Addressing Class Imbalance in COVID-19 Chest X-Ray Classification via Data Augmentation and Focal Loss

1st Anthony Gilles Rudolfo
Computer Science Program
Computer Science Department
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia 11480
anthony.rudolfo@binus.ac.id

2nd Joceline Araki
Computer Science Program
Computer Science Department
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia 11480
joceline.araki@binus.ac.id

3rd Alexander Agung Santoso Gunawan
Computer Science Department
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia 11480
aagung@binus.edu

4th Rilo Chandra Pradana
Computer Science Department
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia 11480
rilo.pradana@binus.ac.id

Abstract— **Imbalanced class distributions in medical imaging datasets can severely degrade the ability of convolutional neural networks (CNNs) to detect underrepresented conditions such as COVID-19 in chest radiographs. This study compares two complementary approaches for mitigating class imbalance: classic image-level augmentation and focal loss, a classifier-level cost function that down-weights well-classified examples. We evaluate both strategies on the publicly available COVID-19 Radiography Database using an EfficientNet-B0 backbone with ImageNet pre-training. Data augmentation experiments employ rotations, flips, crops, and intensity adjustments to oversample minority classes, while focal loss experiments train on the original imbalanced set. Performance is assessed by accuracy, precision, recall, F1-score, area under the ROC curve (AUC) and Equal Error Rate. The results shows that the model using data augmentation achieves better metrics compared to the one using focal loss. These findings highlight data augmentation as an accurate and reliable method for handling class imbalance in medical image classification.**

Keywords—dataset imbalance, convolutional neural networks, image classification

I. Introduction

Convolutional Neural Networks (CNNs) are commonly used for image classification. A common application of CNNs is in medical image analysis, such as for identifying brain tumors using MRI scans [1]. However, there are cases when the available dataset is not ideal, particularly when the sample sizes across classes are disproportionate or imbalanced. Class Imbalance have been proved to impact CNN training, resulting in worsened generalization on the test set and convergence during learning, which in turn impairs prediction performance, particularly for the minority class [2] chest X-rays images of patients infected with COVID-19 are an example of an unbalanced dataset. This dataset was used by Hall et al., in the early phases of the pandemic [3] and has 135 chest X-rays (CXR) images of COVID-19 and 320 CXR images of bacterial and viral pneumonia. An imbalanced dataset may result in models being biased toward the majority class, leading to reduced performance on the minority class. In many applications where the minority classes must be accurately classified (e.g., medical image classification), this problem must be resolved.

There are two primary types of solutions for class imbalance [4]. The first type is the data-level solution where the dataset is altered by either oversampling or undersampling it. Oversampling techniques have been used frequently in deep learning and have proven to be effective. The most common type of oversampling is random minority oversampling, it simply duplicates the samples from minority classes that are randomly selected. This technique has been shown to be efficient but sometimes it may lead to overfitting [2]. A more robust alternative is Data Augmentation [5]. Data augmentation is a process of replicating samples from the original data or creating a synthetic data to increase the dataset size. In the case of image data, this refers various modifications including rotating, cropping, resizing, or altering the color space. The second type is the classifier-level solution which modifies the classification model without altering the dataset. For example, some techniques can change the probability of the prior classifications, while the others add weights to the sample from some classes that are misclassified [6]. In the recent days, Generative Adversarial Networks (GANs) [7] have been rising in popularity due to its ability to handle imbalance datasets by creating artificial samples [5].

The goal of this research is to evaluate whether Focal Loss can produce a better result and performance in classifying imbalanced medical images compared to traditional data augmentation.

II. Literature Review

A. Convolutional Neural Network (CNN)

CNNs were found to perform very well in many kinds of computer vision tasks [11]. Fig. 1 shows this specific neural network. Classifier and the convolution components make up a CNN. The convolutional part, made up of multiple convolution and pooling layers, is responsible for extracting features from the input images [12]. The classification section is made up of a softmax activation function and fully connected layer. The extracted features from the first layer are universal and can be used to solve other problems, whereas those from the final layers are more specialized and tailored to the specific dataset being used.

