International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Investigating Deep Learning Techniques for Automization of X-Ray Image Classification in Healthcare

Hetal B Chauhan[1], Suresh B Patel[2]

Submitted: 17/09/2024 **Revised**: 20/10/2024 **Accepted**: 16/11/2024

Abstract: As a result of technological advancements, machine learning is emerging as a cutting-edge trend in healthcare that aids medical professionals in making decisions and enhances the precision of diagnoses. Investigators are seeking technological solutions to support healthcare providers in their daily tasks. X-ray images are most widely utilized diagnostic imaging methods in healthcare due to low cost and availability. The use of deep learning for the automatic classification of x-ray images can be extremely beneficial in situations where there is a inadequacy of available experts or can assist experts in obtaining a second opinion. The researchers investigated popular CNN architectures, including VGG, RESNET, DENSENET, EFFICIENTNET, and MOBILENET, and optimized them using two datasets of x-ray images. First dataset focuses on heart disease, whereas the second one pertains to breast cancer. Among all models, the MobileNet achieved the highest performance across both datasets when assessing metrics like accuracy, F1 score, recall, and precision..

Keywords: *Machine Learning; Healthcare; Deep Learning; Transfer Learning; Heart Disease (HD); Cancer.*

1. Introduction

X-ray images are most widely utilized diagnostic imaging methods in healthcare. Radiographic images are used to identify heart problems, cancers, lung diseases and injuries. X-ray machines are available easily to everywhere and less expensive compared to other medical imaging like MRI and CT scan. Identifying abnormalities in x-ray images requires specialized expertise. At times, these experts may not be readily available, which can cause delays in diagnosis. Additionally, suggestions from an automated system can assist experts in making decisions. Deep learning models are becoming more widely used for automatically identifying patterns in images without the need for manual feature extraction. As a result, CNN models can assist in the classification of x-ray images [1].

Convolutional Neural Networks (CNNs) are sophisticated multi-layer neural networks aimed at image recognition with minimal preprocessing requirements. They independently learn and identify features from images through convolutional layers (which utilize filters) and pooling layers (which decrease spatial dimensions), thereby avoiding the need of manual feature extraction [2]. As the architecture delves deeper, it recognizes intricate features — starting from simple edges in the initial layers and progressing to complex object shapes in the later layers. The primary elements of CNNs include: Convolutional layers which employ small kernels to examine sections of the image. Non-linear activation functions (ReLU) that address the non-linear characteristics of images. Pooling layers (such as max or mean pooling) that accommodate spatial variations and diminish the image dimensions. These components work in unison, allowing CNNs to attain high precision in applications such as image classification, object detection, and segmentation [3].

2. Literature Survey

Deep learning models have been proven important in detection of diseases using medical images. The research presented by Singh [4] addresses the difficulties associated with conventional medical image analysis and examines recent studies in the domain of medical image analysis that utilize deep learning and leverage large data platforms.

The researchers have explored different methods for diagnosing Covid-19 using chest radiology images. An effort done to predict Covid19 positivity from chest x-ray images using VGG19 and UNET[5]. UNET used for segmenting lungs from the image and removing unnecessary details. An ensemble model including decisions by multiple CNN models to detect Covid19 is proposed in the study done by Das and members[6]. They explored DenseNet201, Resnet50V2 and Inceptionv3 for prediction of COVID19 positive cases through chest radiology. The paper concluded that an ensemble model performed better than basic state of the art models. Ohata[7] and Narin[8] applied concept of transfer learning to detect Covid-19 using pretrained CNN models. Ohata used pre trained model to extract features which are then classified by machine learning classifiers such as k-Nearest

[1]*Information Technology Department, Government Engineering College, Gandhinagar, India*
[2]*Information Technology Department, Government Engineering College, Gandhinagar, India*
E-mail address: hetalchauhan@gecg28.ac.in, sbpatelit@gecg28.ac.in

Neighbor, Bayes, MLP, SVM and Random Forest. A DenseNet model, which was trained on extensive chest X-ray images excluding Covid19 cases, is used to extract features from chest radiology, and a regression method is employed to predict the severity score based on these extracted feature [9].

