$$

The CapsNet model was trained using the Squared Hinge Loss function. The essential premise is that the model "wants" the correct class for each image to score by a predetermined amount more than the erroneous classifications. It is represented as follows:

$$L_c = T_c(\max(0, m^+ - \| V_c \|)^2 + \lambda(1 - T_c)\max(0, \| V_c \| - m^-)^2$$

Where:

T_c = 1 if an image of class k is present, m^+ = 0.9, m^-- = 0.1 and v_j = vector obtained from dense capsule layer. The loss for a correct capsule is represented by the first component of the equation, and the loss for an inaccurate capsule is represented by the second term. The value of λ is 0.5 to limit the loss when some categories do not exist. In order to accomplish correct classification, it is important to make sure that the vector modulus in lieu of the digital capsules of this class is as large as feasible and the vector modulus of other classes is as small as conceivable. Each dense layer capsule loss is added together to create the overall loss.

2.4. Model Training

The key phase in machine learning is the training process, where a model is taught to detect patterns and make predictions based on prepared data. During training, the model acquires knowledge from the data, which improves its predictive abilities over time. Training involves several steps and procedures:

Network Initialization: The CapsNet architecture parameters, including capsule weights and other learnable parameters, were initialized. The Xavier initialization method was used to prevent issues like vanishing or exploding gradients by setting the initial weights of parameters based on input and output dimensions.

Forward Propagation: The training data was passed through the network during forward propagation. Activations and outputs of capsules in each layer were calculated based on input data and parameter values. A squashing function was used to activate capsules, representing the likelihood of specific features.

Loss Calculation: A suitable loss function, such as cross-entropy loss, was defined to measure the discrepancy between the model's predictions and true labels.

Iterative Training: Initial steps were repeated for a predefined number of epochs. The network learned from the training data by adjusting parameters based on gradients and the chosen optimization algorithm.

Validation and Hyperparameter Tuning: The model's performance was periodically evaluated on a separate validation dataset. Metrics like accuracy, precision, and recall were monitored. Validation results guided the fine-tuning of hyperparameters such as learning rate, regularization parameters, and network architecture to enhance the model's performance.

Testing: After completing training, the model's performance was assessed on a separate testing dataset not used in training or validation. This provided an unbiased evaluation of the model's generalization abilities and performance on unseen data.

The training process also employed various techniques:

Batch Processing: Data was processed in batches during each epoch to speed up training and use computational resources efficiently.

Optimizer Selection: The appropriate optimizer, such as the Adam optimizer, was chosen to reduce training time and effort.

Hyperparameters such as the number of epochs and the learning rate were fine-tuned to control the behavior of the model during training and impact the model's performance.

Regularization with Reconstruction: Regularization techniques with reconstruction loss were utilized to improve model stability and generalization.

2.5. Model Evaluation

The performance of the trained CapsNet model was evaluated and compared with outstanding deep learning models using the following metrics:

Accuracy was used to evaluate the overall correctness of a model's predictions by calculating the ratio of correctly classified cases to the total number of cases.

Precision was used to evaluate the proportion of correctly classified images from the total number of images classified as a specific class.

Recall was used to evaluate the proportion of correctly classified images out of the total number of actual cases of ground truth positives.

The F1 Score is the harmonic mean of precision and recall, providing a balanced evaluation metric, making it useful in scenarios where the dataset is imbalanced.

Receiver Operating Characteristic (ROC) curve: The

performance of the model at different categorization thresholds is graphically represented by the ROC curve. The trade-off between true positives and false positives is illustrated by plotting the true positive rate (recall) versus the false positive rate.

Area under the curve (AUC): It is used to express how well a classification model based on the ROC curve performed overall. It gives a single result that sums up the model's performance across all thresholds and indicates the percentage of correctly identified photos.

3. Results and Discussion

The dataset used in this study was secondary data obtained from Mendeley data, developed and uploaded by Kermany et al. (2018a) in 2018. The dataset consisted of 5,856 chest X-ray images. Among these, 1,583 were normal chest X-ray images from pneumonia cases, while 4,273 were pneumonia-infected chest X-ray images from normal cases. The distribution in the dataset reflects the prevalence of pneumonia cases in the dataset. It ensures that both normal and pneumonia-positive cases are adequately represented for training and evaluation. The dataset's composition enabled the CapsNet model to learn and differentiate between normal and pneumonia cases effectively. Appropriate pre-processing techniques were implemented.

