Enhancing Viral Pneumonia Diagnosis Accuracy Using Transfer Learning and Ensemble Technique from Chest X-ray Images

Chandrashekar Uppin cv.uppin@bazeuniversity.edu.ng
HOD, Department of Computer Science,
Baze University
Abuja, Nigeria

Usman Bello Abubakar usman.abu@bazeuniversity.edu.ng
Department of Computer Science,
Baze University
Abuja, Nigeria

Abstract

Pneumonia is an acute pulmonary infection that can be caused by bacteria, viruses, or fungi. It infects the lungs, causing inflammation of the air sacs and pleural effusion: a condition in which the lung is filled with fluid. The diagnosis of pneumonia is tasking as it requires a review of Chest X-ray (CXR) by specialists, laboratory tests, vital signs, and clinical history. Utilizing CXR is an important pneumonia diagnostic method for the evaluation of the airways, pulmonary parenchyma, and vessels, chest walls among others. It can also be used to show changes in the lungs caused by pneumonia. This study aims to employ transfer learning, and ensemble approach to help in the detection of viral pneumonia in chest radiographs. The transfer learning model used was Inception network, ResNet-50, and InceptionResNetv2. With the help of our research, we were able to show how well the ensemble technique, which uses InceptionResNetv2 and the utilization of the Non-local Means Denoising algorithm, works. By utilizing these techniques, we have significantly increased the accuracy of pneumonia classification, opening the door for better diagnostic abilities and patient care. For objective labeling, we obtained a selection of patient chest X-ray images. In this work, the model was assessed using state-of-the-art metrics such as accuracy, sensitivity, and specificity. From the statistical analysis and scikit learn python analysis, the accuracy of the ResNet-50 model was 84%, the accuracy of the inception model was 91% and lastly, the accuracy of the InceptionResNetv2 model was 96%.

Keywords: Viral Pneumonia, Chest X-ray, Transfer Learning, Ensemble Methods, Medical Image Classification, Artificial Intelligence.

1. INTRODUCTION

With more than 4 million fatalities every year, pneumonia ranks as a leading cause of death (Li et al., 2020). Pneumonia is an acute pulmonary infection that infects the lungs, causing inflammation of the air sacs and pleural effusion: a condition in which the lung is filled with fluid. Many people are affected by pneumonia, especially in developing and impoverished countries where there is a high degree of pollution, unclean living conditions, and overcrowding, as well as a lack of proper medical infrastructure (Kundu et al., 2021). The small air sacs (Alveoli) in the lung are affected by a variety of cases of pneumonia, which can be caused by bacterial, viral, or fungal infections. Patients with underlying illnesses like asthma, those with compromised immune systems, hospitalized babies, and elderly patients on ventilators may face a life-threatening situation from pneumonia if it is not detected in time (Bhandary et al., 2020).

Chest X-rays, Computed Tomography (CT) scans, and Magnetic Resonance Imaging (MRI) are just a few of the imaging techniques used in a clinical setting to detect pneumonia. Even though chest X-Ray radiographs are the most affordable diagnostic method for detecting pneumonia, diagnosing this illness from chest X-Rays requires highly competent radiologists because these images frequently overlap with other abnormal lung illnesses (Wang et al., 2019). This means that manually identifying pneumonia takes a long time and frequently results in subjective differences, which could delay diagnosis and treatment. In addition, it could be tasking as it requires a review of Chest radiography (CXR) by laboratory tests, vital signs, and clinical history.

The Computer-Aided Diagnosis (CAD) technique solves these issues by using Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) for automated pneumonia detection. DL has recently received a lot of attention due to its broad applicability to challenges involving automated feature extraction and image classification. Deep learning is a powerful tool for AI that is essential to solving many challenging computer vision issues. Convolutional neural networks (CNNs), in particular, are widely employed for a variety of image classification issues.

We examined the use of transfer learning, and also an ensemble technique to diagnose viral pneumonia from chest X-ray images in this research study. Utilizing previously trained deep learning models that have been trained on substantial datasets, like ImageNet, and modifying them for a particular task or domain is known as transfer learning. Two transfer learning models were used: Inception and ResNet-50. Then, for the ensemble, InceptionResNetv2 was applied which is an incorporation of residual connections and the inception module.