In the research[10], a significant step in the radiology process is automated through the implementation of three convolutional neural network (CNN) models, namely DenseNet121, ResNet50, and EfficientNetB1, to enable quick and precise identification of 14 distinct class labels related with thoracic pathology conditions based on chest X-rays. In relation to the score values achieved by each class in the dataset, DenseNet121 exceeded the performance of the other two models.

Stirenko et al. [11] showcased the efficacy of deep CNNs for computer-aided diagnosis of tuberculosis, especially by utilizing methods such as lung segmentation and both lossless and lossy data augmentation on a small and imbalanced dataset. Rahman and colleagues [12] conducted a study focused on creating a technique for detecting viral and bacterial pneumonia through the analysis of digital X-ray images. They provided an in-depth summary of existing methods for identifying pneumonia and detailed the particular techniques employed in their study.

Due to inadequacy of extensive public datasets in the medical field, the transfer learning approach is often utilized. However, other conditions such as heart issues, cancer, lung diseases, and injuries can also be identified through chest X-rays. The research work on the automation of diagnosing these diseases is limited when compared to the focus on COVID-19 and pneumonia. There has been a lack of thorough research regarding the capability of deep learning models to accurately identify rare or infrequent diseases through chest X-rays. We are looking to investigate CNN models and the principle of transfer learning to automate the diagnosis of heart disease and breast cancer using X-ray images. In order to better assess the effectiveness of these models, it is essential to use more thorough evaluation techniques such as sensitivity, specificity, and F1 score.

3. Material and Method

3.1 Dataset

In this study, used 2 datasets for the experiment. First dataset is for heart disease which includes defects involving heart vessels/wall. Three types of defects are included namely, PDA(patent ductus arteriosus), VSD(ventricular septal defect) atrial septal defect (ASD). Number of samples are 208, 194, 216 and 210 for normal, ASD, PDA, and VSD classes, respectively. Second dataset is for breast cancer having 2 classes. Number of images presenting cancer is 125 and non cancer is 620. Both datasets are publically available. Heart disease dataset is accessible from https://www.kaggle.com/competitions/congenital-heart-disease [13]. Cancer dataset is available on https://data.mendeley.com/datasets/fvjhtskg93/1 [14]. Number of images in the datasets are denoted in Table 1 and Table 2

Table 1: Number of Samples in Heart Disease Dataset

Class	Samples
Normal	208
ASD	194
PDA	216
VSD	210

Table 2: Number of samples in Cancer Disease Dataset

Class	Samples
Cancer	125
Non Cancer	620

3.2 Data Preprocessing

The model incorporates image augmentation techniques, including shear, zoom, rotation, and horizontal flips, to artificially enlarge the training dataset, which enhances generalization and minimizes overfitting [15]. To tackle the issue of low-quality X-ray images, adaptive histogram equalization (AHE) is utilized, which improves local contrast more effectively than global approaches[16]; The AHE clip limit is set at 0.03. During preprocessing, pixel values are normalized (scaled between 0 and 1 by dividing by 255), and images are resized to dimensions of 224×224 pixels. The dataset is divided, allocating 80% for training and 20% for testing to construct the model.

3.3 Model Development

The pre-trained models VGG, RESNET, DENSENET, EFFICIENTNET, and MOBILENET on IMAGENet[17] dataset were retrained using transfer learning for our diagnostic purpose. The final fully connected layers of the pre-trained models were substituted with a flatten layer, succeeded by two dense layers containing 512 and 256 nodes, respectively, utilizing the ReLU activation function. A concluding dense layer featuring 4 units with a softmax function was utilized for the heart disease diagnosis task, while a dense layer with 2 units followed by a sigmoid function was employed for cancer detection. The training parameters were kept consistent for all models across both datasets. Each model was trained for 50 epochs. The categorical_crossentropy loss function employed for multiclass classification, which pertains to heart disease diagnosis, whereas binary_crossentropy was utilized for binary classification in the cancer diagnosis task. An abstract of the training parameters can be found in the Table 3.