B. Pre-trained Model

Training a CNN model from scratch can be computationally expensive and time consuming. Therefore, it is preferred and common to use a pre-trained model in this field. In this research, we employ the baseline model of EfficientNet (i.e., EfficientNet-B0). The EfficientNet models come in several versions, each scaled using distinct compound coefficients, with EfficientNet-B7 as the most advanced and powerful model. It achieved a state of art performance with 84.3% top-1 accuracy on the ImageNet dataset, outperforming other CNNs available at the time, as reported by Tan and Le [13]. Additionally, it was 8.4 times smaller in model size and give a 6.1% improvement in inference speed. A baseline CNN model is taken into consideration to minimize the risk of overfitting, particularly since the task involves a multiclassification with limited image size and complexity. One of the best CNN for classification, the EfficientNet-B0 which is chosen as the baseline of the architecture.

C. Focal Loss

This study adopts focal loss as the primary loss function to address class imbalance in the context of medical image classification. Conventional loss functions such as cross-entropy often favor the majority class in imbalanced datasets, leading to inaccurate predictions for minority classes. To overcome this, we use Focal Loss which is introduced by Lin et al [14]. Focal loss uses a scaling factor that reduce the contribution of correctly predicted examples by modifying the cross-entropy formula and allowing the model to focus more on the misclassified examples.

D. Classic Image Augmentation

Classical Image Augmentation has been particularly useful for improving the model's ability to detect general cases and reduce the class imbalance in medical image classification. By applying several changes into the image, classical image augmentation artificially expands the training set, minimizes the overfitting issue and enhances the image robustness Shorten and Khoshgoftaar [15]. Common basic operations performed on images include the following:

- Flipping: The image is mirrored along the specified axis (i.e. vertically or horizontally) in this process.
- Rotation: A certain angle, such as 90, 180, or 270 degrees, can be used to rotate images.
- Cropping and Scaling: By choosing a random section of the image, random cropping enables the model to concentrate on various aspects of the same image.
- Brightness Adjustment: This technique modifies the brightness of the image.
- Stretching: This process, which is often referred to as shearing, moves a portion of the image along an axis, distorting it.

These augmentation methods jointly enhance the robustness of models trained on medical images, enhancing their performance on unseen data and avoiding overfitting issues [16].

E. Related Works

Several kinds of algorithms based on deep learning have been created to diagnose COVID-19 utilizing the CXR imaging samples. Applying pre-trained deep learning models to improve performance and efficiency, either through transfer learning or by designing specialized deep learning architectures.

Khobahi et al. [9] presented CoroNet, a semi-supervised deep learning designed to detect COVID-19 cases from CXR images. The model was trained using the COVIDX dataset and achieved an overall accuracy of 93,5%, demonstrating high sensitivity and strong positive predictive value (PPV) for COVID-19 detection. Similiarly, Wang et al. [8] developed COVID-Net, a CNN that is specifically trained to detect COVID-19 cases from CXR data. Their work relied on the COVIDX benchmark dataset, comprising 13,975 CXR images. Despite its size, the dataset still presents class imbalance issues. The reported accuracy for COVID-Net reached 93%. The accuracy achieved by the COVID-Net CNN is 93%. Xu et al. [10] designed a deep learning model that can distinguish between COVID-19, influenza-A viral pneumonia (IAVP), and healthy subjects by analyzing the images of computed tomography (CT) scans. The dataset in this study included 219 COVID-19 images, 224 IAVP, and 175 normal cases, split into 85.4% for training-validation and 14.6% for testing. The system attained performance metrics of 86.7% accuracy, 81.5% recall, 80.8% precision, and an F1-score of 81.1%. Additionally, Fedoruk et al. [17] implemented a Generative Adversarial Network (GAN) approach using StyleGAN2-ADA to produce synthetic CXR data for enhancing multi-class classification. They tested this method across three different setups involving COVID-19, viral pneumonia, and normal cases using models like EfficientNet-B0 and Inception-V3. Inception-V3 and EfficientNet-B0 were compared in this research under three distinct augmentation scenarios (i.e., no augmentation, classic augmentation and GAN based augmentation). The result from this research shows that GAN based augmentation improved performance compared to no augmentation, but it was less effective when compared to classical augmentation, with. EfficientNet-B0 achieved the highest accuracy of 90.2% on a balanced dataset with classical augmentation. Chamseddin et al. [18] proposed two approaches using Synthetic Minority Oversampling (SMOTE) and Weighted Categorical Loss (WCL) to balance the datasets and enhance the model's performance. This research explored six state of art CNNs (i.e., DenseNet201, CheXNet, MobileNetV2, ResNet152, VGG19, and Xception) on on three COVID-19 CXR datasets. The results showed that WCL and SMOTE greatly enhance classification accuracy, with DenseNet201 obtaining 98,64% accuracy with SMOTE and CheXNet obtaining 98,87% with 100% precision using WCL. Performance was enchanted by both models, but WCL was more computationally effective since SMOTE needed a significant amount of memory and time. Ullah et al. [19] introduced ChestCovidNet, a compact and robust deep learning framework that used is to detect COVID-19, pneumonia, and lung opacity from CXR images. Their model consists of 11 trainable layers, 3 fully connected layers and 8 convolutional layers which is designed for device with low computational power, and it operates efficiently due to its lightweight architecture. To enhance the performance, the framework uses batch and cross-channel normalization in addition to grouping the conventional layer, Leaky ReLU activation functions, and ShuffleNet units to extract multiscales features. Furthermore, the model's generalization capability was enhanced using data augmentation. For multi-class classification tasks involving COVID-19, normal, lung opacity, and pneumonia cases, the model achieved an impressive accuracy of 98.12%, with a recall and F1-score of

