Comparative Evaluation of Inception V3 and ResNet 50 for Pneumonia Prediction

Anuja Bokhare[1], P.S. Metkewar[2], Sonakshi Ruhela[3], Madhav Avasthi[4], Aafaq A. Rather[5,*], Mohammad Shahnawaz Shaikh[6], Raeesa Bashir[7] and Showkat A. Dar[8]

[1]*School of Data Science, Symbiosis Skills and Professional University, Pune-412101, Maharashtra, India*

[2]*School of Computer Science and Engineering, Department of Computer Science and Applications, Dr. Vishwanath Karad MIT World Peace University, Pune 411038, Maharashtra, India*

[3]*Department of Liberal Arts, Faculty of Psychological Science, Rochester Institute of Technology, Dubai, UAE*

[4]*MIQ Digital India pvt Ltd., Bengaluru, India*

[5]*Symbiosis Statistical Institute, Symbiosis International (Deemed University), Pune, India*

[6]*Faculty of Engineering and Technology, Marwadi University, Rajkot, India*

[7]*Department of Mathematics and Statistics, Faculty of Science and Technology, Vishwakarma University, Pune, India*

[8]*Department of Computer Science and Engineering, GITAM University, Bangalore Campus-561203, India*

Abstract: Pneumonia is a fatal respiratory infection that has become the leading cause of death among many people across the world. Its widespread has grabbed great attention making it a major topic for research under various domains. Its severity has led to the development of systems that can predict whether a patient has chances of being diagnosed with pneumonia or not, this is also called as computer aided diagnosis. However, current study intends to identify an Artificial Neural Network (ANN) model that has been able to provide the highest accuracy when it comes to predicting this life-threatening condition. The prediction was initially done with Machine learning techniques but with the introduction of ANN, it was observed that there are models that provided higher accuracy than the ML models. This study investigates how the concept of deep learning which is a vital part of ANN makes use of one of its most efficient models including Inception V3 and ResNet 50 for the prediction of pneumonia and compare their performance to suggest a better solution to the problem. Results indicate that ResNet50 offers clinically meaningful improvements in sensitivity and specificity, supporting its role as a decision-support tool for early pneumonia detection.

Keywords: Pneumonia Classification, Artificial Neural Network, Deep Learning Algorithms, Inception V3, Resnet50.

1. INTRODUCTION

Pneumonia is a respiratory infection that is contagious and caused by bacteria and viruses. When these pathogens enter the lungs directly or indirectly through the blood flow into an already infected area of the alveoli in one or both lungs, they can cause pneumonia. The symptoms of pneumonia can range from mild to severe, and it can spread quickly through the lungs. The majority of children die from pneumonia. Spirato and faces can transmit the infection, while toys, cups, toothbrushes, and other contaminated items can also spread it. Viruses, bacteria and other infectious organisms such as fungi and parasite cause pneumonia.

Several studies have been undertaken to enhance machine learning models for forecasting pneumonia. Even so, the heavily biased predictions made within hospital or clinical settings have led to unsatisfactory outcomes. Fortunately, deep learning's recent advancements have resolved this issue. This article provides an overview of a successful experiment that utilized deep learning to predict pneumonia in varying population populations.

The Inception V3 model, generated by Google DeepMind, is cutting-edge deep neural network. It underwent training using the ImageNet dataset, which contains more than 1.2 million images spanning 1000 different classes. This model has demonstrated the capacity to identify relevant features in medical scans that can enhance the accuracy of diagnostic software [2].

This case study explains the use of Inception V3 to predict whether or not someone is suffering from pneumonia using deep learning methods. The benefits of using deep learning are utilized as well as what types of images get classified most accurately when using this method [2].

ResNet also called Residual Network is said to be a powerful CNN after it won an ImageNet competition in 2015 as it gave an error of just 3.57%. ResNet was able to tackle the problem of vanishing gradient with the help of skip connection as it allowed the training of extremely deep neural networks (150+ layers). The

*Address correspondence to this author at the Symbiosis Statistical Institute, Symbiosis International (Deemed University), Pune, India;
E-mail: aafaq7741@gmail.com

main focus of ResNet is computational accuracy unlike Inception which focuses on the computational cost.

One of the ResNet variants that have been considered to perform this study is ResNet-50, as the name suggests the version is capable of working with 50 neural network layers. The ResNet 50 model consists of 5 stages and each of these stages consists of the Convolutional block and the Identity block. Each of these blocks has 3 convolutional layers and the blocks of ResNet-50 are built using the bottleneck design.

This study brings novelty by comparing Inception V3 and ResNet 50 for pneumonia prediction, with ResNet 50 showing superior accuracy (95.73%). This unique approach highlights deep learning's potential in medical diagnostics and underscores the importance of optimal model selection for improved patient outcomes.

Pneumonia prediction using deep learning models, specifically Inception V3 and ResNet 50, is the subject of current study. Through a comparative analysis, ResNet 50 has shown greater accuracy, achieving 95.73% compared to Inception V3's 92.03%. This finding suggests the potential for early detection and prognosis of pneumonia, highlighting the transformative role of deep learning in medical diagnostics, particularly in the area of respiratory infections. This development holds great promise for precision medicine and improved patient outcomes. This work frames pneumonia prediction as a medical inference problem where statistical reliability is essential to reduce misdiagnosis. Deep learning models are interpreted as statistical classifiers supporting diagnostic decisions rather than purely computational tools.

The rest of the study is organized as follows: Section 2 presents the motivation for this work through the literature review, and Sections 3 present the methodology used for the analysis, about the dataset used for the research study. Section 4 discuss result analysis and discussion i.e., the results obtained after validating the model. Critical analysis is presented in section 5, which is followed by the Section 6 for conclusion and future scope in section 7.

2. LITERATURE REVIEW

[1] Pneumonia is a serious lung condition that can have severe consequences if not treated promptly. If left untreated, it can result in the death of the patient, leading to an increase in the mortality rate. The present research paper emphasizes the significance of timely detection of pneumonia through X-ray imaging. The researchers have employed various parameters extracted from the image to calculate the accuracy of the model and determine which model produces better results.

