Medical Image Reader powered by Artificial Intelligence

Tanvir Ahmed Palok
Department of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
tanvirahamedx2z@gmail.com

Symum Ahmed
Department of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
symumahmed21@gmail.com

Golam Kibria Anim
Department of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
g.k.anim.tp1999@gmail.com

Shahrear Alam
Department of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
marufrayan31.1.00@gmail.com

Shahed Sharif Bhuiyan Ratul
Department of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
ratul831@gmail.com

Nabuat Zaman Nahim
Department of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
nabuat.zaman@bracu.ac.bd

***Abstract*—Misdiagnosis in medical imaging is a critical concern, risking patients' health due to the pivotal role of radiologists' accuracy in diagnostics. Current cross-checking methods for radiologists' decisions are limited, potentially leading to errors and treatment delays. This study introduces a data processing technique and an advanced prediction system for improving disease detection accuracy in medical images. Our main goal is to contribute to healthcare by developing a system capable of achieving human-level or higher accuracy in disease detection across diverse medical image types. To achieve this, we utilize deep learning techniques, specifically Convolutional Neural Networks (CNNs), and leverage Transfer Learning with pre-trained models. Data processing plays a crucial role, given the importance of image availability and quality. We apply image enhancement techniques such as Histogram Equalization, Adaptive Histogram Equalization (AHE), and Contrast Limited Adaptive Histogram Equalization (CLAHE) to enhance image quality and augment a limited training dataset. The advanced ensemble approach significantly enhances the overall accuracy and reduces individual model variance. Validation of our approach using confusion matrices reveals that selective class-wise voting achieves the highest accuracy at 95.27% on the testing dataset. Additionally, our customized weighted voting approach achieves an accuracy of 94.07% on the test set. These results emphasize the effectiveness of our ensemble techniques in improving disease detection accuracy. Our ensemble techniques offer substantial accuracy improvements, promising more accurate and reliable medical diagnoses.**

***Index Terms*—Misdiagnosis, Deep learning, Ensemble learning, Confusion matrices, Selective class-wise voting, Histogram equalization, AHE, CLAHE, Transfer learning.**

I. Introduction

A. Background

With the advancement of science and technology, the medical sector has too become highly developed in the modern era. Broadly, medical sectors can be divided into three correlated parts, they are- diagnosis, analysis and treatment. After being concerned about certain symptoms of one's physical or mental health, he or she visits a doctor. The doctor typically advises some medical tests or medical images. A medical image is a visual representation of the internal body or organs. Each type of medial image machine uses different types of imaging techniques. Hence, in order to find out the appropriate inner visuals, doctors need a certain type of medical image for diagnosing certain diseases or medical problems. Sadly, this process will not be effective at all, if the diagnosis goes wrong at the very beginning. In many cases, misdiagnosis pushes a life towards death in the end. The study [1] by Seigal et al. (2017) highlights that 1325 claims out of 29,777 medical malpractice filed cases had 'Radiology' as the 'Primary Responsible Service' between the years 2010 and 2014. After a rigorous review, they found that 42% of the claims resulted in high severity clinical injuries including 235 deaths. These numbers are based on a 5-years study over 25 states of the United States of America only. So, it is unimaginable what effect this error from the radiology department has been creating since a very long time. The interpretations from the medical images are made by radiologists by using just their seeing capability. Sometimes, the professional radiologists oversee a tiny nodule inside the organ which can eventually result in malignancy. The images contain many details that it is not impossible for even a professional to omit this mistake. In an X-Ray, the black and white intensity can help to deduce what is the severity of the danger, a small alien structure in the brain image can easily remain unnoticed by a human vision. A mistake made in the first phase of the disease can lead to a very dangerous case eventually.

B. Research Problem

Every life in this world has the right to live in good health, so it is not expected that a tiny mistake will violate or sabotage some of the lives. Therefore, if a system can be built which will work between the radiologists and doctors in order to cross-validate the decisions made by the radiologist, it will be able to send a comparison report to the doctor. Finally, the doctor can see both the radiologist's and system's prediction over the sample. Thus, being quite sure of the disease, the doctor can provide proper treatment to the patient from the earliest phase of diagnosis.