Table 3: Model Training Parameters

Parameter	Value
Learning rate	0.001
Optimizer	Adam
Batch size	64
Number of epochs	50
Loss Function	categorical_crossentropy (Heart disease dataset) binary_crossentropy (Cancer dataset)

4. Result and Discussion

4.1 Evaluation Metrics

Performances of models are evaluated using evaluation criteria like precision, F1-score, accuracy, and sensitivity. Accuracy of the model is the percentage of accurate predictions it makes out of all the inputs. Precision evaluates the number of true positive predictions compared to the total number of positive predictions made by the model. Sensitivity (Recall) quantifies how many accurate positive predictions there are from all real positive input cases. The F1 Score represents the harmonic mean of both precision and recall. A high F1 Score suggests that the model strikes an effective balance between recall and precision [18].

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

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})} \quad \textbf{(2)}$$

$$\text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \quad \textbf{(3)}$$

$$\text{F1 Score} = \frac{2 * \text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}} \quad \textbf{(4)}$$

4.2 Result

Table 4 provides a summary of the outcomes of the models trained on the heart disease dataset that indicates the precision, recall, and F1 score values for all classes predicted by the different models. VGG yields very disappointing results, recording zero scores across all metrics for classes 0, 1, and 3. Class 2 is the only one that shows some success, achieving an F1-score of 0.42. EfficientNet displays variable and low performance, with F1-scores fluctuating between 0.00 and 0.37, and it completely fails to predict classes 2 and 3. DenseNet performs slightly better, particularly in class 0 with an F1-score of 0.39, but still faces challenges across all classes, exhibiting low recall in classes 1 and 3. ResNet shows a more balanced and moderate performance, attaining highest F1-score of 0.58 in class 1 along with reasonable outcomes in the other classes. MobileNet stands out as the top-performing model overall, demonstrating relatively high precision and recall in classes 0 and 1, with F1-scores of 0.60 and 0.65 respectively. However, it does struggle in class 2, where it records an F1 score of 0.21, while performing adequately in class 3 with an F1 score of 0.51. Performances of models for binary classification is noted in table 5. Outcome of VGG and EfficientNet is poor. Outcome of ResNet and DenseNet is acceptable. MobileNet outperformed all models with 0.98, 0.99 and 0.98 values of precision, recall and F1-score, respectively.

Table 4: Values of evaluation metrics for HD dataset

Model	Class	Precision	Recall	F1-score
VGG	0	0	0	0
	1	0	0	0
	2	0.26	1	0.42
	3	0	0	0
EfficientNet	0	0.18	0.13	0.15
	1	0.24	0.78	0.37
	2	0	0	0
	3	0	0	0
DenseNet	0	0.29	0.58	0.39
	1	0.67	0.1	0.17
	2	0.17	0.33	0.23
	3	0.67	0.05	0.09
ResNet	0	0.39	0.34	0.37
	1	0.41	1	0.58
	2	0.57	0.28	0.38
	3	0.78	0.17	0.27
MobileNet	0	0.5	0.76	0.6
	1	0.51	0.9	0.65

Model	Class	Precision	Recall	F1-score
	2	1	0.12	0.21
	3	0.64	0.43	0.51

Table 5: Values of evaluation metrics for Cancer dataset

Model	Class	Precision	Recall	F1-score
VGG	0	0	0	0
	1	0.83	1	0.91
EfficientNet	0	0.2	1	0.34
	1	1	0.22	0.36
DenseNet	0	0.74	1	0.85
	1	1	0.93	0.96
ResNet	0	0.95	0.8	0.87
	1	0.96	0.99	0.98
MobileNet	0	0.96	0.88	0.92
	1	0.98	0.99	0.98

4.3 Performance Comparison and Discussion

Pictorial representations of F1 scores by all the models applied to the heart disease and cancer datasets are shown in figure 1 and 2, respectively. According to figure 1, MobileNet attained the highest F1 scores of 60, 51, and 65 for the ASD, VSD, and normal classes in the heart disease dataset. For the PDA class, the highest F1 score observed was 42, achieved with the VGG model. In the cancer dataset, MobileNet recorded the highest F1 score of 92 for the cancer class as demonstrated in figure2. Additionally, both ResNet and MobileNet achieved the same F1 score of 98 for the non-cancer class. Reviewing the accuracy value in table 2, it shows that MobileNet outperformed the other models across both datasets.. ResNet and MobileNet are best performers for both datasets. DenseNet is slightly less effective than ResNet/MobileNet. EfficientNet and VGG struggle with class balance and consistency, making them less reliable for these tasks. Both models overfit or fail to generalize properly.