The methods that have been compared in this study were Multi-layer perceptron, Logistic regression, classification, random forest, and sequential minimal optimization. The NIH chest X-ray14 dataset has been used to test and train the classifier. The concept of ROI (Region Of Interest) was used for segmenting the image to find the main area of the image that needs to be focused on to extract the required features. The precision, sensitivity, accuracy, specificity, the area under the curve (AUC), and F1 score were considered as the performance metrics. It was observed that the accuracy was higher for all the classifiers when the classification was done using ROI-confined feature extraction (segmentation) when compared to its performance before segmentation, that is with the full chest image. However, the overall observation from this study was that the area under the curve of Multilayer Perceptron and the Logistic Regression classifier has provided the highest level of accuracy that was 95.388% and 95.631% respectively using segmentation, putting them both at a higher level when compared to the classifier models. [1] Although these classifier models have provided such a high level of accuracy, medical experts were yet to approve it as the results provided are entirely based on the features extracted from the image and not on any clinical background of the patient.

In [3] authors have discussed how the deep learning model DCNN (Deep Convolutional Neural Network) have proved to be a better approach to finding out if a patient has been infected with pneumonia or not over the regular classifier models such as Naive Bayes, random forest, decision tree, SVM, etc. based on some certain metrics like accuracy. The author [3] discusses the implementation of ML classifiers and DCNN framework on chest X-rays, medical experts can generate meaningful outputs, such as whether a patient has pneumonia or not. Feature extraction is performed on the images as a part of image pre-processing, and that data is then sent to the neural network. The last layer of the DCNN will give either '1' or '0' as the output, '1' for pneumonia and '0' for normal. The authors have made use of Python's sklearn library to build the models and to pre-process the images, while the Keras library was used for implementing the CNN model. To find out which model has performed the best, the performance metrics have considered the parameters; sensitivity, specificity, accuracy, MCC, f1-score, error rate, FP rate, TP rate and area under a curve. After experimenting on all the models, it was observed that DCNN had outperformed under all the parameters making it the best model for predicting pneumonia. [3] The DCNN classifier gave an accuracy of 84%. The future scope of this research was to implement optimization techniques for higher accuracy. Additionally, a large amount of image data used for testing which will improve the model.

The authors realized that even after introducing the Alexnet model [4], which significantly decreased the error rate (nearly 10%), and observed a small awareness into the internal operation and behavior of complex models. This lack of knowledge limited the scope of improvement and recognition of the problem inside the model. Thereafter, another study [5] have been tried to overcome this issue and suggested a graphical method to depict which in every layer stimulate a distinct feature map. This graphical method assisted to know how the features are generated in each layer. This helped authors to recognize the problem in the model. Researcher intended to use convolutions to map the features back to the input pixel space to inspect a specific activity. To achieve this, authors initialized the other activations to zero and performed a deconvolution on the features to get the results in the image pixel expense. The technique authors described proved to be successful since it crosses the outcome of AlexNet by 1.7% i.e., maximum in the model. Still the result on the PASCAL dataset showed a decline in the accuracy of the algorithm, which proved that this model could not be used for other datasets.

Authors [5] have introduced the ImageNet dataset, an indispensable resource in the field of computer vision research. This extensive collection of labeled images, encompassing a multitude of categories, has greatly furthered the progress of image recognition and machine learning. The authors meticulously describe the hierarchical organization of the dataset and its utility in the training and evaluation of various computer vision algorithms. The outcome establishes a vital basis for subsequent investigations and advancements in visual recognition systems.

Authors [6] have realized that the availability of relatively small datasets limited the efficiency and capacity of the algorithms used for the object recognition task. To learn about a large number of objects from thousands of images with relatively high efficiency, a model with a higher learning capacity was required. However, due to the extreme complexity ImageNet unable to solve the problem. So the model having previous knowledge about unseen data is required. Convolutional Neural Network (CNN) used to solve this problem. Authors have used the ReLU (Rectified Linear Units) nonlinearity function [6] instead of the tanh functions, that resolved the issue of saturation and maximise the training speed by multiple times. A limited GPU memory encouraged them to spread the net across two GPUs connected in cross-GPU parallelization architecture. The GPUs were connected in specific layers where they could read and write into one another's memory directly without the interference of the host machine. The architecture proposed by them proved to be successful as top-1, and top-5 error rates were significantly less (16.4% from 26.1%) than the previously available state-of-art technology. It has been observed that small dataset and inadequate GPU memory will be the problems during the research.

Researchers explore the rationale behind developing various CNN i.e., convolutional neural network models to predict pneumonic lungs based on chest X-rays obtained from a dataset. To evaluate the efficiency of these models, four CNN models were constructed, and the number of convolutional layers increased with each model. While larger datasets typically yield more accurate results, there is no theoretical proof that higher numbers of convolutional layers will lead to increased model accuracy. Despite the availability of several pre-existing convolutional networks, the authors have developed seven CNN models from scratch, with their performance evaluated based on recall, accuracy and F1 scores. The Keras library, built on top of TensorFlow, was utilized for constructing the models. The architecture of the models includes four key elements: pooling layer, convolutional layers, flattening layers, and fully connected layers. In the convolution layer, an input image is treated as a 3D matrix, with a filter or kernel function referred to as a feature detector applied to extract features. The resulting output is a feature map. These models employ the ReLU and softmax activation functions, with ReLU used in hidden layers and softmax applied in the output layer to classify data into multiple classes [7].

The researchers opted for max pooling, where the filter picks out the greatest value to be included in the output array. Max pooling is the more commonly used pooling approach. Flattening layer, at this step, the pooled feature map is then converted into an input vector and this is the process of flattening. The vectors obtained after flattening are the input for the fully connected layers, where all the layers are fully connected to one another and perform forward propagation. After forward propagation, backward propagation is performed and this is done until the error is highly reduced, for the experiment. The 4 models make use of Adam optimization algorithm. The conclusion, derived was that classifier model 3 provides the highest level of accuracy, that is 92.31% when compared to the other models. The future aim was to come up with models such as transfer learning models [17]. These models to be used on the data to provide even more refined outputs than the ones provided by these CNN models.