II. Literature Review

Yimer et al. researched [2] to improve the treatment of lung diseases from CXR images using an AI-based multi-class classification technique. The results showed that the Median filter effectively removed noise without compromising edge preservation. The CLAHE method performed better than global histogram equalization and AHE in enhancing image contrast, which is why they chose it as their study's processing technique.

The study [3] by Hussein et al. created a hybrid architecture of CLAHE and CNN, outperforming traditional methods by roughly 20% accuracy. By evaluating three different CNNs for classifying lung diseases from CXR images, they discovered that the model can classify three types of lung diseases, including COVID-19, pneumonia and tuberculosis.

Taresh et al. presented a study [4] on using transfer learning to automatically train CNN models to detect COVID-19 from CXR images. To do so, the authors used a 5,000 chest X-ray images dataset to train and evaluate three different trained CNNs: VGG16, ResNet50, and MobileNet, and they also applied Transfer Learning. The authors believe the study's results are still substantial and provide evidence that transfer learning can be used to train effective CNNs for COVID-19 detection.

Khan et al. portrayed a study [5] that proposed a DL model based on the Xception architecture for detecting COVID-19 cases from CXR images. CoroNet demonstrated higher accuracy in classification tasks than CovidNet, VGG19 and other CNN models. In summary, the proposed CoroNet model based on the Xception architecture showed prominent performance in detecting COVID-19 cases from chest X-ray images, outperforming other state-of-the-art deep learning models.

Another study regarding this matter by Rajpurkar et al. represents a CNN named CheXNet [6], which outperformed radiologists on this task, with an F1 score of 0.435 (95% CI 0.387, 0.481). The authors also extended CheXNet to classify 14 thoracic diseases and achieved state-of-the-art results on all 14 classes. CheXNet can accurately identify many thoracic problems, including pneumonia, consolidation, effusion, emphysema, and fibrosis.

Alsubai et al. chose DL models, CNN and CNN-LSTM, to apply their proposed technique for brain tumor classification [7]. On the other hand, the hybrid CNN-LSTM model performed even better, with an accuracy of 99.1%, precision of 98.8%, recall of 98.9%, and F1-measure of 99.0%. In summary, the proposed CNN-LSTM model exhibits outstanding results in brain tumor classification, outperforming previous techniques and demonstrating its potential for accurate and efficient diagnosis.

Khan et al. [8] proposes two deep-learning models for brain tumor classification. The" Fine-tuned CNN with VGG16" achieved 100% accuracy on dataset 2, although an overfitting issue was seen. The proposed models for binary and multi-class tumor classification outperformed existing state-of-the-art methods.

Noreen et al. [9] used DL models, specifically DenseNet and Inception-v3 architectures and presented a study on the classification of brain tumors. The results of this study demonstrate that a combination of features extracted from various layers of Inception-v3 and DenseNet201 significantly improves classification accuracy compared to individual block feature extraction methods. The study further analyzes feature maps and discusses the challenges of classifying brain tumors due to their diverse shapes, sizes, and positions within MR images.

Yildirim et al. presented in their study [10] that a DL model was developed to detect kidney stones from CT images of the abdomen. The model successfully saw kidney stones in most cases. However, some images were misclassified due to factors like the presence of rib tips, calcified areas, and other organs in the image. The authors expect future work to involve collecting images from different sources to validate the model's performance in diverse settings.

The study conducted by Shakeel et al. [11] presented that the Improved Profuse Clustering Technique (IPCT) and Deep Learning with Instantaneously trained neural networks (DITNN) can improve lung cancer detection and classification systems using CT images. They tested both approaches regarding various segmentation metrics, outperforming other segmentation methods like fuzzy c-means, global thresholding, watershed, and Sobel. This study's outcome suggests that the DITNN approach is highly reliable for lung cancer detection and classification.