95.75%, as well as high precision. Al-Shargabi et al. [20] proposed COVID-CGAN, a deep learning apprach that uses a CGAN to detect COVID-19 and resolve imbalance problems in the dataset. TTo augment the training data, the authors generated synthetic CXR images using a customized CGAN, resulting in a dataset composed of 84.8% synthetic and 15.2% original images. Five distinct deep learning models (i.e., InceptionResNetV2, Xception, SqueezeNet, VGG16, and AlexNet) were trained using this enriched dataset over 10 epochs. Among them, InceptionResNetV2 outperformed the others, achieving a notable accuracy of 99.72%, thereby demonstrating the effectiveness of synthetic data in boosting classification performance.

In this research, we evaluate the performance of focal loss compared with classic data augmentation like flipping, cropping or stretching. Traditional loss functions often struggle with class imbalance, which results in skewed model predictions. We addressed this problem using focal loss which was first introduced for object detection tasks and modifies the loss contribution according to the sample difficulty. This research provides insights into whether weighted loss functions give a better alternative to a classic data augmentation technique in handling an imbalanced dataset for medical image classification.

III. Methodology

The flow of our research is outlined in Fig. 1. Initially, the dataset undergoes preprocessing before it is divided into different sets for different purposes (i.e., training, validation, and testing). In the first experiment, we address the class imbalance problem using focal loss. In the second experiment, we apply standard image augmentation techniques to manage the class imbalance. Following that, the balanced dataset is utilized to train the CNN model. During this stage, we trained and evaluated the model to identify cases of COVID-19, viral pneumonia, and healthy lungs. Lastly, we compared the evaluation metrics of each model to identify the method that yields the most accurate results for detecting COVID-19 in CXR images.

A. Dataset

The dataset used in this research is "COVID-19 Radiography Database" Chowdhury et al [21]. Rahman et al[22]. A collaborative team that includes researchers from Qatar University and Dhaka University, partners from Malaysia and Pakistan created the dataset. The dataset comprises chest X-ray (CXR) images sorted into multiple categories: 3,617 images of confirmed COVID-19 cases, 10,193 images labeled as normal, and 1,346 images of viral pneumonia cases. This dataset is known for its class imbalance, where normal cases are significantly greater than the other categories. Addressing this imbalance is important to enhance model performance in medical image classification tasks.

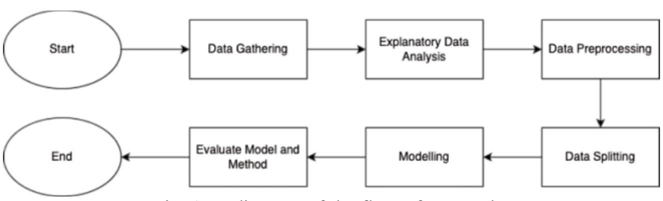


Fig. 1. A diagram of the flow of our study.

B. Data Preprocessing

In both experiments, the images were resized to 224x224 pixels. Then, we apply Contrast-Limited Adaptive Histogram Equaliztion (CLAHE) to enhance the structural details. CLAHE was developed to resolve the limitation of conventional Histogram Equalization (HE) which work by dividing each image into small sections, equalizes the histogram to each block independently and blends them together [24]. Fig. 2 shows the outcomes after applying CLAHE to a CXR image both before and after.