2. LITERATURE REVIEW

The researchers gathered chest X-ray images of people with pneumonia and healthy people from the Guangzhou Women's and Children's Medical Center using a publicly available dataset (Kermany et al., 2018). The researchers, (Yang & Mei, 2022) manually screened the dataset to remove the more inconsistent and redundant data and only preserved the high-quality X-ray images of pneumonia in order to ensure the model training and testing correctness. Additionally, they eliminated some of the pneumonia X-ray images from the dataset because there were more of them than there were of other X-ray conditions in this paper's dataset. In the end, they chose an equal number of images of patients with pneumonia and healthy individuals. The train test split ration used was 80:20 where 80% was for training and 20% for testing. LeNet5, AlexNet, MobileNet, ResNet18, and Vision Transformer models were established as five deep-learning models for pneumonia recognition (Yang & Mei, 2022). They compared and analyzed these five deep-learning models for pneumonia recognition to determine the best models for pneumonia recognition in various scenarios.

(Chandra & Verma, 2020)analyzed the lung areas of chest X-ray images, segmented the lung regions, and then extracted eight statistical traits from these regions. They applied and examined using five benchmarked classifiers named Multilayer Perceptron, Random Forest, Sequential Minimal Optimization (SMO), Logistic Regression, and Classification via Regression. A dataset of a total of 412 chest X-ray images containing 206 normal and 206 pneumonic cases from the ChestX-ray14 dataset was used in their research. With an accuracy of 95.63% with the Logistic Regression classifier and 95.39% with the Multilayer Perceptron, experimental findings show that the suggested method outperformed the current method (Chandra & Verma, 2020).

(Kuo et al., 2019) used a retrospective cohort study to identify pneumonia, using machine learning techniques, in 185 schizophrenic patients who were admitted to a Taiwanese district mental hospital between 2013 and 2018. Eleven predictors, including gender, age, clozapine use, drug-drug interaction, dosage, duration of medication, coughing, change of leukocyte count, change of neutrophil count, change of blood sugar level, change of body weight, were used to predict the onset of pneumonia (Kuo et al., 2019). Seven machine learning algorithms were used in the study to build predictive models: regression tree, decision tree, k-nearest neighbors, naïve Bayes, random forest, support vector machine, and logistic regression. On a dataset of 150 patients, the machine learning models were trained, and on a dataset of 35 patients, they were tested. They achieved the highest accuracy rate of 92%, using a decision tree classifier whilst the other models fell short by large margins (Kuo et al., 2019).

The CNN model proposed by(Sharma et al., 2020) achieved a classification accuracy of 90%. In their study, they classified images of chest X-rays for pneumonia using a convolutional neural network (CNN). They used a balanced dataset of 1000 images consisting of 500 chest X-ray images of patients with pneumonia and 500 images of patients without pneumonia. These images made up the dataset used to train and test the CNN. Five convolutional layers and three fully connected layers made up the CNN. Fully connected layers were used to classify the radiographic images, while convolutional layers were utilized to extract features from the radiographic images. The authors also noted that the CNN could be used to improve the efficiency of the diagnosis process. The CNN could be used to screen chest X-ray images for pneumonia, and only the images that are classified as suspicious by the CNN would need to be reviewed by a human expert. This would free up the time for human experts to focus on the more difficult cases (Sharma et al., 2020).

The authors, (Jaiswal et al., 2019), used deep learning techniques for identifying pneumonia in chest X-ray images in their research. The authors used the Radiological Society of North America (RSNA) Pneumonia Detection Challenge dataset, which contains over 26,000 chest X-ray images with annotations indicating the presence or absence of pneumonia. They split the dataset into training, validation, and test sets. Their approach was based on Mask-RCNN, a deep neural network that incorporates global and local features for pixel-wise segmentation. Initially, they illustrated how they employed a faster region-based convolutional network with pixel-wise instance segmentation for classification and localization to build the model. The mask-RCNN model takes a chest X-ray image as an input and predicts the bounding boxes of the image, label, and mask including classes. It extends the algorithm of F-RCNN by adding a branch that induces a binary mask predicting whether the given image pixel contributes to the given part of the object or not. The final model employed an ensemble approach where the Mask-RCNN was ensembled on ResNet-50 and ResNet-101 for image thresholding, and a mask region-based CNN for the detection of pneumonia traces using segmentation (Jaiswal et al., 2019).

Our research offers a concise and innovative approach to pneumonia classification by utilizing transfer learning and also, the ensemble method with InceptionResNetv2 and applying the Non-local Means Denoising algorithm. By using this ensemble and the Non-local Means Denoising technique to efficiently reduce the image noise, we fill a research gap and tend to reduce the detrimental effects of noise by improving the X-ray image quality. This, in turn, improves the effectiveness of our classification model. These contributions provide a foundation for more accurate and reliable pneumonia diagnosis, addressing key challenges in the field.