Islam et al. [12] reported that the Swin Transformer can become one of the most effective options for diagnosing kidney diseases from CT images. The Swin Transformer gained the highest recall for kidney cyst, normal, stone, and tumor class images, with values of 0.996, 0.981, 0.989, and 1, respectively. The Swin Transformer focused on small regions of interest, leading to more accurate predictions than others.

Gharaibeh et al. [13] DL has shown promising results compared to traditional ML methods in classifying kidney tumors from CT images. Their segmentation studies focused on detecting the tumor nodules within the kidney, and they achieved accuracy rates of 97.7% and 96.9% using V-Net and 3D U-Net, both based on 210 CT scans. The authors expect future works based on this work to include trials using smaller models and properly auguring the images.

The research [14] done by Saeedi et al. tried to develop two DL networks and six machine learning techniques for classifying MRI images into brain tumor categories: glioma, meningioma, pituitary gland tumor, and healthy brain. Comparisons were made with studies that employed other neural networks for tumor classification, and the proposed 2D CNN performed better, achieving higher accuracy with 96.47%. However, a limitation was the small size of medical image databases, which restricted the availability for training deep neural networks.

Mohsen et al. presented that the performance evaluation of their proposed methodology [15] was based on a comprehensive set of evaluation metrics, including average classification rate, average recall, average precision, average F-Measure, and average area under the ROC curve (AUC). Compared with other classifiers, the DNN classifier emerged as the most prominent performer, exhibiting excellence across all measured performance criteria. On the other hand, KNN and SMO played relatively lower performance in the given evaluation metrics. Through the model: DNN classifier in accurately classifying brain MRI images, resulting in ease at further research and detection of diseases.

Gaur et al. highlighted an in-depth comparison [16] of methodologies and state-of-the-art models for analyzing brain MRI images and classifying brain tumors. The comparison shows the effectiveness of DL-based approaches, particularly CNNs, when combined with advanced techniques like PCA and KSVM, improving accuracy and reducing computation time in brain tumor diagnosis. By prioritizing the explainable AI, they used SHAP and LIME methods to visualize the learning of CNN models.

Kumar et al.discussed in their study [17] that the development of an Android app allowing users to predict diseases and view disease trends was named" Disease Prediction using Artificial Intelligence" (DPAI). The model they used outperformed all the existing models by 1.2746% in terms of accuracy and 1.3926% in terms of F-measure. They evaluated their proposed model's performance against existing ML models on binary classification problems for predicting diabetes, heart disease, and COVID-19.

III. Research Methodology

A. Dataset Description

For our research, we have collected various datasets with 3 different types of medical images and they are- X-ray, MRI, CT Image. We collected these datasets from publicly accessible online sources. Combining all of them, we have generated a unique dataset containing 18 classes including both the patients' and normal ones. We named this dataset *mir18*. We split all the images into a ratio of 70:30 for training and validation purposes respectively. In summary, *mir18* dataset contains 18 classes and 34,178 training samples. Notably, *mir18* is a highly imbalanced dataset. Notably, among the 18 classes, 13 of them pertain to various diseases, while the remaining 5 classes describe the normal condition of the organs- Brain, Lungs, Kidney, and Knee.

B. Data Processing

The number of training samples in *mir18* exhibits a notable class imbalance, with some classes having under 1,000 instances, while another class surpasses 7,134 instances. This imbalance should be carefully managed when developing machine learning models to ensure robust performance. Methodically, this type of data cannot leverage the training of our models. Besides, the model will encounter overfitting as well as cannot classify those classes with low numbers of samples with a satisfying accuracy. Due to this reason, we have applied a combination of different Image Processing Techniques (Histogram Equalization, AHE, CLAHE) and Normalization to the classes under 1000 samples to expand the number of samples in them to 1000 samples per class. In this method, for every original image we get 4 different processed images using different techniques. The motivation behind applying this method is generally the ability of Image Processing Techniques that can create new variant of those samples which may derive new features that can help the model to learn new patterns. In the next stage, we applied a series of augmentation techniques including rotation, gaussian filter, scale, shear, jitter, flipping, sharpening on the train data to make them expand to a certain threshold (4,000). This method will introduce a large number of images that may assist the model in learning new patterns but may not be as useful as the first stage we applied. Finally, we resized all the images to a dimension of (224, 224, 3) with a batch size of 32. Applying this data processing, we got another set of training samples containing a total of 74,366 images in 18 classes. We named this dataset as *mir18_v2*. This new dataset had less imbalance among the classes. Fig. 1 portrays the Data Processing method.