A sample of images used can be viewed as follows:

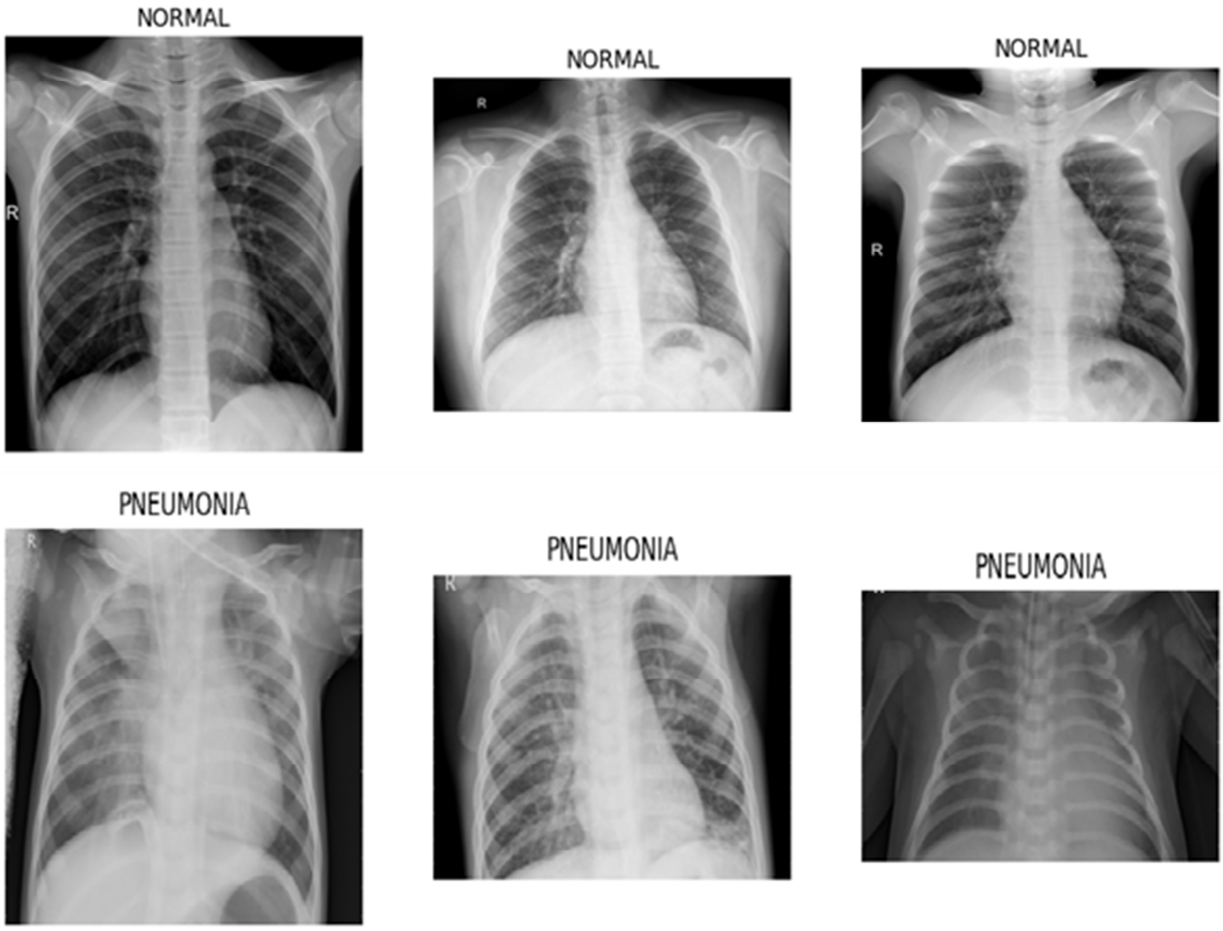

Figure 3. *Sample X-ray images.*

The images were pre-processed, the initial step they were resized to change their dimensions. This ensured that they had the same dimensions, making them compatible for further processing and analysis. The resized images were then standardized using min-max normalization to a consistent scale. This helped to bring all the pixel values to a similar scale and improved the model's convergence during training. Gaussian blur was applied to the images as a noise reduction technique. It is a widely used image-processing technique for reducing noise and smoothing images. It is based on the mathematical concept of convolving an image with a Gaussian kernel. Gaussian blur effectively reduces random and high-frequency noise in an image.