3. MATERIALS AND METHODS

In this study, we developed a pneumonia classification model from X-ray images using a combination of deductive and inductive approaches. We collected a labeled dataset of X-ray images and preprocessed them by resizing, normalizing, and reducing noise. Relevant features associated with pneumonia were extracted using established image processing techniques. Through deductive analysis, we validated the significance of these features by consulting medical literature and conducting experiments. An inductive approach was then employed to use pre-trained transfer learning classification models. The model's performance was evaluated using metrics like accuracy, sensitivity, specificity, and so on.

3.1 Dataset

The dataset used for this study was made publicly available by (Rahman et al., 2021), making it accessible for research purposes. The dataset consists of 1341 normal, and 1341 viral pneumonia chest X-ray (CXR) images. Each class contains a diverse range of X-ray images, allowing for comprehensive analysis and evaluation of machine learning models for pneumonia detection. Fig 1 shows an image of pneumonia affected and normal CXR images.

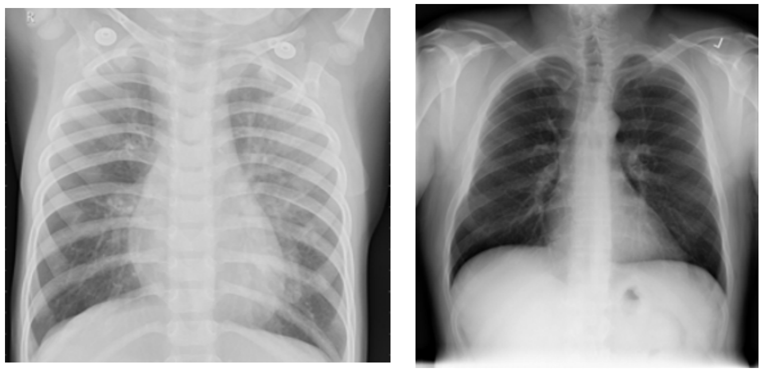

FIGURE 1: Pneumonia and Normal CXR image.

When training the deep learning model, we utilized the common organization of data using the train-test-val folder structure. The train-test-val folder structure is important because it helps avoid overfitting.

Overfitting occurs when the model learns the specific details of the training data too well, and as a result, it does not generalize well to new data. Table I shows the splitting of image data into a train, test and validation data.

Class	Total (T)	Training	Validation	Testing
0(Normal)	1341	858	214	261
1(pneumonia)	1341	858	214	261

TABLE 1: Image Distribution/

3.2 Pixel Normalization

The pixel values of the CXR images were normalized to bring the pixel intensity values within a consistent range or scale. This was done to improve the convergence and stability of the training process of the deep learning classifier. It ensures that all input features (pixels in this case) have similar scales, preventing any particular feature from dominating the learning process. We used the ImageDataGenerator class from the Keras library to perform pixel value normalization as part of our data preprocessing pipeline. We passed the value of 1.0/255.0 as the argument of the rescale named parameter of the ImageDataGenerator function. This implied that each pixel value in the image will be divided by 255, resulting in pixel values between 0 and 1.

3.3 Image Denoising

The images underwent a denoising phase, which refers to the process of reducing or removing noise from an image while preserving the important details. This is an essential pre-processing step because it helps to improve data quality, feature extraction, generalization, and performance of deep learning models. Firstly, we identified the kind of noise present in the image to make sure we used the appropriate denoising technique. To do this, the images were converted to grayscale and then their standard deviation was calculated using the standard deviation(std) function in the Python NumPy package.

We analyzed the standard deviation of the grayscale image, and classified the type of noise in the X-ray image accordingly: The noise gotten was the speckle noise. This type of noise is also common in X-ray images and is characterized by the presence of small bright and dark spots in the image (Manson et al., 2019).

To remove the noise, we employed a denoising algorithm called the Non-local Means Denoising algorithm, which is effective in reducing speckle noise while preserving important image details (Buades et al., 2011). Fig. 2 shows, from left to right, two images of a noisy CXR image and its denoised equivalent.