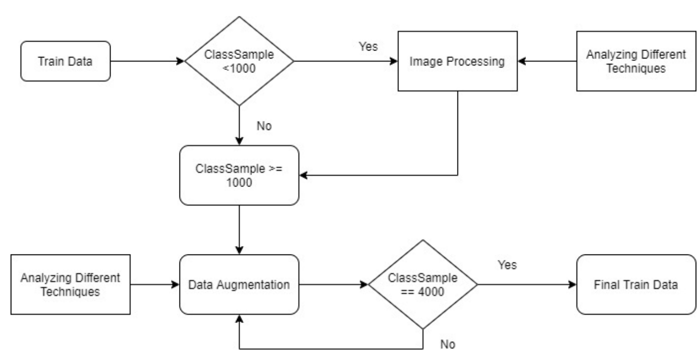


Fig. 1: Flow-chart of the Data Processing Method

C. Models

After a thorough research, we picked several pre-trained models and added 2 convolutional layers and a dropout of 0.5 onto them. In our study, we have applied several pre-trained models such as: VGG16, VGG19, DenseNet121, ResNet50, EfficientNetV2S, InceptionResNetV2. We used all the models' stored weights from imagenet and trained them using both of our datasets. After checking the predictions, we finally picked two of them. These models are- EfficientNetV2S and InceptionResnetV2. While using these models, we followed two approaches. First approach being the typical keeping the base layers of the models frozen and in the other approach we kept the layers trainable. In some cases, we found that keeping the layers trainable had a good impact on both the training accuracy and the cross-validations. Also, it helped in keeping the validation loss very low in numbers. Hence, we finally came up with two approaches of training using two CNN models on two different datasets. Table I portrays the overview of all our trials and the naming convention for better understanding of the Result Analysis section.

TABLE I: Overview of the models.

Model #	***Pre-trained Model***	***Dataset used***	***Base Layer***
Model 1	EfficientNetV2S	*mir18_v2*	Trainable
Model 2	EfficientNetV2S	*mir18_v3*	Frozen
Model 3	InceptionResNetV2	*mir18_v2*	Trainable
Model 4	InceptionResNetV2	*mir18_v3*	Trainable
Model 5	EfficientNetV2S	*mir18*	Trainable
Model 6	InceptionResNetV2	*mir18*	Trainable
Model 7	EfficientNetV2S	*mir18*	Frozen

Here, *mir18_v3* refers to another dataset that we produced by applying another variant of image processing techniques which was specialized for only x-ray images (other two types of images were still processed by the previously mentioned combination). However, we dropped this dataset later because it was not contributing sufficient enough in our study. Therefore, we finally picked Model 1, 3, 5, 6 and 7 from the Table I to carry our research on.