Histogram equalization was applied to the images; it is a widely used image enhancement technique that helps improve an image's contrast and visibility. It redistributes the pixel intensities in an image to cover a broader range of values, enhancing its overall appearance. Histogram equalization was applied to the images to help improve the visibility of details to capture the presence of any pneumonia infiltrations. The pre-processed images were viewed as follows where notable variations and improvements can be

observed.

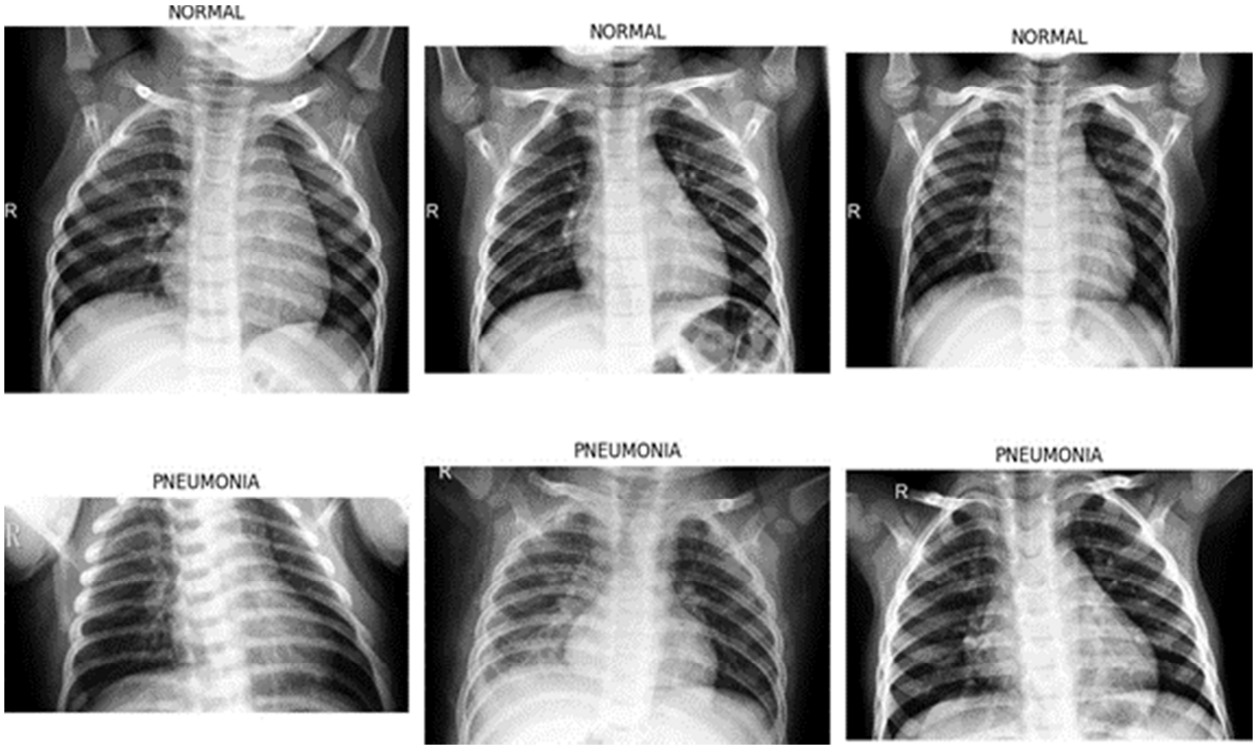

Figure 4. *Sample pre-processed X-ray images.*

The pre-processed data was split into 80% to 20% proportions for model training and evaluation, respectively. The 80% proportion was used as the training set, while 10% was used as the validation set and 10% as the testing set. There were 4684 chest x-ray images in the training set and 586 in the validation and testing sets. They were distributed as follows:

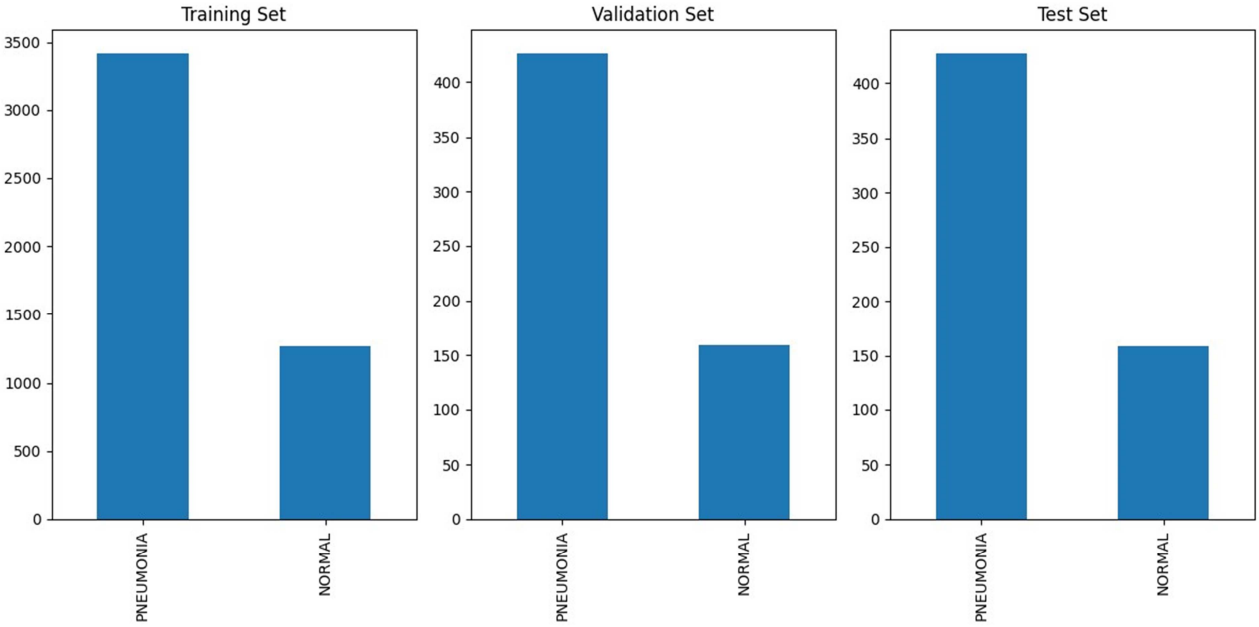

Figure 5. *Distribution of partitioned data.*

The distribution was imbalanced for the normal and pneumonia cases. Still, the split ensured a balanced distribution of proportions of cases in all the sets, which is key in the model validation and evaluation after training.

3.1. Model Training

The Capsule Neural Network (CapsNet) was meticulously trained throughout 50 epochs, each processing a batch size 32. A learning rate of 0.001 was employed throughout the training process. The output below was obtained to mark the training process:

Table 1. *Output of model training first 10 epochs.*

Epoch	Steps-time	Training		Validation	
		Loss	Accuracy	Loss	Accuracy
1/50	147/147– 422s	0.3227	0.8574	0.1432	0.9556
2/50	147/147– 433s	0.1611	0.9372	0.1646	0.9403
3/50	147/147– 423s	0.1540	0.9417	0.1284	0.9625
4/50	147/147– 429s	0.1259	0.9522	0.1136	0.9608
5/50	147/147– 415s	0.1087	0.9584	0.1183	0.9488
6/50	147/147– 427s	0.0943	0.9652	0.1062	0.9642
7/50	147/147– 414s	0.0752	0.9725	0.1636	0.9437
8/50	147/147– 413s	0.0627	0.9778	0.1114	0.9573
9/50	147/147 – 415s	0.0429	0.9855	0.1934	0.9300
10/50	147/147 – 407s	0.0395	0.9863	0.1475	0.9454

During the model training, several noteworthy observations were made. Initially, there was a significant improvement in model training accuracy as the number of epochs increased, peaking at epoch 16. Beyond this point, the model's training accuracy reached a plateau at a training accuracy of 1, indicating that further epochs did not significantly enhance its performance. The validation accuracy was 0.9539 at this level. Interestingly, it was also observed that the time required for each successive training phase decreased as we progressed through the epochs. Additionally, the model demonstrated its proficiency in minimizing the loss function, where the training loss was consistently decreased with each epoch to achieve a loss of 0.0016 at epoch 16. This pattern indicated that the model effectively penalized deviations from the expected outcomes, reflecting a steady linear learning rate. The accuracy and loss plots were observed as follows:

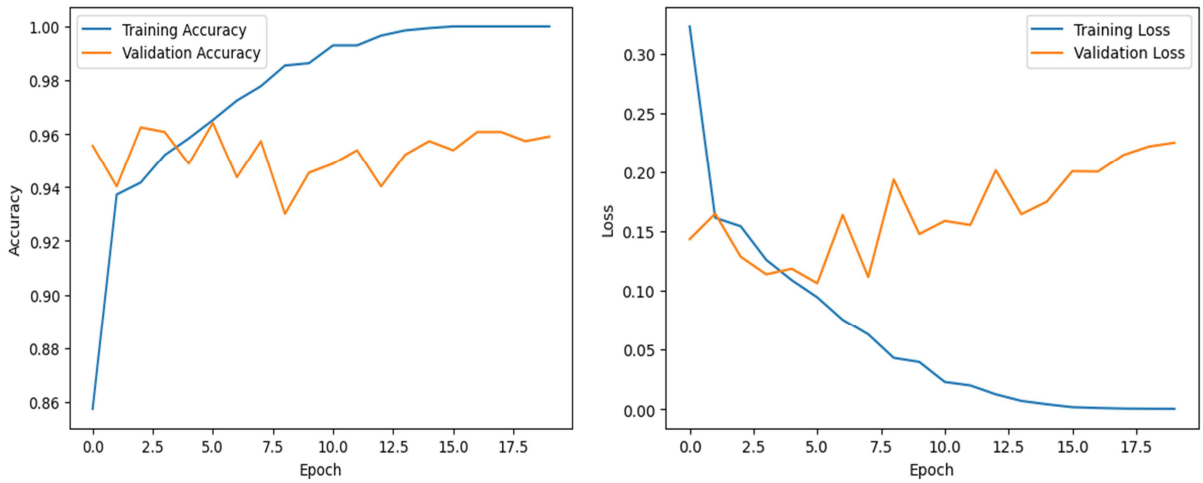

Figure 6. *Training and validation accuracy and loss plots.*

3.2. Model Evaluation

The classification metrics were obtained, the Capsule Network achieved an accuracy of 0.96 a precision of 0.9697, a recall of 0.9742 and a F1 score of 0.972. An accuracy of 0.96 implies that the model correctly predicted the classes of 96% of the samples in the dataset, which is a high overall correctness rate. This implies the CapsNet model performed well in diagnosing pneumonia from the chest X-ray images. A precision of 0.9742 suggests that 96.97% of the predicted pneumonia cases were indeed true positives, implying a reasonably low rate of false positive predictions. A recall of 0.9742 means that the model correctly captured 97.42% of the positive cases, indicating that it can detect pneumonia cases. The model F1 score considers both false positives and false negatives, as follows:

$$F\ 1\ score = 2 * (Precision * Recall) / (Precision + Recall) = 2 * (0.9697 * 0.9742) / (0.9697+0.9742) = 0.972$$

With an F1 score of 0.972, the model demonstrates a good trade-off between precision and recall, indicating overall effective performance.

The Receiver Operating Characteristic Curve was obtained as follows:

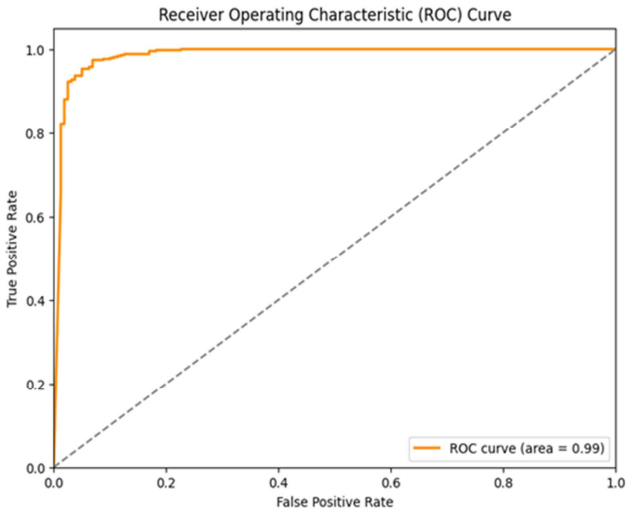

Figure 7. *ROC curve.*

The ROC curve shown above shows larger true positives and smaller false negatives since the true positive rate is represented by large numbers on the y-axis. The classification curve was deeply elongated to the top left, indicating a high performance for the classification model. The area under the curve for the classification was also obtained from the curve. The AUC was 0.99, which implies outstanding discrimination in the classification. The CapsNet model performed well in classifying pneumonia cases hence, it is very suitable for early pneumonia diagnosis.