In another study [8], Object detection performance calculated using PASCAL VOC shows poor result as complex ensemble features were combined with simpler HOG-like features in the current best methods. To solve this problem, Ross Girshick proposed a simpler and more scalable solution by bridging the gap between object detection and image recognition. In the proposed approach, a comprehensive analysis was conducted on a given image, resulting in the extraction of around 2000 region proposals. By employing a selective search algorithm, the image was divided into smaller segments, which were then combined in a bottom-up fashion to identify regions with similar features, such as texture, size, and fill information. However, it is important to note that not all of the regions identified through this process were of interest to the researchers. Therefore, some of the regions are positive, and a majority are negative (from the background). These regions are warped from the image and fed into five convolutional layers and 2 fully connected layers to draw out relevant information. Support vector machines were used to classify deviated images from the extracted images. To enhance the speed of the CNN model, the authors divided the training process into two steps. Initially, the authors conducted pre-training on the ILSVRC2012 dataset, followed by domain-specific fine-tuning. During this process, the model utilized positive samples from the warped images, where the intersection of the real image and the warped region was greater than 0.5. This led to a 10% improvement in the results. The proposed model achieved a mean average precision of 31.4% on the ILSVRC 2013 dataset, surpassing the then state-of-the-art OverFeat model (24.3%). However, the authors did not take into account the potential benefits of applying CNN to each region separately, such as increased processing time and memory requirements.

The authors have [9] constructed upon prior research into visual recognition tasks and elucidated strategies that improved detection accuracy while concurrently enhancing training and testing speed. The authors have additionally identified limitations within the RCNN model. The training process was executed through a multi-step approach, during which a convolutional neural network (CNN) was applied to the outcomes of each step. This approach necessitated substantial memory to store the features and consequently rendered the process slower. The proposed fast RCNN model resolved the drawbacks discussed earlier by applying convolutions once to an image to extract image features rather than applying it to each object proposal. The resulting image were added to feature map of convolutional layer. These changes proved to be successful as the score of mean precision rise from 4 to 6% w.r.t., various datasets. The region proposal state was ignored by the author, which became a bottleneck region in the model and hampered the speed of the object detection model.

In [10], the authors have focused on using RCNN and CNN for visual recognition tasks. It was found that CNN outperformed all the best methods used previously for classification tasks with a single class label. However, for biomedical image processing, the desired output should include localization. The configuration used to solve this problem had its drawbacks, such as being slow and having redundancy. There was a trade-off between the use of context and localization. To address this issue, the authors proposed a new U-shaped architecture. This architecture first used down sampling to decrease the dimensions of the image and obtain a well-defined context. Then, it up sampled the image, resulting in a large number of feature channels. The authors decoded and then encoded the data and passed it through a bottleneck. They also used a skip connection, which helped to train the model even with a small amount of available data. The new architecture was found to be efficient, as it had the lowest error rate among the three types of datasets it was experimented on. However, the authors did not consider the negative impact of the bottleneck structure, which made the learning process slow. Additionally, U-Net was a 2D architecture, which did not allow for full volumetric processing of the image data.

In [11], the authors recognised that the region proposal state became a bottleneck region in the model and hampered the speed of the object detection model. To solve this problem and further enhance the speed of the algorithm, author have proposed RPN i.e., region proposal network for selective search. The outcome then passed as input to convolutional layer in this model. A feature map was obtained as outcome from it. To create a region proposal, a small convolutional layer to this feature map generated a box proposal. K box proposals were developed, which were generated as output, the box, and class information. The k class gave the probability of a box containing an object. The results [11] on PASCAL VOC 2007 and 2012 datasets show that the proposed method with RPN made only 300 region proposals while with selective search, it made 2000 region proposals, at the same time, decreasing the run time by 1,630 milliseconds and increasing the efficiency. However, the author [11] could not consider the fact that the RPN was trained at the anchor point, and since the features in a single image would be similar, it would increase the training time of the RPN.

The study [12] states which network algorithm gives the best output, to find that out the authors have considered 5 CNN-based pneumonia diagnosis algorithms; MobileNet which applies 3 × 3 depthwise separable convolutions, a regular CNN, ResNet-18, ResNet-50 and VGG19. A Kaggle dataset has been considered, which consists of chest X-rays of children with and without pneumonia, between the age of 1 to 5 from Guangzhou Women and Children's Medical Centre. The images had to be pre-processed to get more clarity and to get rid of any noise. The 5 models considered were run on the test dataset using Python. [12] It was observed that the MobileNet model provided the highest level of accuracy with a minimum loss, that is an accuracy of 92.79%. This technique was being seen as an efficient solution for early detection of pneumonia, thereby reducing the work of medical experts as well as this can help reduce the mortality rate of pneumonia patients. Due to the lightweight nature of MobileNet and the fact that it has comparatively lesser parameters when compared to the other models, this can be implemented on mobile devices and was seen as a feasible approach to resolving the problem of diagnosing a patient with pneumonia.

In a study conducted by the authors [13], the transfer learning method was utilized to develop and integrate five separate models, based on existing models found in the literature. Each of the developed models incorporated dropout layers, softmax function, RMSProp optimizer, and categorical cross-entropy. After evaluating the proposed models with the transfer learning+CNN model, it was determined that VGG16 achieved the highest accuracy rate of 96.81%, while VGG19, NasNetMobile, ResNet152V2, and InceptionResNetV2 attained accuracy rates of 96.58%, 83.37%, 96.35%, and 94.87%, respectively. Future research opportunities include enhancing the model's optimization and expanding the dataset.

The research proposes a computer-assisted diagnostic system for pneumonia detection, leveraging deep transfer learning and a combination of convolutional neural networks. The method achieved accuracy rates of 98.81% and 86.85% on the Kermany and RSNA datasets, respectively, surpassing current state-of-the-art approaches. The study employs novel weighting techniques for ensemble models to improve diagnostic reliability, which is validated through extensive statistical analysis [18].

In [14], the authors have implemented some pre-trained CNN models, they are Inception-v3, ResNet 50, and VGG 16. The idea of transfer learning [17] is highlighted here because an already used model is going to be reused to resolve the issue, which means a CNN model will be implemented along with the pre-trained models to develop an ensemble model. This promote the concept of transfer learning which in turn leads to better model performance. The authors [22] have considered a chest X-ray image dataset from Kaggle, the images are then pre-processed to get them to a standard size. After that the image data was augmented to increase the size of the dataset to avoid overfitting eventually in the process, this was mainly done for the normal chest images. The training data is sent to the ensemble model, and further on the testing data along with the ensemble model was sent to the trained model. The CNN model was used to convert the 3D data that was received as input into 2D data by implementing filtering for feature extraction, the models make use of the appropriate activation function and the most common one being ReLU. The convolutional layer is compiled with the help of Adam optimizer and the binary_cross entropy loss function.