At the end of this step, the dataset is then split into 80% for training, 10% for validation, and 10% for testing. After that, in the first experiment the model is trained using focal loss method. In the second experiment, the classic image augmentation technique was applied on the minority class within the training data. Additionally, categorical cross entropy loss in this experiment. The image is randomly rotated by 5%, randomly darkened or brightened by up to 15%, randomly zoomed in the image up to 10%, and shifted the image and cardinal axes at random within the 8% range.

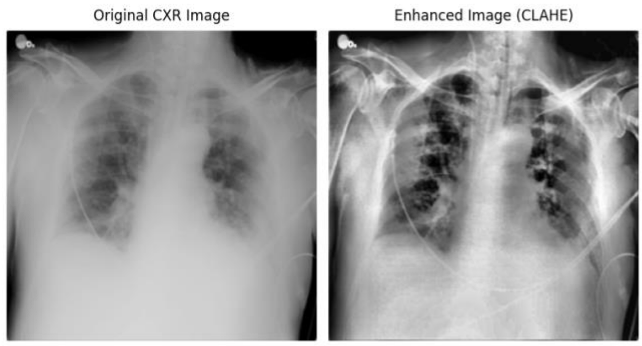


Fig. 2. Original image (left) and enhanced image with CLAHE (right)

C. Training and Evaluation Metrics

The model utilizes transfer learning, employing a pre-trained EfficientNet-B0 based on ImageNet. A completely connected layer designed for three class classification takes the place of the last classification layers, and a softmax activation function comes next. Additionally, global average pooling is applied, along with two dropout layers with values of 0.3 and 0.4, and L2 regularization with a value of 0.001. During the initial epochs, only the new classification layer is trained, while the rest of the EfficientNet-B0 remains frozen. Afterward, the entire model is fine-tuned. Training is conducted using the Adam optimizer with a learning rate of 0.0001, a batch size of 32, and learning rate decay. To assess the classification performance—particularly in addressing class imbalance—a wide range of evaluation metrics is applied. These include precision, recall, F1-score, and AUC-ROC [23]. Since this is a multiclass imbalance problem, we also use confusion matrix and Equal Error Rate.

Training and evaluation are implemented using Python in Google Collab with 12GB RAM and T4 GPU.

IV. Results and Discussions

A. Focal Loss

The model trained using Focal Loss demonstrates strong performance in COVID classification, as seen in both the evaluation metrics and visual plots. The results are shown in Fig. 3 and Table 1. Focal Loss works by reducing the relative loss for well-classified examples and increasing it for hard-to-classify samples. This shifts the model's learning focus toward minority or misclassified examples—an approach that is particularly effective when dealing with class imbalance, which is common in medical image datasets.

In the classification report, this model achieved a COVID accuracy of 0.93, with precision of 0.87, recall of 0.93, and an F1-score of 0.90, indicating a high capability in correctly identifying COVID cases while maintaining balance in prediction quality. The model also achieved an AUC of 0.9833, showing strong class separability, and an EER of 0.0532, suggesting low misclassification near the decision threshold. In the ROC curve shown in Fig. 4 the COVID class reaches an AUC of 0.9624, with comparably high scores for the NORMAL (0.9776) and viral pneumonia (0.9958) classes, indicating that the model can reliably distinguish among all categories. The confusion matrix in Table 2 supports this performance, with 316 COVID cases correctly identified and relatively few misclassifications (39 to NORMAL, 7 to viral pneumonia). NORMAL and viral pneumonia classes also show strong classification accuracy, supporting the model's well-rounded performance.

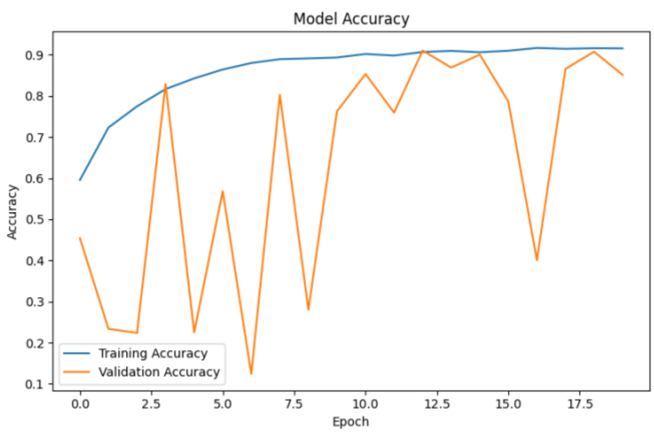
The curves displayed in Fig. 3(a) and Fig. 3(b) display some instability during the early epochs—likely due to the dynamic sample weighting introduced by focal loss. Despite this, the model converges well after several epochs, achieving high training accuracy and maintaining a low, consistent validation loss, confirming good generalization capability. Overall, the model trained with Focal Loss performs robustly across all metrics and visual evaluations, making it suitable for use in challenging classification tasks involving imbalanced medical datasets.