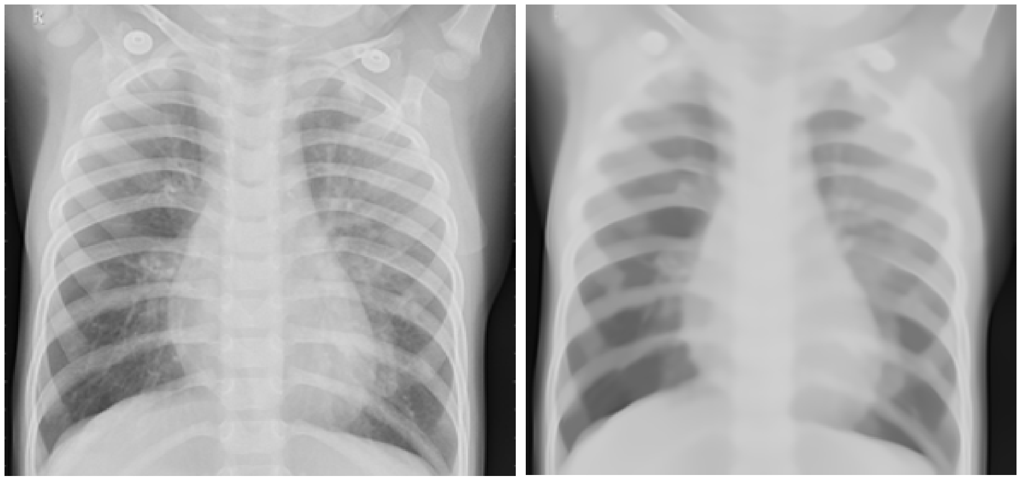

FIGURE 2: Noisy CXR Image and Denoised CXR Image.

3.4 Model Architecture

We proposed a transfer learning approach, using the ImageNet weights to be transferred to three pre-trained models: Inception Network, ResNet-50, and, an InceptionResNetV2. This enables the model to learn effectively over a smaller dataset due to its transferrable learning from the ImageNet dataset.

The GoogleNet is a 22-layer deep convolutional neural architecture that addressed computer vision issues such as object recognition and image classification in the ImageNet It has achieved 93.9% accuracy in the top 5 results (Abubakar et al., 2022). ResNet-50 is a 50-layer CNN. Five stages, each with a convolution and an identity block, make up the ResNet-50 model. As with the identity blocks, each convolution block has three convolution layers. The identity connection between the layers is the single addition to the basic network that transforms it into a residual network (Abubakar et al., 2022).

The InceptionResNetv2 is a convolutional neural network that was trained using more than a million photos from the ImageNet collection. The 164-layer network can categorize photos into 1000 different image categories (Zahra Elhamraoui, 2020). It is constructed using an ensemble technique of both the Residual connection and the Inception structure. Multiple convolutional filters of various sizes are incorporated with residual connections in the Inception-Resnet block. In addition to avoiding the degradation problem caused by deep structures, the inclusion of residual connections shortens training time (Zahra Elhamraoui, 2020). The architecture of the models are depicted in Figure 3, Figure 4, and Figure 5.

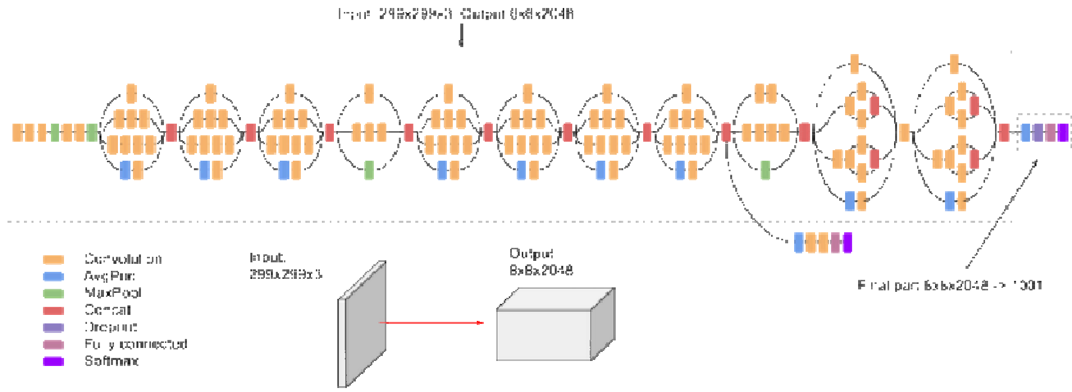

FIGURE 3: Inception Model (Abubakar et al., 2022).