ResNet50 Model as the name suggests is a CNN model with 50 layers, [16] here with the ImageNet database which has been trained on an extremely large number of images, has multiple pre-trained model versions that can be loaded as per the requirement. VGG-16 Model is another CNN model consisting of 16 layers. It provides decent outputs even with a small dataset due to its massive training. This model was mainly applied in the case of object detection. Here all three models are implemented with CNN making them ensemble models. The authors [14] have used Jupiter Notebook as the environment to experiment with the help of the Tensor Flow Library. When CNN model was compared with the other models i.e., VGG-16, Inception-v3, Resnet50; it was seen to give an accuracy of 98.25%, which was the highest. While the results of ensemble models were slightly different. It was seen that the accuracy of the models had increased, and it was a close call between Inception-V3 with custom built CNN model and ResNet50 with custom built CNN. Their accuracy was 99.29% and 99.08% respectively. K-fold cross-validation was done for all the standard models as well as the ensemble models and it was derived that the proposed approach is significant for pneumonia classification. It was observed that the ensemble methods provided more accuracy, and their usage must be made more often by field experts [16].

The authors [15] have made use of the classic CNN model as well as some of its pre-trained models such as AlexNet, VGG-16, and VGG-19 for feature extraction. The feature set has then been fed to the machine learning models like decision tree, KNN, linear discriminant analysis, linear regression, and support vector machine. It was observed that when the CNN models were applied to the five classifiers after the images were augmented provided a higher accuracy, than the original images were used. Here the authors [15] have made use of mRMR as the feature selection approach. When the output was obtained by applying mRMR on the features that were extracted from the various CNN models that made use of the augmented images, it was observed that a different machine learning algorithm out of the 5 gave a higher accuracy for the different combinations of CNN models. Proposed approach was used on both augmented and non-augmented data and it was found that SVM+mRMR provided an accuracy of 98.21% for non-augmented data and LDA+mRMR provided an accuracy of 99.41% for an augmented data set. Future research will examine how well the suggested technique performs on various datasets. [15] Designing high-performance computer-aided diagnosis systems for various medical imaging activities can be done generically using the presented method. Table **1** highlights the gap in existing research. study is conducted.

It is observed from the table that future research aims to enhance diagnostic reliability through transfer learning models and ensemble methods, design high-performance computer-aided diagnosis systems for broader medical imaging applications, and investigate novel image preprocessing techniques for improved feature extraction.

3. METHODOLOGY

Performance evaluation employs accuracy, sensitivity, specificity, precision, and F1-score derived from confusion matrices. These measures function as estimators of diagnostic reliability. Although formal hypothesis testing was not applied, metric comparisons provide inferential insights. Upon closely studying the literature; the two deep learning models that are widely implemented for the pneumonia prediction problem are Inception V3 and Resnet50, to choose the best model with the best accuracy in terms of predicting pneumonia samples from the healthy samples, this comparative study is conducted.

Image recognition model Inception V3 has been recognised to achieve greater than 78.1% accuracy on the ImageNet dataset. [1] The model signifies the result of numerous perceptions discovered over the time by several academics. Convolutions, average pooling, max pooling, concatenations, dropouts, and fully linked layers are among the symmetric and asymmetric building pieces that comprise the model.

The research design shown in Figure **1** outlines a systematic approach for comparing the performance of two prominent deep learning models, Inception V3 and ResNet50, on a Pneumonia dataset. Initially dataset is loaded and pre-processed to ensure it is suitable for training and testing. The dataset is then split into two parts, with 80% allocated for training and 20% for testing.

After splitting the dataset, the Inception V3 and ResNet50 models are trained separately using the 80% training data. After training, each model is tested using the 20% testing data, and their accuracies and other measures are recorded for further comparison purpose. The outcome helps in identifying better model for the given dataset.

3.1. Dataset Discussion

The dataset comprises 3278 chest X-ray images with a near-balanced class distribution (Pneumonia: 51.7%, Normal: 48.3%). This balance minimizes classification bias and supports meaningful statistical

Table 1: Summary of the Related Work

<table>
<tr><th>Strengths</th><th>Weaknesses</th><th>Gap in research</th><th>Reference</th></tr>
<tr><td>MLP and logistic regression achieved high accuracy (95.388% and 95.631%).</td><td>Model accuracy depends on image preprocessing and feature extraction methods.</td><td>Implement optimization techniques for higher accuracy in prediction models and perform the comparison within it.</td><td>[1]</td></tr>
<tr><td rowspan="2">DCNN demonstrated superior performance over traditional classifiers, achieving 84% accuracy.</td><td>Limited dataset size and GPU memory posed challenges for training deep learning models.</td><td>Explore transfer learning models and ensemble methods to improve diagnostic reliability.</td><td>[3]</td></tr>
<tr><td>Lack of clinical validation for model predictions raise concerns among medical experts.</td><td>Investigate the impact of proposed techniques on various datasets and medical imaging tasks.</td><td>[6]</td></tr>
<tr><td>MobileNet model showed high accuracy (92.79%) and efficiency for pneumonia detection</td><td>Bottleneck regions in models, such as region proposal state, hampered algorithm speed.</td><td>Design high-performance computer-aided diagnosis systems for broader medical imaging applications.</td><td>[12]</td></tr>
<tr><td rowspan="2">Ensemble models combining pre-trained CNNs like Inception-v3, ResNet 50, and VGG 16 achieved high accuracy (99.29% and 99.08%)</td><td rowspan="2">Nil</td><td>Develop methods for clinical validation, discuss the comparative analysis of existing methods.</td><td>[13]</td></tr>
<tr><td>Investigate novel image preprocessing techniques for improved feature extraction.</td><td>[14]</td></tr>
</table>