TABLE 1: COVID classification performance using Focal Loss.

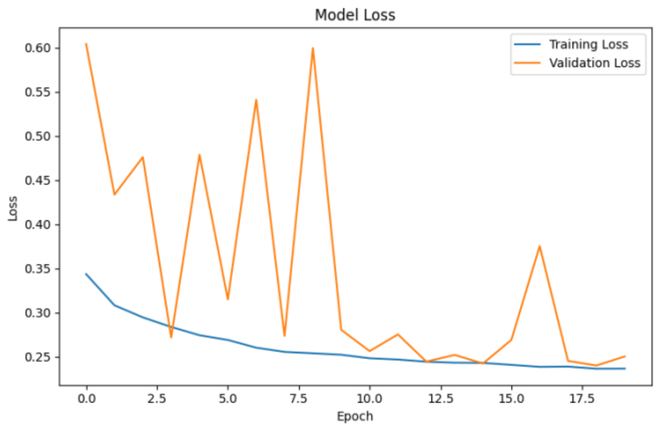
Accuracy	Precision	Recall	F1-Score	AUC	EER
0.93	0.87	0.93	0.90	0.9833	0.0532

TABLE 2: Confusion matrix

True/Predicted	Covid-19	Normal	Viral Pneumonia
Covid-19	316	39	7
Normal	29	949	44
Viral Pneumonia	1	2	139



(a) Model accuracy



(b) Model loss

Fig. 3. Model performance plots using Focal Loss: (a) accuracy, (b) loss.

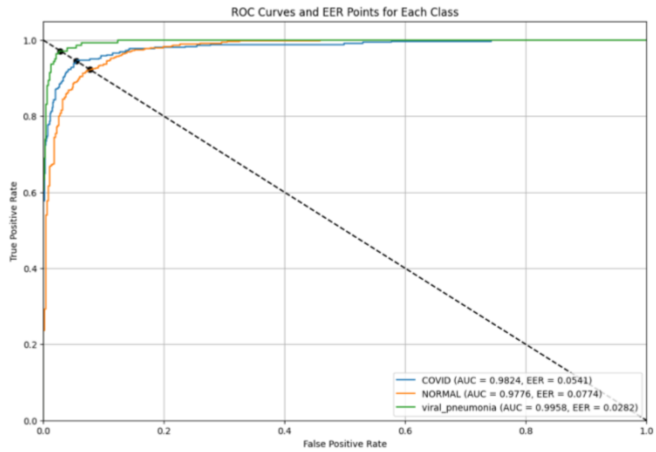


Fig. 4. Model performance plots using Focal Loss: ROC Curves.

B. Data Augmentation

The model trained using data augmentation demonstrates strong performance in COVID classification, as seen in both the evaluation metrics and visual plots. The result are shown in Fig. 5 and Table 3. Augmentation techniques such as flipping, scaling, and rotation were used to artificially expand the training data variety. These transformations improve the model's ability to generalize and handle real-world data inconsistencies. Performance metrics from the classification report show that the model attained a COVID detection accuracy of 94%, precision of 92%, recall of 95%, and an F1-score of 93%, reflecting its high capability in identifying COVID cases while minimizing false positives.

The ROC curve in Fig. 6, COVID class achieves an excellent AUC of 0.9936, with similarly high AUC values for the other two classes (0.995 for NORMAL and 0.999 for viral pneumonia), suggesting strong separability across all categories. Additionally, the macro-average Equal Error Rate (EER) is 0.0253, confirming that the model maintains a very low overall balance point between false positives and false negatives. The confusion matrix in Table 4 further supports this, showing that most COVID samples are correctly classified, with relatively few misclassifications (23 to NORMAL and only 2 to viral pneumonia). The Normal and viral pneumonia categories also showed strong predictive accuracy, emphasizing that the model's performance is consistent across all categories.