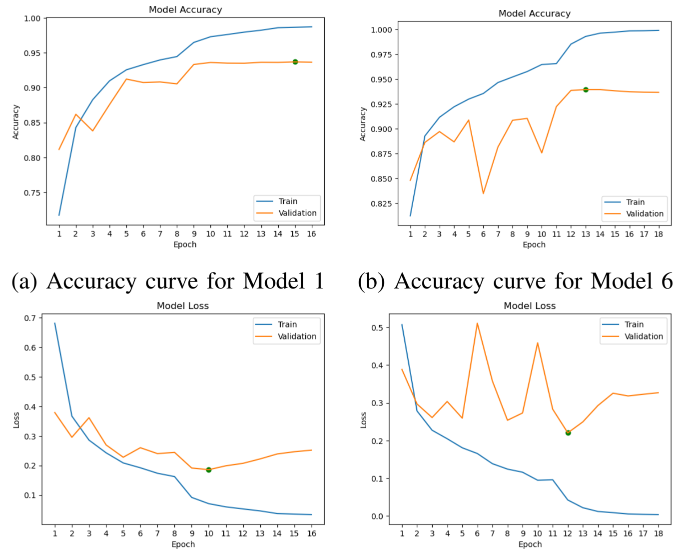
D. Training and Testing

Our research is carried on using Python 3.9.13 in Jupyter Notebook 6.4.12. and Tensorflow v2.10.0 on a system consisting Intel(R) Core(TM) i5-8400 CPU @ 2.80GHz with 16 GB RAM. We used NVIDIA Gefore RTX 3060 GPU for an efficient training purpose.

For testing, it is made sure that a set of unseen samples has each time been used so that we can gain a perfect prediction result. Hence, we have split the validation set into two halves and kept one half as our validation set and the other for only testing purpose. Fig. 2 represents the accuracy and loss curves of both our pre-trained models respectively while keeping both of their base layers trainable.

E. ***Result Analysis***

1) Performance of Models: The main purpose of our approaches is to gain better accuracy on all the 18 classes.



(a) Accuracy curve for Model 1 (b) Accuracy curve for Model 6

(c) Loss curve for Model 1 (d) Loss curve for Model 6

Fig. 2: Curve Analysis of Model 1 and Model 6

In the beginning, we trained all the selected models on the *mir18* dataset. This dataset was highly imbalanced in a sense that it was a mix of multiple datasets from multiple sources available online. Besides, some types of images were very rare to collect, for example: 'Benign Lung Cancer' has only 120 samples on the whole dataset. In this case, we only used 84 samples for training, 18 for validation and 18 for testing purposes. After training on the base image we get 3 best model including EfficientNetV2S (keeping base layers frozen), EfficientNetV2S (keeping base layers trainable), InceptionResnetV2 (trainable) with a test accuracy of 88.78%, 92.68%, 93.77%. But the main drawback here is that, from the beginning of the training we have encountered overfitting. We have applied early stopping to lessen the overfitting and improve generalization of the model during the training process. However, due to the highly imbalanced class distribution of our dataset, the overfitting issue still remains. The training accuracies were 92.07%, 96.75%, 99.28% whereas the validation accuracies were 89.63%, 92.96%, 93.93% respectively. In order to overcome overfitting for a highly imbalanced dataset, we have applied the data processing techniques in a different way.

TABLE II: A comparison chart for understanding the performances of the models before and after applying our data processing techniques.

Model	***mir18*** **(%)**			***mir18_v2*** **(%)**		
-	Train	Valid	Test	Train	Valid	Test
DenseNet121	90.9	87.67	85.79	94.14	88.05	87.37
EfficientNetV2S	92.07	89.63	88.78	96.3	91.58	91.43
EfficientNetV2S*	96.75	92.96	92.68	98.64	93.69	93.66
InceptionResNetV2*	99.28	93.93	93.77	98.66	94.11	92.26
VGG19	92.77	87.03	86.41	96.71	89.65	88.97

2) Confusion Matrix: In our study, we have trained multiple models with *mir18* and *mir18_v2* datasets to find an optimal solution to our problem. In order to analyze their result, we carefully examined the confusion matrices generated from the predictions of these models to determine which ones were more effective at classifying maximum classes. Confusion matrix is a vital tool for evaluating classification model performance. It breaks down predictions into four categories: true positives, true negatives, false positives, and false negatives. This technique enables the computation of key metrics like accuracy, precision, recall, and F1 score. Analyzing the confusion matrix helps refine models and enhances their decision-making capabilities. The diagonal values of the confusion matrix denote correct predictions: true positives are instances correctly classified as positive, and true negatives are instances correctly classified as negative.