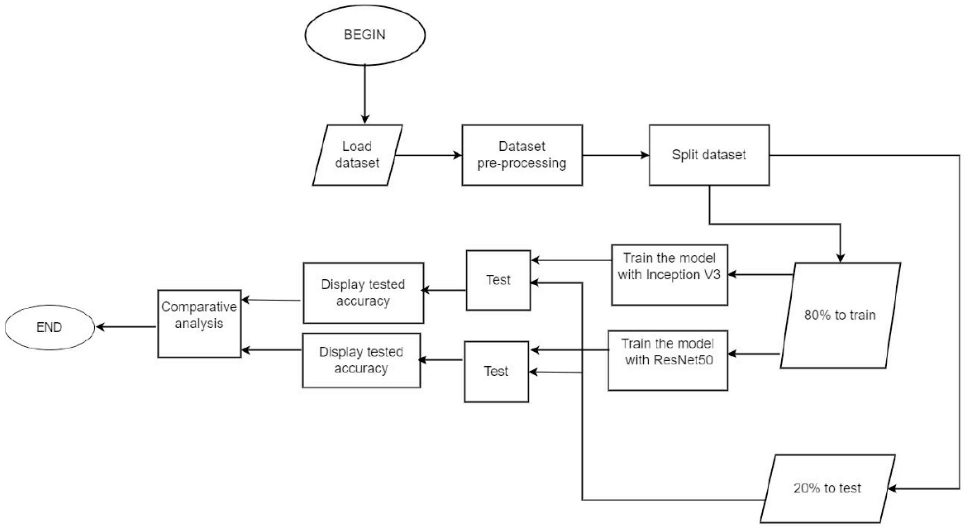

Figure 1: Framework of Research Design.

inference for medical analysis. The dataset is divided into two categories (Pneumonia and Normal). There are 3278 JPEG X-Ray pictures and two categories (Pneumonia/Normal) of which 1583 belonged to normal and 1695 belonged to pneumonia class. Figure **2** represents a sample chest X-ray belonging to the normal class and Figure **3** represents a sample belonging to Pneumonia class

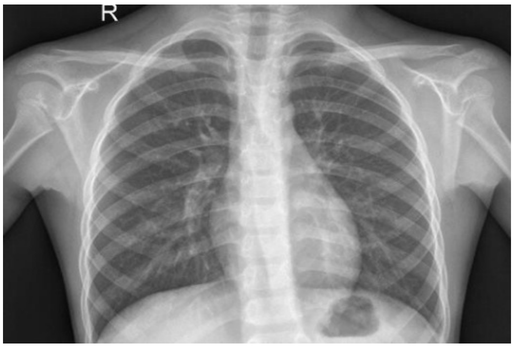

Figure 2: Normal Sample.

All chest X-ray imaging was performed as part of the patient's routine clinical care. All chest radiographs were originally filtered for quality control by deleting any low-quality or illegible scans before being analysed. The picture diagnoses were then assessed before being approved for training in the AI system. The evaluation set was also verified individually to account for any grading problems.

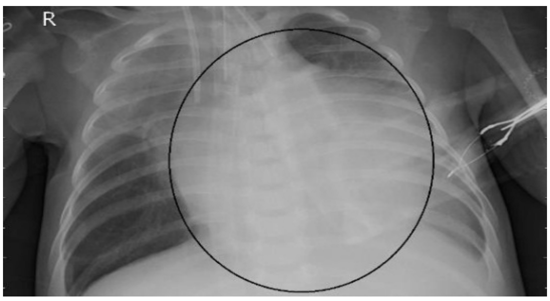

Figure 3: Pneumonia Sample.

Batch normalization is employed frequently throughout the model and is applied to activation inputs. [20] & [27] Softmax is used to calculate the loss. Inception V3 is a CNN model that works on low-powered computers, this is so because it reduces the number of parameters in the network making it a lightweight network and a more cost-effective method to undertake which is computationally efficient [14].

To fit the images with the neural network model, the images were zoomed and horizontally flipped to make them uniform and the shear range was kept at 0.2. This provided an even shape of the image for the entire dataset thereby removing the biases in the dataset due to the varying size of the images. Furthermore, each image was converted into (180, 180, and 3) sizes to enhance the speed of the model. This dataset was divided into 80-20 for the training and validation process and provided as input to the model.

3.2. Proposed Model Description

The model included *inception_v3_input* as the input layer, *inception_v3* as the functional layer followed by flatten which flattened the results, and *dens*e layer, which is followed by *dense_1* or the output layer as shown in Figure **4**.

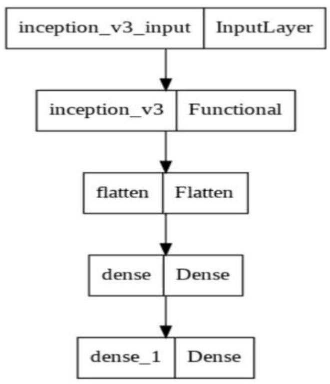

Figure 4: Inception V3.

For this project, authors have used a pre-trained *Inception V3 Model*. To receive high accuracy, the weights were set as per the processing done on the Imagenet dataset. The top layers of inception V3 were excluded from the learning process to prevent it from overfitting.

To make the results comparable for Inception V3 and Resnet50, all the other layers and parameters were kept constant, only the model was changed to Resnet50 as shown in Figure **5**.

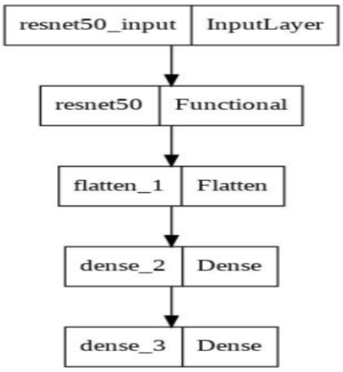

Figure 5: ResNet50.

The overall results of the model were flattened to comply with the next layers. Further, a layer with 512 nodes was added to the system which enabled the model to learn features according to the given dataset. The output layer has 2 nodes. The activation function for the output layer was set as softmax. The accuracy was recorded and compared with that obtained from Inception V3with Resnet50.

4. RESULT ANALYSIS AND DISCUSSION

Both models were validated on 657 samples and its accuracies were recorded with an accuracy based on the confusion matrix. As seen in Figure **6**, the Inception V3 model correctly identified 295 normal and 310 pneumonia images, with 21 false positive and 31 false negative predictions giving an accuracy of 92.03%. Higher sensitivity achieved by ResNet50 reduces false-negative diagnoses, which is critical for early clinical intervention. Improved specificity minimizes unnecessary follow-up procedures, enhancing clinical efficiency.

As observed in Figure **7**, the model's accuracy for the train and validation set is visualized graphically. The accuracy stabilizes beyond 34 epochs. The proficiency of this model proves to be promising as per Figure **8** which depicts the model loss for the training and validation set.