3.3. Comparison

The CapsNet model findings were evaluated and compared with the outstanding deep learning models. The following results were obtained.

Table 1. *Model comparison metrics.*

Model	Metrics				
	Accuracy	Precision	Recall	F1 score	Time
CapsNet	0.96	0.9697	0.9742	0.9720	21
CNN	0.92	0.8920	0.9212	0.9352	42
Autoencoder	0.91	0.8820	0.9012	0.8965	36
ResNet	0.86	0.8420	0.8624	0.8862	48
CNN-LSTM	0.89	0.864	0.8456	0.8461	56

The trained CapsNet model outperformed the CNN, Autoencoder, a transfer learning ResNet model and an ensemble CNN-LSTM model in the diagnosis of pneumonia. The model competed fairly in terms of classification accuracy with the CNN and the Autoencoder but was able to outstand in the computation time. The time taken to train our model was 21 minutes compared to the large amount of time required to train the ensemble and transfer learning models. These results establish the foundation of this study to obtain a model that was accurate and efficient for pneumonia diagnosis. The Capsule network model is indeed accurate and efficient in the automatic detection of pneumonia from chest x-ray images.

4. Conclusion and Recommendations

The results and findings highlight the effectiveness of the CapsNet model as the main model for early diagnosis of pneumonia from chest X-ray images. With its powerful architecture and advanced features, the CapsNet model performs better in accurately classifying pneumonia cases. The high accuracy, precision, recall, and F1 score achieved by the CapsNet model emphasize its potential as a robust tool for pneumonia diagnosis. These findings underscore the importance of leveraging innovative deep learning techniques, such as CapsNet, in developing cutting-edge automated image classification systems for improved early detection and treatment of pneumonia.

The findings of this study recommend that there is need to conduct further studies to validate the performance of the models in real-world clinical settings. Collaborating with medical professionals and evaluating the models' effectiveness on patient data can help assess their practical utility and integration into the healthcare workflow. The major issue with health science is the clinical validation of developing technologies. There is, therefore, a dire need to come up with a way to ensure validation and further initiation.

The deep learning field has a swamp of upcoming future research that can enhance the capabilities and applicability of automated systems for early pneumonia diagnosis, ultimately contributing to improved patient care and outcomes. The findings provide valuable insights into the development and evaluation of models for early diagnosis of pneumonia from chest X-ray images.

Acknowledgments

I would like to appreciate the support of the faculty department of statistics and actuarial science in Jomo Kenyatta University of Agriculture and Technology together with my supervisor's Dr Antony Wanjoya and Thomas Mageto. The knowledge and skills they have instilled contributed towards the success of this research. I also appreciate my family for the moral and financial support throughout this journey.

References

[1] Aledhari, M., Joji, S., Hefeida, M., and Saeed, F. (2019). Optimized cnn-based diagnosis system to detect the pneumonia from chest radiographs. In 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pages 2405–2412. IEEE.

[2] Ayan, E., Karabulut, B., and Ünver, H. M. (2022). Diagnosis of pediatric pneumonia with ensemble of deep convolutional neural networks in chest x-ray images. Arabian Journal for Science and Engineering, pages 1–17.

[3] Chassagnon, G., Vakalopoulou, M., Paragios, N., and Revel, M.-P. (2020). Artificial intel- ligence applications for thoracic imaging. European journal of radiology, 123: 108774.

[4] Curigliano, G., Burstein, H. J., Winer, E. P., Gnant, M., Dubsky, P., Loibl, S., Colleoni, M., Regan, M. M., Piccart-Gebhart, M., Senn, H.-J., et al. (2017). De-escalating and escalating treatments for early-stage breast cancer: the st. gallen international expert consensus conference on the primary therapy of early breast cancer 2017. Annals of Oncology, 28(8): 1700–1712.

[5] Dadonaite, B., Ritchie, H., and Roser, M. (2018). Diarrheal diseases. Our World in Data.

[6] Erdem, E. and AYDİN, T. (2021). Detection of pneumonia with a novel cnn-based approach. Sakarya University Journal of Computer and Information Sciences, 4(1): 26–34.