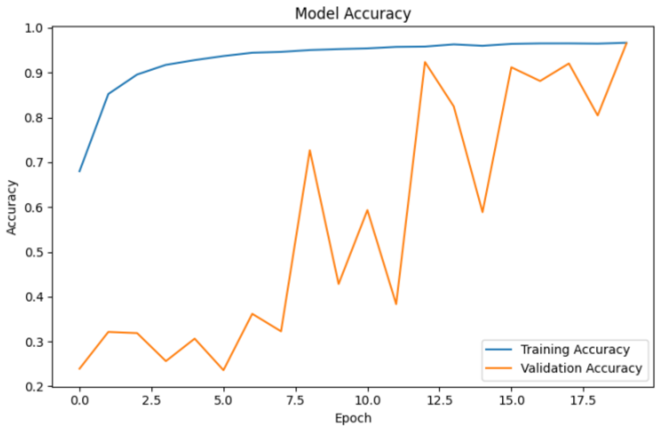
However, the training and validation curves in the accuracy/loss plots in Fig. 5(a) and Fig. 5(b) reveal some volatility early in training—likely due to the increased variability introduced by augmentation. Despite that, the model converges well with high final accuracy and stable validation loss after epoch 10, confirming the effectiveness of augmentation in enhancing generalization and reducing overfitting. Overall, these findings highlight the effectiveness and resilience of the data-augmented model, especially in scenarios involving class imbalance and complex visual patterns in medical imagery.

TABLE 3: COVID Classification Result using Data Augmentation

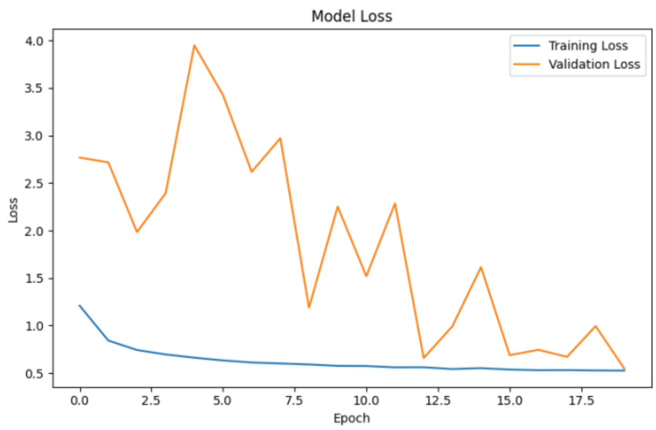
Accuracy	Precision	Recall	F1-Score	AUC	EER
0.94	0.92	0.95	0.93	0.9936	0.0253

TABLE 4: Confusion matrix on Data Augmentation Method

True/Predicted	Covid-19	Normal	Viral Pneumonia
Covid-19	337	23	2
Normal	14	1018	7
Viral Pneumonia	0	3	139



(a) Model accuracy



(b) Model loss

Fig. 5. Model performance plots using Data Augmentation: (a) accuracy, (b) loss.

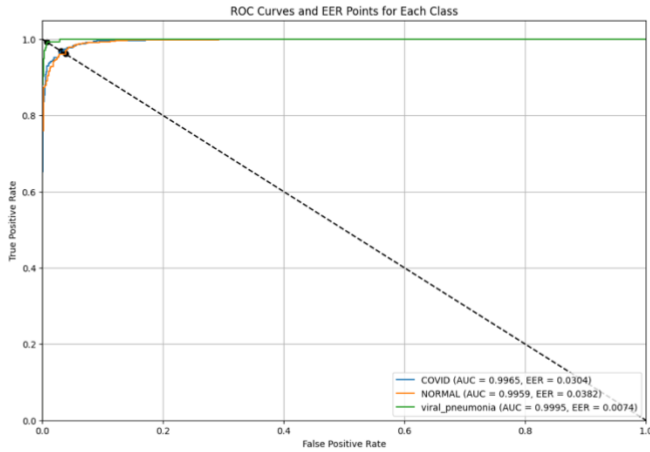


Fig. 6. Model performance plots using Data Augmentation: ROC Curves.