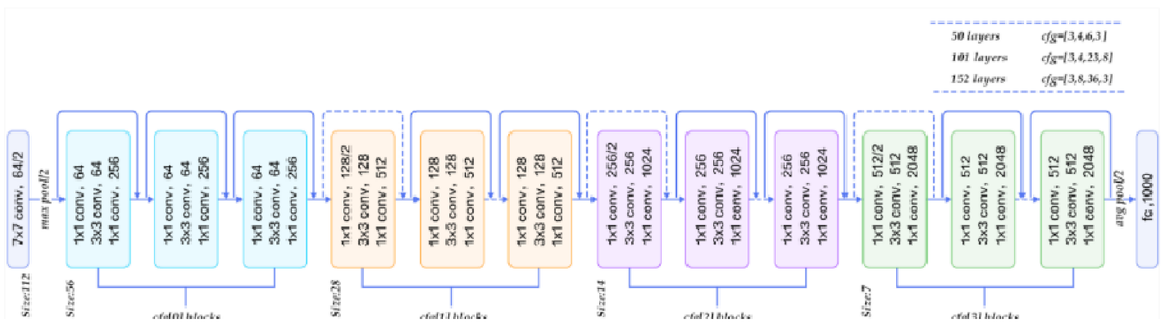

FIGURE 4: ResNet-50 Model (Abubakar et al., 2022).

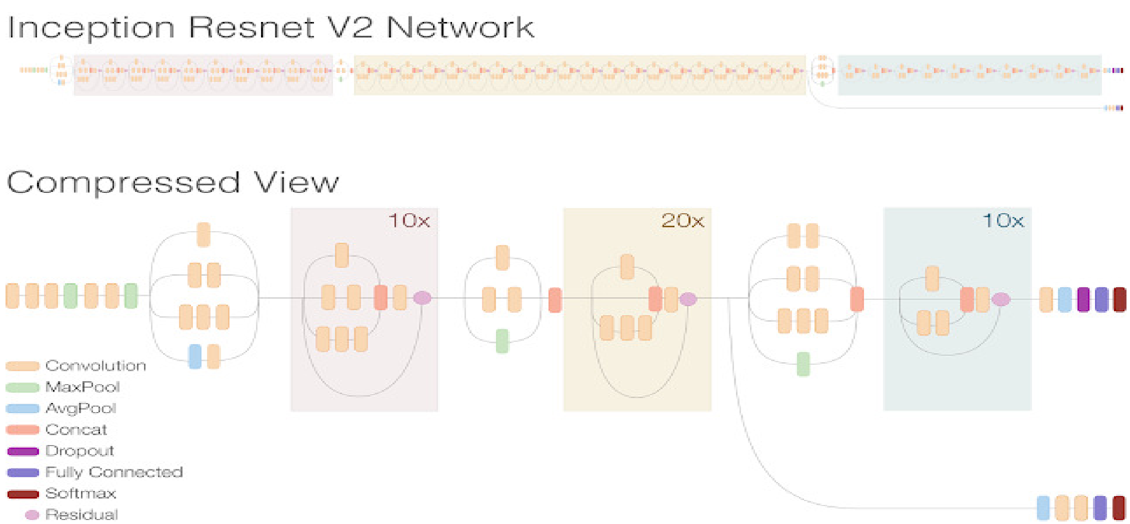

FIGURE 5: InceptionResNetv2 Model (Zahra Elhamraoui, 2020).

All the layers of the pre-trained model were made to be non-trainable. However, some of the layers could be re-trained to increase performance but at the cost of a higher chance of model overfitting. For the models, as the loss metric, binary_crossentropy was used as the dataset target has two classes (i.e., binary classification problem). Adaptive Moment Estimation (Adam) is the chosen optimizer, and the learning rate used was 0.001. The model underwent 30 epochs of training.

3.5 Model Training

The pre-trained models were loaded using a deep learning framework. The model's weights are frozen to prevent updates during training. The last classification layer of the model was replaced with a new fully connected layer to match the number of classes in the pneumonia classification task (i.e., two classes).

During training, the models were trained on the training dataset using an appropriate optimizer which minimizes the loss function. The optimizer used was the Adaptive Moment Estimation (Adam) optimizer. This was used because it often converges faster than other optimization algorithms and it combines the advantages of both RMSprop and momentum optimization. The learning rate, which controls the weight update step size, is set based on experimentation and we decided to use 0.001. The models were trained for 30 epochs, iterating over the training dataset in batches.