3) Ensemble Learning: In order to gain a more satisfying prediction, the ensemble techniques have worked to achieve better accuracy. In terms of majority voting, the ensemble models have an accuracy of 93.75%. By applying "Sum Rule Ensemble", the accuracy increases to 94.02%. Next, we tried to predict using the "Mean Argmax" ensemble which gains an accuracy of 94.02% as well. After analyzing our used models' classification performance on all target classes, we have tried to implement a solution of our own.

We have introduced a new type of voting where according to the actual label we select the best model for prediction. This is a completely customized approach to show the best prediction of the combined models for each label. Though this approach will not give accurate predictions for a random sample or a new test set. Because of this, we have tried a new way where neglecting the actual label for selecting the model, we have used the prediction of our best model (maximum average accuracy) to select the best model for a particular class to maximize accuracy which may work on a random sample or a new test set. We named this technique *"Selective Class-wise Voting"*.

To introduce a better approach, we have applied a customized where predefined weight gives priority to the best model for voting. After analyzing the results of these models, we specify a weighted array of values for each model containing values from 0 to 7. For every class, based on the model performance we have assigned a weight. For every model, there is a weighted array for 18 classes. The combination of the weight is most important here. Now, for each random sample, every model will give its prediction. Based on the weight assigned for that predicted class the model's prediction and weight will be stored in a dictionary with the model name as a key. After that, we removed the model's prediction from the dictionary which has a 0 weighted value. Now iterating over that dictionary, different logic is set to get a better prediction. Firstly, there can be a set of 3 models or less with the best weight for a particular class. Besides, before removing the models from the dictionary we stored the best-performing model's output in a separate dictionary. If after removing the model with 0 weight in the specific class, the dictionary becomes empty then we will take the best-performing model's prediction. We have also used a reducer function to keep the models in the dictionary limited to 3. This method is most suitable for a random sample or a group of samples without any type of label assigned to it. This technique is named as *"Customized Weighted Voting"*.

Firstly, the *"Selective Class-wise Voting"* technique scored an accuracy of 95.27% overall and the second proposed method scored 93.6%. To overcome this problem and handle random samples or a set of random samples our proposed method of ensemble *"Customized Weighted Voting"* is introduced which has gained a better accuracy of 94.07%. Though the accuracy difference is very less, combining the best weight for the models per class can increase this accuracy which can outperform *"Selective Class-wise Voting"* technique (Actual label-wise selection of model) which gained an accuracy of 95.27%. Fig. 3 shows the confusion matrices after applying both these approaches.

4) Validity: Our custom data processing techniques have greatly improved training outcomes. Training and validation accuracy now show minimal differences from the start, thanks to our innovative approach. After applying our methods, most models have seen notable increases in both validation and testing accuracy. Specifically, the models EfficientNetV2S (frozen), EfficientNetV2S (trainable), and InceptionResnetV2 (trainable) achieved impressive test accuracies of 91.43%, 93.66%, and 92.26%, along with corresponding validation accuracies of 91.58%, 93.69%, and 94.11%, respectively. We've summarized the accuracy percentages for each class before and after data processing in Table I for reference. Of particular note is the improvement in classifying Benign Lung Cancer. Before our data processing approach, these models often struggled, but after implementation, they achieved significantly better results. In addition, when we tested DenseNet121 and VGG19, both initially classifying only a few samples in the *mir18* dataset, their accuracy increased substantially after our data processing technique was applied. Our analysis also suggests that EfficientNet-based models may struggle with smaller datasets compared to Inception-Resnet architecture models, which tend to perform better with limited data.