In binary classification, the performance of CNN models is outstanding. In contrast, for multiclass classification, the performance tends to be less effective. Convolutional Neural Networks (CNNs) generally excel at binary classification rather than multi-class classification because binary classification, which aims to categorize data into one of two categories, is typically easier to model than multi-class classification. This is due to the fact that the decision boundary separating the two classes is simpler to understand and establish, unlike the intricate decision boundaries required to differentiate among several classes in a multi-class situation.Multi-class classification requires the development of several decision boundaries to distinguish each class from the others which increases the complexity of the model and may result in over fitting, particularly when training data is limited.

Table 6: Prediction accuracy of models

Model	Cancer	Heart disease
VGG	83	26
EfficientNet	35	23
DenseNet	94	26
ResNet	96	45
Mobilenet	97	54

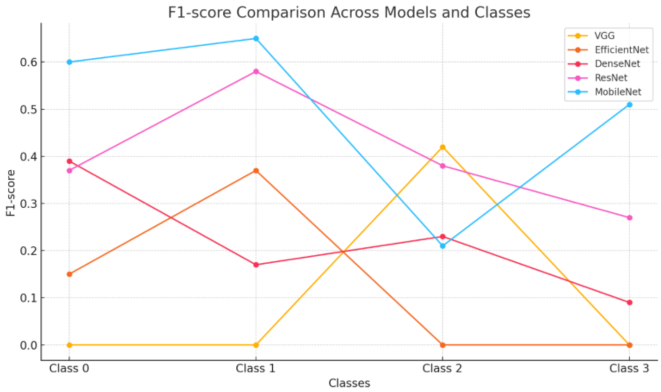

Figure 1: F1 scores for different classes on HD datset

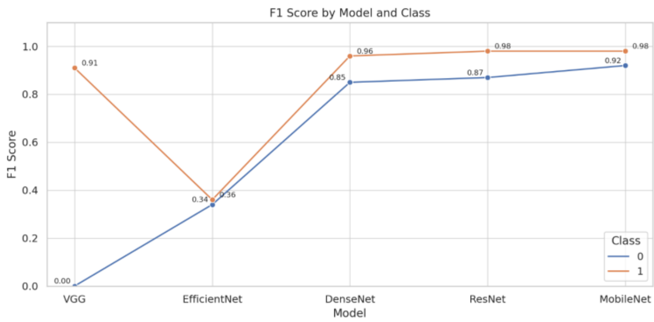

Figure 2: F1 Scores for Different Classes on Score on Cancer Dataset.

5. Conclusion

The primary objective of this study was to automate disease identification through X-ray images by leveraging various deep learning models and evaluating their performance based on different parameters. There has been insufficient research on the ability of deep learning models to accurately detect rare or uncommon diseases through chest X-rays. A key aspect of our study includes the prediction of heart defects and breast cancer. To accomplish this, the researchers employed well-known CNN architectures like DenseNet, ResNet, VGG, EfficientNet, and MobileNet by implementing the transfer learning approach. ResNet and MobileNet emerged as the top performers for both datasets. When available data is limited, applying transfer learning by beginning with a pretrained model and subsequently fine-tuning it on your restricted dataset can prove advantageous. Furthermore, CNN models typically perform better in binary classification scenarios than in multiclass classification. Multiclass classification requires establishing multiple decision boundaries to distinguish between each category, which increases the model's complexity and may lead to overfitting, especially when the training dataset is not extensive. Nonetheless, the developed models can aid both professionals in making informed decisions and individuals without access to expert guidance.

6. Future Scope

Pretrained models that are optimized for ImageNet often face challenges when applied to areas like medical imaging. Domain adaptation techniques adjust these pretrained models to improve their performance in new domains, often using unsupervised or semi-supervised datasets. More advanced adaptation methods can be explored to narrow the domain gap and mitigate negative transfer. Instead of conventional models such as VGG, DenseNet, ResNet, models based on Vision Transformers (ViT), and hybrid CNN-Transformer architectures, which are trained on wider and more diverse datasets can show improved generalization across various tasks. Future research might focus on leveraging more extensive pretrained models beyond typical CNNs to enhance performance in downstream applications.