After training, the models were evaluated on the test dataset to assess its performance on unseen data. Classification metrics like accuracy, sensitivity, specificity, and precision are computed to evaluate the model's performance.

4. RESULTS

4.1 Classification Metric

The following metrics were established for each model to fully assess its performance: (1) sensitivity, (2) specificity, (3) accuracy, (4), and precision. The formula for the specified metrics is expressed below.

$$\text{Sensitivity} = \frac{truepositive}{truepositive+falsenegative} \quad (1)$$

$$\text{Specificity} = \frac{truenegative}{truenegative+falsenegative} \quad (2)$$

$$\text{Accuracy} = \frac{truenegative+truepositive}{allcases} \quad (3)$$

$$\text{Precision"} = \frac{truepositive}{truepositive+falsepositive} \quad (4)$$

Table 2 shows the sensitivity, accuracy, and specificity of the models.

	Accuracy	Sensitivity	Specificity	Precision
ResNet-50	0.84	0.85	0.85	0.84
Inception	0.91	0.92	0.92	0.90
InceptionResNetv2	0.96	0.97	0.96	0.95

TABLE 2: Result Obtained.

4.2 Confusion Matrix

In addition to giving insight into the mistakes, the classifier is making, a confusion matrix is utilized to reveal the specific mistakes that are occurring. The confusion matrix helps to overcome the limitation of using classification accuracy alone. Figure 6, Figure 7, and Figure 8, shows the confusion matrix of the models.

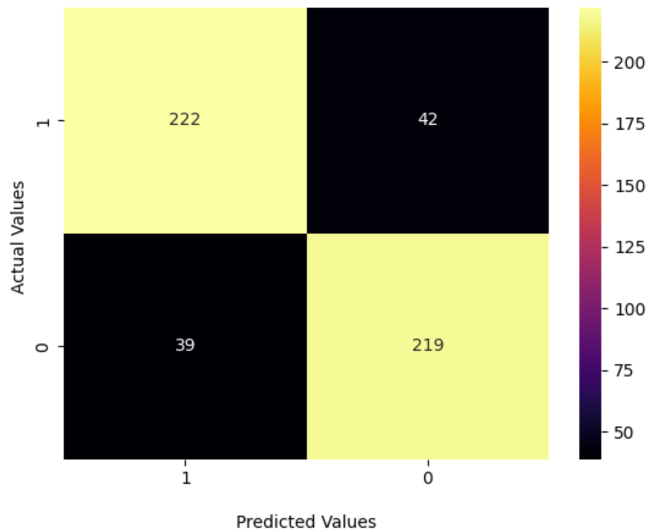

FIGURE 6: Confusion Matrix for ResNet-50.

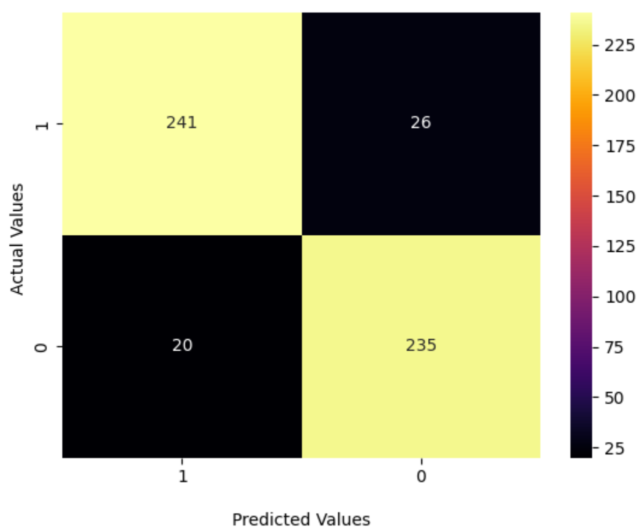

FIGURE 7: Confusion Matrix for Inception.

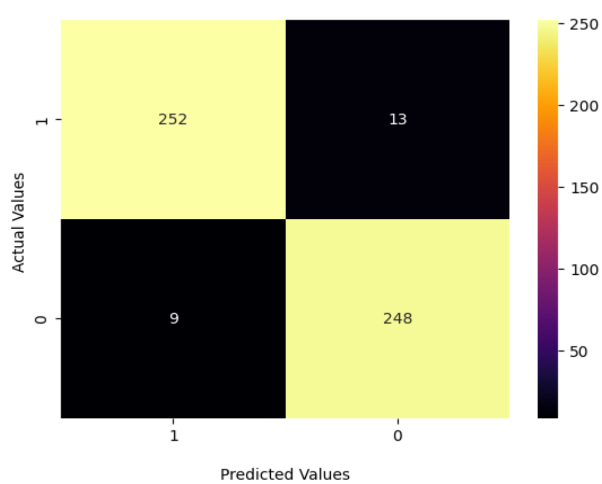

FIGURE 8: Confusion Matrix for InceptionResNetv2.