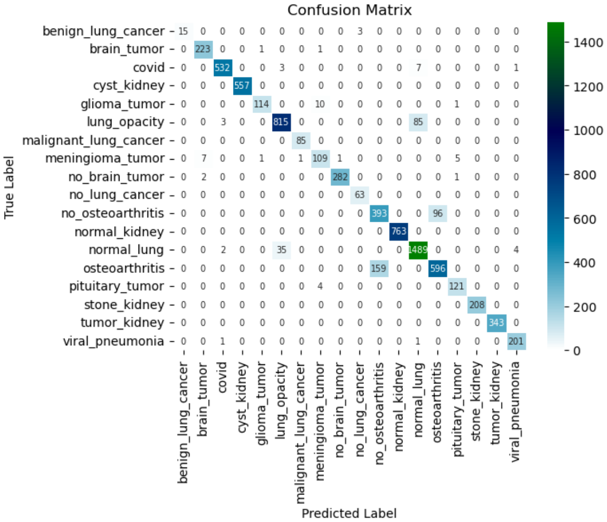
IV. Conclusion

A. Challenges

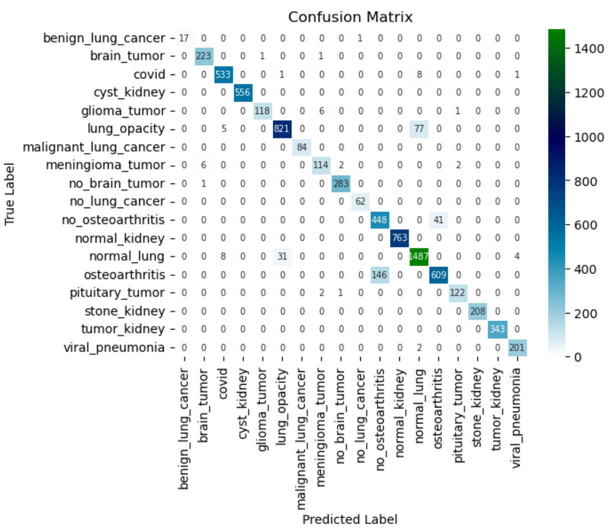
Our research aimed to improve disease classification using medical images with a unique data processing method and an effective ensemble technique. However, data scarcity and class imbalance posed challenges, leading to overfitting and difficulties in classifying diseases like Osteoarthritis. Future research should prioritize Explainable AI to enhance trust and transparency in AI-driven medical diagnosis for successful integration into healthcare practice.

B. Future Scope

Our aim is to enhance our methodology for more effective disease diagnosis by addressing data scarcity and overfitting issues, expanding the range of diseases classified, and improving



(a) Confusion Matrix from the Selective Class-wise Voting Technique



(b) Confusion Matrix from the Customized Weighted Voting Technique

Fig. 3: Analysis of Confusion Matrices

the precision of challenging classifications like 'Osteoarthritis' and 'No Osteoarthritis.' Future research can focus on refining data augmentation techniques for medical images, incorporating advanced transfer learning methods, integrating multiple data sources for a holistic diagnosis approach, developing real-time prediction systems for healthcare, and conducting rigorous clinical trials to ensure the safety and reliability of AI-assisted diagnosis in practical healthcare settings. These avenues promise to advance medical science significantly.

References

[1] D. Siegal, L. M. Stratchko, and C. DeRoo, "The role of radiology in diagnostic error: a medical malpractice claims review," Diagnosis, vol. 4, no. 3, pp. 125–131, Sep. 2017, doi: https://doi.org/10.1515/dx-2017-0025.

[2] F. Yimer, A. W. Tessema, and G. L. Simegn, "Multiple Lung Diseases Classification from Chest X- Ray Images using Deep Learning approach," International Journal of Advanced Trends in Computer Science and Engineering, vol. 10, no. 5, pp. 2936–2946, Oct. 2021, doi: https://doi.org/10.30534/ijatcse/2021/021052021.

[3] F. Hussein et al., "Hybrid CLAHE-CNN Deep Neural Networks for Classifying Lung Diseases from X-ray Acquisitions," Electronics, vol. 11, no. 19, p. 3075, Sep. 2022, doi: https://doi.org/10.3390/electronics11193075.

[4] M. M. Taresh, N. Zhu, T. A. A. Ali, A. S. Hameed, and M. L. Mutar, "Transfer Learning to Detect COVID-19 Automatically from X-Ray Images Using Convolutional Neural Networks," International Journal of Biomedical Imaging, vol. 2021, pp. 1–9, May 2021, doi: https://doi.org/10.1155/2021/8828404.