References

[1] Mall, P. K., Singh, P. K., Srivastav, S., Narayan, V., Paprzycki, M., Jaworska, T., &Ganzha, M. (2023). A comprehensive review of deep neural networks for medical image processing: Recent developments and future opportunities. Healthcare Analytics, 4, 100216.

[2] Taye, M. M. (2023). Theoretical understanding of convolutional neural network: Concepts, architectures, applications, future directions. Computation, 11(3), 52.

[3] Rahman, T., Chowdhury, M. E., Khandakar, A., Islam, K. R., Islam, K. F., Mahbub, Z. B., ...&Kashem, S. (2020). Transfer learning with deep convolutional neural network (CNN) for pneumonia detection using chest X-ray. Applied Sciences, 10(9), 3233.

[4] Singh, Y., Tripathi, N., Yadav, S., Gupta, N., Kumar, A. U., & Ramesh, J. V. N. (2024). Transfer Learning and Chest X-ray-Based Image Processing and Modeling to Detect COVID-19. In Smart Technologies in Healthcare Management (pp. 240-263). CRC Press.

[5] Arias-Garzón, D., Alzate-Grisales, J. A., Orozco-Arias, S., Arteaga-Arteaga, H. B., Bravo-Ortiz, M. A., Mora-Rubio, A., ...&Tabares-Soto, R. (2021). COVID-19 detection in X-ray images using convolutional neural networks. Machine Learning with Applications, 6, 100138.

[6] Das, A. K., Ghosh, S., Thunder, S., Dutta, R., Agarwal, S., &Chakrabarti, A. (2021). Automatic COVID-19 detection from X-ray images using ensemble learning with convolutional neural network. Pattern Analysis and Applications, 24, 1111-1124.

[7] Ohata, E. F., Bezerra, G. M., das Chagas, J. V. S., Neto, A. V. L., Albuquerque, A. B., De Albuquerque, V. H. C., &ReboucasFilho, P. P. (2020). Automatic detection of COVID-19 infection using chest X-ray images through transfer learning. IEEE/CAA Journal of AutomaticaSinica, 8(1), 239-248.

[8] Narin, A., Kaya, C., &Pamuk, Z. (2021). Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. Pattern Analysis and Applications, 24, 1207-1220.

[9] Cohen, J. P., Dao, L., Roth, K., Morrison, P., Bengio, Y., Abbasi, A. F., ...& Duong, T. (2020). Predicting covid-19 pneumonia severity on chest x-ray with deep learning. Cureus, 12(7).

[10] Mann, M., Badoni, R. P., Soni, H., Al-Shehri, M., Kaushik, A. C., & Wei, D. Q. (2023). Utilization of deep convolutional neural networks for accurate chest X-ray diagnosis and disease detection. Interdisciplinary Sciences: Computational Life Sciences, 15(3), 374-392.

[11] Stirenko, S., Kochura, Y., Alienin, O., Rokovyi, O., Gordienko, Y., Gang, P., & Zeng, W. (2018, April). Chest X-ray analysis of tuberculosis by deep learning with segmentation and augmentation. In 2018 IEEE 38th International Conference on Electronics and Nanotechnology (ELNANO) (pp. 422-428). IEEE.

[12] Rahman, T., Chowdhury, M. E., Khandakar, A., Islam, K. R., Islam, K. F., Mahbub, Z. B., ...&Kashem, S. (2020). Transfer learning with deep convolutional neural network (CNN) for pneumonia detection using chest X-ray. Applied Sciences, 10(9), 3233.

[13] https://www.kaggle.com/competitions/congenital-heart-disease

[14] https://data.mendeley.com/datasets/fvjhtskg93/1

[15] Shorten, C., &Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. Journal of big data, 6(1), 1-48.

[16] Nia, S. N., & Shih, F. Y. (2024). Medical X-ray image enhancement using global contrast-limited adaptive histogram equalization. International Journal of Pattern Recognition and Artificial Intelligence, 38(12), 2457010.

[17] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., &Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition (pp. 248-255). IEEE.

[18] Chauhan, H., &Modi, K. (2023). AMSFMap Methodology to improve prediction accuracy of CNN model for Covid19 using X-ray images. Procedia Computer Science, 218, 1394-1404.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.