V. CONCLUSION

This study evaluates the effectiveness of two methods to address class imbalance using the "COVID-19 Radiography database" dataset. The data augmentation model outperforms the focal loss model specifically on the COVID class, achieving higher precision, recall, F1-score, and AUC. Its ability to generalize well to varied input forms likely contributes to its superior performance in detecting COVID, a class with subtle and variable imaging features. This is further supported by its lower Equal Error Rate (EER) for COVID (0.0304), compared to the focal loss model (0.0541), and fewer misclassifications in the confusion matrix. However, this comes at the cost of early training instability, as evident in the erratic validation curves before epoch 10. The focal loss model, on the other hand, offers a more stable training process. Although its metrics are slightly lower, it still maintains strong recall and decent precision. Its performance on minority classes like viral pneumonia is especially notable, with a lower EER (0.0282) and a more balanced confusion matrix, showing its strength in handling class imbalance. Visually, the ROC curves confirm that both models perform well overall, but the model with augmentation has a clear edge in COVID classification. This study acknowledges its shortcoming, including the use of one dataset and one model architecture which restrict how broadly the result can be applied into different datasets or different model architectures. The data augmentation was based on standard settings rather than being refined and the optimal focal loss hyperparameter was not carried out due to computational limitations. In practice, the best results could likely be achieved by combining both techniques, leveraging data augmentation to improve generalization and focal loss to enhance sensitivity to underrepresented classes. Combining both techniques is left for future work. In conclusion, classic data augmentation show better performance compared to focal loss, especially for the underrepresented COVID-19 class.

AUTHOR CONTRIBUTION AND AI USAGE DECLARATION

A.G.R. initiated and designed the research, took lead in writing the manuscript and conducted the experiment. J.A. assisted in writing and drafting the manuscript. A.A.S.G. and R.C.P. supervised the research and offered ongoing guidance during the study. All authors have reviewed and approved the final manuscript. We declare that we use AI tools that is Quillbot and ChatGPT in this manuscript to check for grammar, paraphrasing and support the coding task. The result from AI is not used directly and is under human supervision for error checking.

Availability Data and Materials

The COVID-19 Radiography Database is available for public and can be accessed in https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database. All Associated code is available on google colab https://colab.research.google.com/drive/1R5kaZoel5PXU5pXfi7XOrdW2iruOB_Va?usp=sharing https://colab.research.google.com/drive/1Ky297pzwDDV-gJfod0WsFC2nVqOHxvBq?usp=sharing

References

[1] Chattopadhyay A, Maitra M. MRI-based brain tumour image detection using CNN based deep learning method. *Neuroscience Informatics*. 2022;2(4):100060. doi:10.1016/j.neuri.2022.100060.

[2] Buda M, Maki A, Mazurowski MA. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*. 2018;106:249–259. doi:10.1016/j.neunet.2018.07.011.

[3] Hall LO, Paul R, Goldgof DB, Goldgof GM. Finding Covid-19 from Chest X-rays using Deep Learning on a Small Dataset. *arXiv*. 2020. Available from: https://arxiv.org/abs/2004.02060.

[4] Weiss GM. *Foundations of Imbalanced Learning*. In: *Imbalanced Learning*. John Wiley & Sons, Ltd; 2013. p. 13–41. doi:10.1002/9781118646106.ch2.

[5] Goceri E. Medical Image Data Augmentation: Techniques, comparisons and interpretations. *Artif Intell Rev*. 2023 Mar;56(11):12561–12605. doi:10.1007/s10462-023-10543-z.

[6] Thabtah F, Hammoud S, Kamalov F, Gonsalves A. Data imbalance in classification: Experimental evaluation. *Inf Sci.* 2020;513:429–441. doi:10.1016/j.ins.2019.11.004.

[7] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial networks. *Commun ACM.* 2020 Nov;63(11):139–144. doi:10.1145/3422622.

[8] Wang L, Wong A. COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-Ray Images. *arXiv*. 2020. Available from: https://arxiv.org/abs/2003.09871.

[9] Khobahi S, Agarwal C, Soltanalian M. CoroNet: A Deep Network Architecture for Semi-Supervised Task-Based Identification of COVID-19 from Chest X-Ray Images. *medRxiv*. 2020 Apr 14. doi:10.1101/2020.04.14.20065722.