[5] A. I. Khan, J. L. Shah, and M. M. Bhat, "CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images," Computer Methods and Programs in Biomedicine, vol. 196, p. 105581, Nov. 2020, doi: https://doi.org/10.1016/j.cmpb.2020.105581.

[6] P. Rajpurkar et al., "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning," arXiv.org, 2017. https://arxiv.org/abs/1711.05225

[7] S. Alsubai, H. U. Khan, A. Alqahtani, M. Sha, S. Abbas, and U. G. Mohammad, "Ensemble deep learning for brain tumor detection," Frontiers in Computational Neuroscience, vol. 16, Sep. 2022, doi: https://doi.org/10.3389/fncom.2022.1005617.

[8] Md. S. I. Khan et al., "Accurate brain tumor detection using deep convolutional neural network," Computational and Structural Biotechnology Journal, vol. 20, pp. 4733–4745, Aug. 2022, doi: https://doi.org/10.1016/j.csbj.2022.08.039.

[9] N. Noreen, S. Palaniappan, A. Qayyum, I. Ahmad, M. Imran, and M. Shoaib, "A Deep Learning Model Based on Concatenation Approach for the Diagnosis of Brain Tumor," IEEE Access, vol. 8, pp. 55135–55144, 2020, doi: https://doi.org/10.1109/access.2020.2978629.

[10] K. Yildirim, P. G. Bozdag, M. Talo, O. Yildirim, M. Karabatak, and U. Rajendra. Acharya, "Deep learning model for automated kidney stone detection using coronal CT images," Computers in Biology and Medicine, vol. 135, p. 104569, Aug. 2021, doi: https://doi.org/10.1016/j.compbiomed.2021.104569.

[11] P. M. Shakeel, M. A. Burhanuddin, and M. I. Desa, "Lung cancer detection from CT image using improved profuse clustering and deep learning instantaneously trained neural networks," Measurement, vol. 145, pp. 702–712, Oct. 2019, doi: https://doi.org/10.1016/j.measurement.2019.05.027.

[12] M. N. Islam, M. Hasan, Md. K. Hossain, Md. G. R. Alam, M. Z. Uddin, and A. Soylu, "Vision transformer and explainable transfer learning models for auto detection of kidney cyst, stone and tumor from CT-radiography," Scientific Reports, vol. 12, no. 1, Jul. 2022, doi: https://doi.org/10.1038/s41598-022-15634-4.

[13] M. Gharaibeh et al., "Radiology Imaging Scans for Early Diagnosis of Kidney Tumors: A Review of Data Analytics-Based Machine Learning and Deep Learning Approaches," Big Data and Cognitive Computing, vol. 6, no. 1, p. 29, Mar. 2022, doi: https://doi.org/10.3390/bdcc6010029.

[14] S. Saeedi, S. Rezayi, H. Keshavarz, and S. R. Niakan Kalhori, "MRI-based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques," BMC Medical Informatics and Decision Making, vol. 23, no. 1, Jan. 2023, doi: https://doi.org/10.1186/s12911-023-02114-6.

[15] H. Mohsen, E.-S. A. El-Dahshan, E.-S. M. El-Horbaty, and A.-B. M. Salem, "Classification using deep learning neural networks for brain tumors," Future Computing and Informatics Journal, vol. 3, no. 1, pp. 68–71, Jun. 2018, doi: https://doi.org/10.1016/j.fcij.2017.12.001.

[16] L. Gaur, M. Bhandari, T. Razdan, S. Mallik, and Z. Zhao, "Explanation-Driven Deep Learning Model for Prediction of Brain Tumour Status Using MRI Image Data," Frontiers in Genetics, vol. 13, Mar. 2022, doi: https://doi.org/10.3389/fgene.2022.822666.

[17] N. Kumar, N. Narayan Das, D. Gupta, K. Gupta, and J. Bindra, "Efficient Automated Disease Diagnosis Using Machine Learning Models," Journal of Healthcare Engineering, vol. 2021, pp. 1–13, May 2021, doi: https://doi.org/10.1155/2021/9983652.