[7] GM, H., Gourisaria, M. K., Rautaray, S. S., and Pandey, M. (2021). Pneumonia detection using cnn through chest x-ray. Journal of Engineering Science and Technology (JESTEC), 16(1): 861–876.

[8] Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., Venu- gopalan, S., Widner, K., Madams, T., Cuadros, J., et al. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. jama, 316(22): 2402–2410.

[9] Gupta, P. (2021). Pneumonia detection using convolutional neural networks. Science and Technology, 7(01): 77–80.

[10] Hashmi, M. F., Katiyar, S., Keskar, A. G., Bokde, N. D., and Geem, Z. W. (2020). Efficient pneumonia detection in chest xray images using deep transfer learning. Diagnostics, 10(6): 417.

[11] Kareem, A., Liu, H., and Sant, P. (2022). Review on pneumonia image detection: A machine learning approach. Human-Centric Intelligent Systems, 2(1-2): 31–43.

[12] Kermany, D., Zhang, K., Goldbaum, M., et al. (2018a). Labeled optical coherence tomography (oct) and chest x-ray images for classification. Mendeley data, 2(2): 651.

[13] Kermany, D. S., Goldbaum, M., Cai, W., Valentim, C. C., Liang, H., Baxter, S. L., McKeown, A., Yang, G., Wu, X., Yan, F., et al. (2018b). Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell, 172(5): 1122–1131.

[14] Lakhani, P. and Sundaram, B. (2017). Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. Radiology, 284(2): 574–582.

[15] Lee, S. M., Seo, J. B., Yun, J., Cho, Y.-H., Vogel-Claussen, J., Schiebler, M. L., Gefter, W. B., Van Beek, E. J., Goo, J. M., Lee, K. S., et al. (2019). Deep learning applications in chest radiography and computed tomography. Journal of thoracic imaging, 34(2): 75– 85.

[16] Liang, G. and Zheng, L. (2020). A transfer learning method with deep residual network for pediatric pneumonia diagnosis. Computer methods and programs in biomedicine, 187: 104964.

[17] Liz, H., Sánchez-Montañés, M., Tagarro, A., Domínguez-Rodríguez, S., Dagan, R., and Camacho, D. (2021). Ensembles of convolutional neural network models for pediatric pneumonia diagnosis. Future Generation Computer Systems, 122: 220–233.

[18] Milletari, F., Navab, N., and Ahmadi, S.-A. (2016). V-net: Fully convolutional neural networks for volumetric medical image segmentation. In 2016 fourth international conference on 3D vision (3DV), pages 565–571. Ieee.

[19] Nasrullah, N., Sang, J., Alam, M. S., Mateen, M., Cai, B., and Hu, H. (2019). Automated lung nodule detection and classification using deep learning combined with multiple strategies. Sensors, 19(17): 3722.

[20] Opuba, E. N. and Onyango, P. O. (2022). Health facility practices and patterns of prescription influencing health-seeking behaviour among caregivers of children diagnosed with pneumonia in endebess sub-county, kenya. Journal of Global Health Reports, 6: e2022029.

[21] Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Langlotz, C., Shpanskaya, K., et al. (2017). Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv preprint arXiv: 1711.05225.

[22] Rehman, A., Naz, S., Khan, A., Zaib, A., and Razzak, I. (2022). Improving coronavirus (covid-19) diagnosis using deep transfer learning. In Proceedings of International Conference on Information Technology and Applications, pages 23–37. Springer.

[23] Ruuskanen, O., Lahti, E., Jennings, L. C., and Murdoch, D. R. (2011). Viral pneumonia. The Lancet, 377(9773): 1264–1275.

[24] Sabour, S., Frosst, N., and Hinton, G. E. (2017). Dynamic routing between capsules.

[25] Sirazitdinov, I., Kholiavchenko, M., Mustafaev, T., Yixuan, Y., Kuleev, R., and Ibragimov, B. (2019). Deep neural network ensemble for pneumonia localization from a large-scale chest x-ray database. Computers & electrical engineering, 78: 388–399.

[26] Yee, S. L. K. and Raymond, W. J. K. (2020). Pneumonia diagnosis using chest x-ray images and machine learning. In proceedings of the 2020 10th international conference on biomedical engineering and technology, pages 101–1.