4.2 Discussion: Impact and Improvement

The findings of our research on the classification of pneumonia using several transfer learning models, such as ResNet-50, Inception, and InceptionResNetv2, show a considerable impact and improvement in accuracy when compared to current methods.

First, an accuracy of 84% was attained with the ResNet-50 model. Although encouraging, this result falls short of the accuracy attained by more sophisticated models. However, it shows the necessity for more complex designs and provides a baseline for comparison.

With an accuracy of 91%, the Inception model significantly outperforms ResNet-50. The model can recognize complex dependencies and patterns in the X-ray pictures thanks to the adoption of the Inception architecture, which combines many parallel convolutional layers with various filter sizes. This increase in accuracy shows how well the Inception model distinguishes between cases of pneumonia and non-pneumonia.

The most remarkable advancement is observed with the InceptionResNetv2 model, which achieved an accuracy of 96%. This state-of-the-art architecture combines the strengths of both Inception and ResNet by incorporating residual connections and inception modules. The InceptionResNetv2 model surpasses both the ResNet-50 and Inception models, showcasing the significant impact of leveraging advanced architectural design on pneumonia classification accuracy. The higher accuracy of the InceptionResNetv2 model indicates its ability to capture fine-grained details in the X-ray images and effectively distinguish pneumonia cases.

These advancements have significant effects on medical diagnosis and treatment. A precise and effective classification of pneumonia can help health care professionals make prompt, informed decisions that enhance patient outcomes. The InceptionResNetv2 model exhibits considerable potential as a useful tool to aid radiologists and healthcare professionals in reliably and promptly detecting pneumonia thanks to its accuracy of 96%.

5. CONCLUSION

Our research endeavors to answer the fundamental research question: "Can transfer learning and an ensemble method of deep learning, and the implementation of the Non-local Means Denoising algorithm improve the accuracy of pneumonia classification from X-ray images?" Through our study, we have successfully addressed this research question and demonstrated the effectiveness of the ensemble method with InceptionResNetv2 in combination with the Non-local Means Denoising algorithm. By leveraging these approaches, we have achieved significant advancements in pneumonia classification accuracy, paving the way for enhanced diagnostic capabilities and improved patient care.

The achieved accuracy serves as a measure of how effectively the transfer learning method employs pre-trained models to categorize pneumonia from chest X-ray images. Our results indicate that the ResNet-50 model attained an accuracy of 84%, highlighting its potential as a practical option for pneumonia detection. However, the Inception model surpassed ResNet-50 with an accuracy of 91%, showcasing its enhanced capability to extract crucial information and differentiate between individuals with pneumonia and those in good health.

The InceptionResNetv2 model achieved the highest accuracy of 96%, underscoring its superior performance in pneumonia detection. This can be attributed to the incorporation of residual connections and the inception module in the InceptionResNetv2 architecture, which enhances its ability to discriminate and represent features.

The higher accuracy achieved by InceptionResNetv2 suggests its potential for real-world applications, where accurate and reliable diagnosis is crucial. Nevertheless, to promote widespread adoption and improve the overall effectiveness of pneumonia detection systems, future work should focus on enhancing and evaluating these models using larger and more diverse datasets.

Our findings have two practical consequences. First of all, the improved accuracy is an aid for radiologists and healthcare practitioners, lowering human error and increasing diagnostic confidence. By incorporating these models into their workflow, they can streamline the diagnostic process, resulting in increased efficiency and faster turnaround times. Secondly, the incorporation of automated pneumonia classification models can help healthcare organizations by increasing resource allocation and patient care. Radiologists, healthcare organizations, and medical researchers are the intended audience for this study.