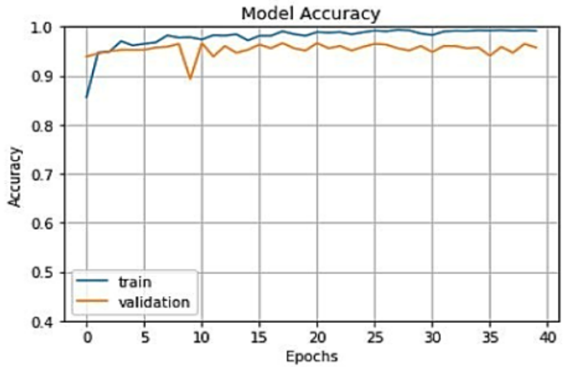

Figure 7: Model Accuracy - Inception V3.

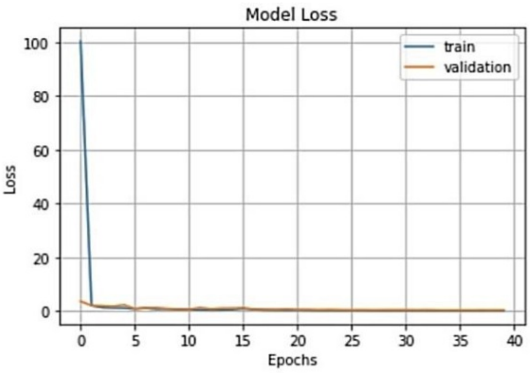

Figure 8: Model Loss - Inception V3.

The ResNet-50 model was validated on 657 samples which yielded an accuracy of 95.73%, which is higher than that provided by the Inception V3 model. This can be witnessed in the confusion matrix (Figure **9**) that has been produced for the model, where the

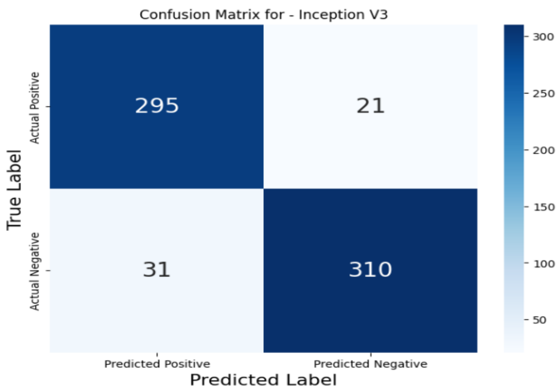

Figure 6: Confusion Matrix - Inception V3

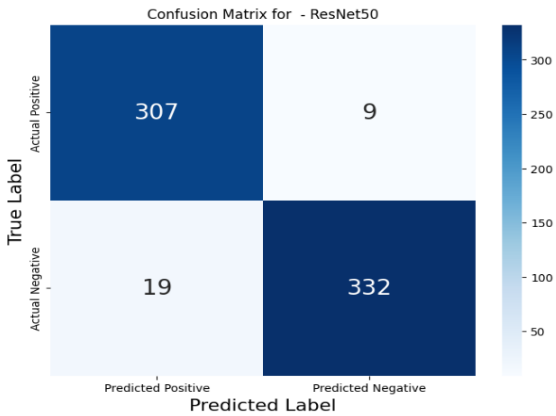

Figure 9: Confusion Matrix - ResNet50.

model correctly identified 307 images as normal and 322 as pneumonia infected. While 9 were false negatives and 19 were false positives.

From Figure **10**, which graphically represents the ResNet50 model's accuracy for the training and validation set it can be observed that the accuracy begins to become steady after 30 epochs. The efficiency of the model can be well understood from the graph in Figure **11** which represents the model loss for both the training and validation sets hence this model can also be used in cases of tuberculosis or any other lung-related diseases. This can help in preventing deaths from pneumonia and it can also help in regions with limited healthcare facilities available as well as early detection of pneumonia.

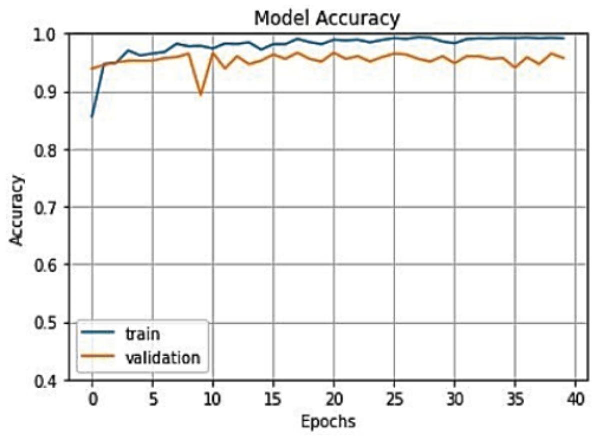

Figure 10: Model Accuracy - ResNet50.

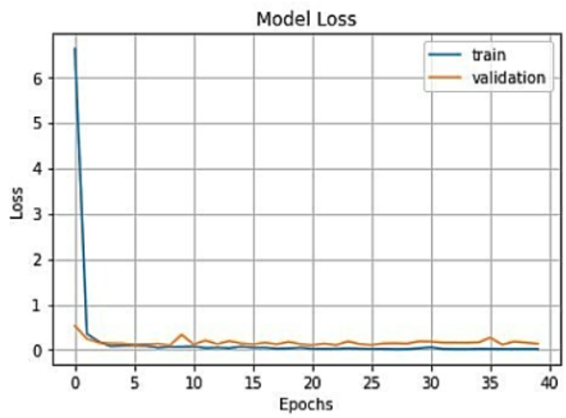

Figure 11: Model Loss- ResNet50.

Figure **6** and **9** indicate the inference that ResNet50 has fewer false positives and false negatives compared to Inception V3, indicating better performance in this context.