[10] Xu X, Jiang X, Ma C, Du P, Li X, Lv S, Yu L, Ni Q, Chen Y, Su J, Lang G, Li Y, Zhao H, Liu J, Xu K, Ruan L, Sheng J, Qiu Y, Wu W, Liang T, Li L. A Deep Learning System to Screen Novel Coronavirus Disease 2019 Pneumonia. *Eng*. 2020;6(10):1122–1129. doi:10.1016/j.eng.2020.04.010.

[11] Bengio Y. Learning Deep Architectures for AI. *Found Trends® Mach Learn.* 2009;2(1):1–127. doi:10.1561/2200000006.

[12] O'Shea K, Nash R. An Introduction to Convolutional Neural Networks. *arXiv*. 2015. Available from: https://arxiv.org/abs/1511.08458.

[13] Tan M, Le QV. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *arXiv*. 2020. Available from: https://arxiv.org/abs/1905.11946.

[14] Lin TY, Goyal P, Girshick R, He K, Dollar P. Focal Loss for Dense Object Detection. In: *2017 IEEE International Conference on Computer Vision (ICCV)*; 2017 Oct 22–29; Venice, Italy. IEEE. p. 2999–3007. doi:10.1109/ICCV.2017.324.

[15] Shorten C, Khoshgoftaar TM. A survey on image data augmentation for Deep Learning. *J Big Data.* 2019 Jul;6(1):19. doi:10.1186/s40537-019-0197-0.

[16] Fedoruk O, Klimaszewski K, Ogonowski A, Kruk M. Additional look into GAN-based augmentation for deep learning COVID-19 image classification. *Machine Graphics and Vision.* 2023 Dec;32(3/4):107–124. https://mgv.sggw.edu.pl/article/view/5930.

[17] Fedoruk O, Klimaszewski K, Ogonowski A, Możdżonek R. Performance of GAN-based augmentation for deep learning COVID-19 image classification. In: *International Workshop on Machine Learning and Quantum Computing Applications in Medicine and Physics: WMLQ2022.* 2024. doi:10.1063/5.0203379.

[18] Chamseddine E, Mansouri N, Soui M, Abed M. Handling class imbalance in COVID-19 chest X-ray images classification: Using SMOTE and weighted loss. *Appl Soft Comput.* 2022;129:109588. doi:10.1016/j.asoc.2022.109588.

[19] Ullah N, Khan JA, Almakdi S, Alshehri MS, Al Qathrady M, Anwar MS, Syed IK. ChestCovidNet: An Effective DL-based Approach for COVID-19, Lung Opacity, and Pneumonia Detection Using Chest Radiographs Images. *Biochem Cell Biol.* 2023;0(ja):null. doi:10.1139/bcb-2023-0265.

[20] Al-Shargabi AA, Alshoibahi JF, Alabdulatif A, Alrobah N. Covid-CGAN: Efficient deep learning approach for COVID-19 detection based on CXR images using conditional GANs. *Appl Sci.* 2021 Aug;11(16):7174. doi:10.3390/app11167174.

[21] Chowdhury MEH, Rahman T, Khandakar A, Mazhar R, Kadir MA, Mahbub ZB, Islam KR, Khan MS, Iqbal A, Emadi NA, Reaz MB, Islam MT. Can AI Help in Screening Viral and COVID-19 Pneumonia? *IEEE Access.* 2020;8:132665–132676. doi:10.1109/ACCESS.2020.3010287.

[22] Rahman T, Khandakar A, Qiblawey Y, Tahir A, Kiranyaz S, Abul Kashem SB, Islam MT, Al Maadeed S, Zughaier SM, Khan MS, Chowdhury MEH. Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images. *Comput Biol Med.* 2021 May;132(C):104319. doi:10.1016/j.compbiomed.2021.104319.

[23] Hicks SA, Strümke I, Thambawita V, Hammou M, Riegler MA, Halvorsen P, Parasa S. On evaluation metrics for medical applications of artificial intelligence. *medRxiv*. 2021 Apr 7. doi:10.1101/2021.04.07.21254975. Available from: https://www.medrxiv.org/content/early/2021/04/09/2021.04.07.21254975

[24] Pizer SM, Amburn EP, Austin JD, Cromartie R, Geselowitz A, Greer T, Haar Romeny B, Zimmerman JB, Zuiderveld K. Adaptive histogram equalization and its variations. *Comput Vision Graph Image Process.* 1987;39(3):355–368. doi:10.1016/S0734-189X(87)80186-X