6. REFERENCES

Abubakar, U. B., Boukar, M. M., & Adeshina, S. (2022). Comparison of Transfer Learning Model Accuracy for Osteoporosis Classification on Knee Radiograph. *2022 2nd International Conference on Computing and Machine Intelligence, ICMI 2022 - Proceedings*. https://doi.org/10.1109/ICMI55296.2022.9873731

Bhandary, A., Prabhu, G. A., Rajinikanth, V., Thanaraj, K. P., Satapathy, S. C., Robbins, D. E., Shasky, C., Zhang, Y. D., Tavares, J. M. R. S., & Raja, N. S. M. (2020). Deep-learning framework to detect lung abnormality – A study with chest X-Ray and lung CT scan images. *Pattern Recognition Letters*, *129*, 271–278. https://doi.org/10.1016/J.PATREC.2019.11.013

Buades, A., Coll, B., & Morel, J.-M. (2011). Non-Local Means Denoising. *Image Processing On Line*, *1*, 208–212. https://doi.org/10.5201/IPOL.2011.BCM_NLM

Chandra, T. B., & Verma, K. (2020). Pneumonia Detection on Chest X-Ray Using Machine Learning Paradigm. *Advances in Intelligent Systems and Computing*, *1022 AISC*, 21–33. https://doi.org/10.1007/978-981-32-9088-4_3/COVER

InceptionResNetV2 Simple Introduction | by Zahra Elhamraoui | Medium. (n.d.). Retrieved June 13, 2023, from https://medium.com/@zahraelhamraoui1997/inceptionresnetv2-simple-introduction-9a2000edcdb6

Jaiswal, A. K., Tiwari, P., Kumar, S., Gupta, D., Khanna, A., & Rodrigues, J. J. P. C. (2019). Identifying pneumonia in chest X-rays: A deep learning approach. *Measurement*, *145*, 511–518. https://doi.org/10.1016/J.MEASUREMENT.2019.05.076

Kermany, D., Zhang, K., & Goldbaum, M. (2018). *Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification*. *2*. https://doi.org/10.17632/RSCBJBR9SJ.2

Kundu, R., Das, R., Geem, Z. W., Han, G. T., & Sarkar, R. (2021). Pneumonia detection in chest X-ray images using an ensemble of deep learning models. *PLOS ONE*, *16*(9), e0256630. https://doi.org/10.1371/JOURNAL.PONE.0256630

Kuo, K. M., Talley, P. C., Huang, C. H., & Cheng, L. C. (2019). Predicting hospital-acquired pneumonia among schizophrenic patients: A machine learning approach. *BMC Medical Informatics and Decision Making*, *19*(1), 1–8. https://doi.org/10.1186/S12911-019-0792-1/TABLES/5

Li, Y., Zhang, Z., Dai, C., Dong, Q., &Badrigilan, S. (2020). Accuracy of deep learning for automated detection of pneumonia using chest X-Ray images: A systematic review and meta-analysis. *Computers in Biology and Medicine*, *123*, 103898. https://doi.org/10.1016/J.COMPBIOMED.2020.103898

Manson, E., Ampoh, V. A., Fiagbedzi, E., Amuasi, J. H., Flether, J. J., & Schandorf, C. (2019). Curr Trends Clin Med Imaging Image Noise in Radiography and Tomography: Causes, Effects and Reduction Techniques. *Current Trends in Clinical & Medical Imaging*, *3*(4), 86–91. https://doi.org/10.19080/CTCMI.2019.02.555620

Rahman, T., Khandakar, A., Qiblawey, Y., Tahir, A., Kiranyaz, S., Abul Kashem, S. Bin, Islam, M. T., Al Maadeed, S., Zughaier, S. M., Khan, M. S., & Chowdhury, M. E. H. (2021). Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images. *Computers in Biology and Medicine*, *132*, 104319. https://doi.org/10.1016/j.compbiomed.2021.104319

Sharma, H., Jain, J. S., Bansal, P., & Gupta, S. (2020). Feature extraction and classification of chest X-ray images using CNN to detect pneumonia. *Proceedings of the Confluence 2020 - 10th International Conference on Cloud Computing, Data Science and Engineering*, 227–231. https://doi.org/10.1109/CONFLUENCE47617.2020.9057809

Wang, Y., Chen, Y., Yang, N., Zheng, L., Dey, N., Ashour, A. S., Rajinikanth, V., Tavares, J. M. R. S., & Shi, F. (2019). Classification of mice hepatic granuloma microscopic images based on a deep convolutional neural network. *Applied Soft Computing*, *74*, 40–50. https://doi.org/10.1016/J.ASOC.2018.10.006

Yang, Y., & Mei, G. (2022). Pneumonia Recognition by Deep Learning: A Comparative Investigation. *Applied Sciences 2022, Vol. 12, Page 4334*, *12*(9), 4334. https://doi.org/10.3390/APP12094334