The performance of both the classifiers was evaluated based on accuracy, precision, recall, specificity, and F1-score. Authors initially calculated the number of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) labels from the confusion matrix, then following performance metrics were formulated for further comparison based on Equations (1)–(5);

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

Accuracy calculates the entire correctness of predictions given by the model, specifying the amount of correct predictions out of all predictions made.

$$Precision = \frac{TP}{TP+FN} \quad (2)$$

Precision calculates the correctness of positive predictions given by the model. A greater precision shows smaller amount of false positives, that means when the model predicts a positive outcome, it is more expected to be correct.

$$Recall = \frac{TP}{TP+FP} \quad (3)$$

Recall calculates the model's skill to correctly classify all positive outcomes. A greater recall shows that the model is capturing a bigger amount of actual positive cases, minimizing the probabilities of missing important outcomes.

$$Specificity = \frac{TN}{TN+FP} \quad (4)$$

Specificity calculates the accuracy of negative predictions given by the model. A higher specificity shows fewer false negatives, indicates that when the model predicts a negative outcome, it is more likely to be correct.

$$F1 - sccore = \frac{2TP}{2TP+FP+FN} \quad (5)$$

F1-score considers both false positives and false negatives, making it useful for evaluating models in conditions where there is an imbalance between positive and negative outcomes. A larger F1-score shows better performance in terms of both recall and precision.

Table 2: Performance Metrics for InceptionV3 and ResNet50

Model Name	Performance Measure								
	TP	TN	FP	FN	Accuracy	Precision	Recall	Specificity	F1-Score
InceptionV3	295	310	21	31	92.03%	93.20%	90.49%	93.62%	91.82%
ResNet50	307	322	9	19	95.73%	97.15%	94.18%	97.37%	95.64%

Table **2** describe the comparison between Inception V3 and Resnet50.

The Receiver Operating Characteristic (ROC) which shows the trade-off between False Positive Rate and True Positive Rate, helping in assessing the model's performance visually. Figure **13** indicates that ResNet50 has a very high recall and a very low false positive rate, suggesting excellent performance in correctly identifying positive cases with very few false positives as compared to ROC of Inception V3 in Figure **12**. The ROC analysis conclude that ResNet50 outperforms Inception V3, with a higher True Positive Rate and a lower False Positive Rate.

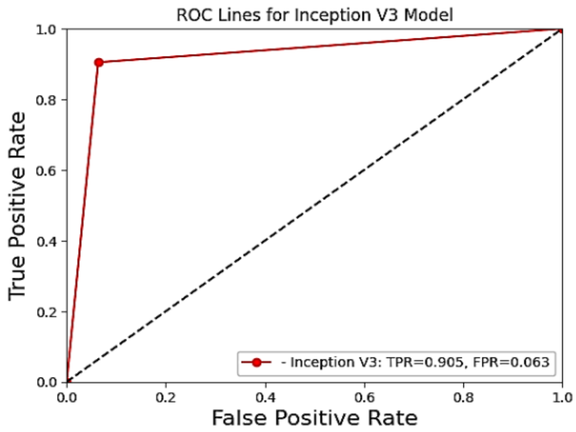

Figure 12: ROC for- InceptionV3.

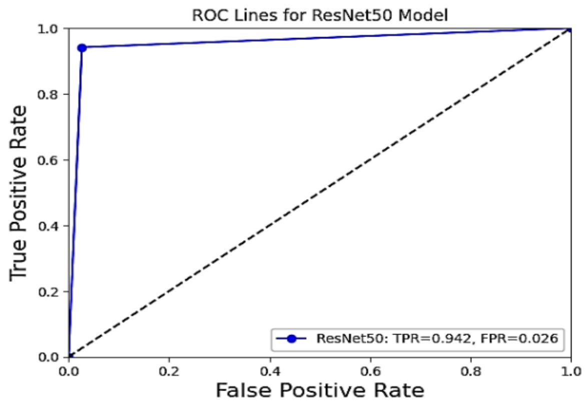

Figure 13: ROC for - ResNet50.

This indicates that ResNet50 is more effective at correctly classifying positive cases while making fewer false positive errors. Overall, ResNet50 provides superior classification performance.

The computation time observed for Inception V3 is 0.2505 seconds and for ResNet50 is 0.2563 seconds.

Limitations include dataset representativeness and absence of multi-center validation. Statistical uncertainty and generalizability should be addressed in future work using confidence intervals and hypothesis testing.

5. CRITICAL ANALYSIS

The Inception V3 and ResNet50 models have been validated and compared on 657 samples, providing significant insight into their effectiveness for forecasting pneumonia. Inception V3 model was 92, good for accuracy. distinguishing between 295 normal and 310 pneumonia images, but with 21 false positives and 31 false negatives. Observations of its accuracy and loss paths reveal consistent performance, with accuracy remaining stable beyond 34 epochs. Conversely, ResNet50 model outperformed Inception V3 with an accuracy of 95. 73%. ResNet50 was successful in capturing and transmitting 9 false negatives and 19 false positives. Around 307 normal images and 322 pneumonia images were transmitted, reflecting its robust performance over time. High precision, recall, specificity and F1-score are based on the capabilities of the models. Correctness is measured by accuracy, while precision emphasizes the accuracy of positive predictions, and recall measures model retrieval for capturing all positive outcomes.

The work's benefits include, its use of deep learning techniques and high accuracy, highlighting the potential of automated predictive models for medical imaging analysis. However, constraints such as the size of the dataset and potential imbalances within predictive strength; the practical use of these models in practice is significant, resulting in improved diagnostic accuracy and resource efficiency in healthcare. The advancement of respiratory medicine and the improvement of patient care worldwide is facilitated by this research.

CONCLUSION

Pneumonia, characterized by inflammation of the lungs' air sacs, presents diagnostic complexities, especially in distinguishing between viral and bacterial origins. Computer-Aided Diagnosis (CAD) systems play a crucial role in automating pneumonia identification through chest radiographic images, yet the task remains challenging.

To surmount this challenge, researchers have devised a deep learning-based prediction model that can accurately diagnose pneumonia in patients. The foundation of our predictive model is built on reports derived from chest X-ray imaging scans and the application of deep learning, renowned for its proficiency in processing visual data. Two deep learning models, the widely used Inception V3 and ResNet50 were incorporated and testing conducted on both models to check in diagnosing pneumonia.

Our comparative study results some intriguing findings. The Inception V3 model attained an accuracy of 92.03%, representing its strength to predict pneumonia. Whereas, the ResNet50 model dominated this performance with an accuracy rate of 95.73%. The study underscores the best effectiveness of the ResNet50 model for early-stage pneumonia identification and its possible application in clinical settings.

The "quantity" and "quality" of the dataset also to be considered during the training of the predictive model. Therefore, enhancement and expanding the dataset will be focused in future which will result in improving the trained predictive model helping in early detection of pneumonia. This study contributes to statistical medical research by demonstrating how deep learning models can support reliable diagnostic inference. Emphasis is placed on clinical relevance and statistical interpretability rather than accuracy alone.

FUTURE SCOPE

The proposed model has a potential for future refinement. It would be improved in future with focus on other respiratory conditions, including tuberculosis.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to the Editor and the anonymous reviewers for their valuable comments, constructive suggestions and insightful feedback, which significantly improved the quality and clarity of the manuscript.

REFERENCES

[1] Bahadur CT, Kesari V. Pneumonia detection on chest x-ray using machine learning paradigm. Proceedings of 3rd international conference on computer vision and image processing. Springer, Singapore 2020. https://doi.org/10.1007/978-981-32-9088-4_3

[2] Google. Advanced Guide to Inception V3 | Cloud TPU | Google Cloud. Google, 2022. Available at https://cloud.google.com/tpu/docs/inception-v3-advanced.

[3] Karan J, Nishtha H. Big data deep learning framework using keras: a case study of pneumonia prediction. 2018 4th International Conference on computing communication and automation (ICCCA). IEEE 2018. https://doi.org/10.1109/CCAA.2018.8777571

[4] Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. In Proc Eur Conf Comput Vision 2014; 818-833. https://doi.org/10.1007/978-3-319-10590-1_53

[5] Deng J, Dong W, Socher R, Li L-J, Kai L, Li F-F. ImageNet: A large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition 2009; 248-255. https://doi.org/10.1109/CVPR.2009.5206848

[6] Krizhevsky A, Ilya S, Geoffrey EH. Imagenet classification with deep convolutional neural networks. In Proc Adv Neural Inf Process Syst 2012; 1097-1105.

[7] Sirish Kaushik V, *et al*. Pneumonia Detection Using Convolutional Neural Networks (CNNs). Lecture Notes in Networks and Systems 2020; 471-83. https://doi.org/10.1007/978-981-15-3369-3_36

[8] Girshick R, Donahue J, Darrell T, Malik J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. 2014 IEEE Conference on Computer Vision and Pattern Recognition 2014; 580-587. https://doi.org/10.1109/CVPR.2014.81

[9] Girshick R. Fast R-CNN. In IEEE International Conference on Computer Vision (ICCV) 2015; arXiv:1504.08083. https://doi.org/10.1109/ICCV.2015.169

[10] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In Proc. Medical Image Comput. Comp.-Assis. Interv. – MICCAI 2015. Lecture Notes in Computer Science, vol. 9351, Cham: Springer. https://doi.org/10.1007/978-3-319-24574-4_28

[11] Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence 2017; 39(6): 1137-1149. https://doi.org/10.1109/TPAMI.2016.2577031

[12] Zhenjia Y, Liangping M, Runfeng Z. Comparison and validation of deep learning models for the diagnosis of pneumonia. Computational Intelligence and Neuroscience 2020. https://doi.org/10.1155/2020/8876798

[13] Hicham M, *et al*. Convolutional neural network based classification of patients with pneumonia using X-ray lung images. Transfer 2020; 2(99.41): 16.

[14] Muhammad M, *et al*. Pneumonia Classification from X-Ray Images with Inception-V3 and Convolutional Neural Network. MDPI 2022. https://doi.org/10.3390/diagnostics12051280

[15] Toğaçar M, *et al*. A deep feature learning model for pneumonia detection applying a combination of mRMR feature selection and machine learning models. Irbm 2020; 41(4): 212-222. https://doi.org/10.1016/j.irbm.2019.10.006

[16] Militante Sammy V, Dionisio Nanette V, Sibbaluca Brandon G. Pneumonia detection through adaptive deep learning models of convolutional neural networks. 2020 11th IEEE Control and System Graduate Research Colloquium (ICSGRC). IEEE 2020. https://doi.org/10.1109/ICSGRC49013.2020.9232613

[17] Vikash C, *et al*. A novel transfer learning based approach for pneumonia detection in chest X-ray images. Applied Sciences 2020; 10(2): 559. https://doi.org/10.3390/app10020559

[18] Rohit K, *et al*. Pneumonia detection in chest X-ray images using an ensemble of deep learning models. Plos One 2021; 16(9): e0256630. https://doi.org/10.1371/journal.pone.0256630

[19] Al Mamlook RE, Chen S, Bzizi HF. Investigation of the performance of Machine Learning Classifiers for Pneumonia Detection in Chest X-ray Images. 2020 IEEE International Conference on Electro Information Technology (EIT) 2020; 098-104. https://doi.org/10.1109/EIT48999.2020.9208232

[20] Christian S, *et al*. Rethinking the inception architecture for computer vision. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2016. https://doi.org/10.1109/CVPR.2016.308

[21] Erhan D, Szegedy C, Toshev A, Anguelov D. Scalable object detection using deep neural networks. In Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on 2014; 2155-2162. https://doi.org/10.1109/CVPR.2014.276

[22] Ji-Yul K, Soo-Young Y. Diagnostic Classification of Chest X-ray Pneumonia using Inception V3 Modeling. Journal of the Korean Society of Radiology 2020; 14(6): 773-780.

[23] Enes A, Murat ÜH. Diagnosis of pneumonia from chest X-ray images using deep learning. 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT) IEEE 2019. https://doi.org/10.1109/EBBT.2019.8741582

[24] Dong C, Loy CC, He K, Tang X. Learning a deep convolutional network for image super-resolution. In Computer Vision–ECCV. Springer 2014; 184-199 https://doi.org/10.1007/978-3-319-10593-2_13

[25] Agostinelli F, Hoffman M, Sadowski P, Baldi P. Learning activation functions to improve deep neural networks. 2014. arXiv:1412.6830. Available at https://arxiv.org/abs/ 1412.68302.

[26] Singh B, Alam I, Rather AA, Alam M. Linear Combination of Order Statistics of Exponentiated Nadarajah–Haghighi Distribution and Their Applications. Lobachevskii Journal of Mathematics 2023; 44(11): 4839-4848. https://doi.org/10.1134/S1995080223110318

[27] Rather AA, Subramanian C. A New Exponentiated Distribution with Engineering Science Applications. Journal of Statistics Applications & Probability 2020; 9(1): 127-137. https://doi.org/10.18576/jsap/090112

Received on 22-12-2025 Accepted on 22-01-2026 Published on 13-02-2026

https://doi.org/10.6000/1929-6029.2026.15.06