Examination of imbalance approach with prototypical Machine Learning methods to detect COVID-19 from chest X-ray images

Binit George
School of Computing
Dublin City University
Dublin, Republic of Ireland
binit.george2@mail.dcu.ie

Aneeta Charly V C
School of Computing
Dublin City University
Dublin, Republic of Ireland
aneeta.vazhappillycharly2@mail.dcu.ie

***Abstract*—The emergence of the COVID-19 pandemic raised significant health concerns all around the world. As the number of cases increased, radiology examination using chest radiography proved to be one of the vital screening approaches considering the lack of resources and labour required for RT-PCR screening. Many research studies in detecting COVID-19 rely on deep learning methods. In this study, we perform a multi-level classification on chest x-ray images to detect normal, COVID-19 and pneumonia cases using Convolutional Neural Network(CNN) and state-of-the-art transfer learning models(VGG-16 and DenseNet-121). Additionally, a comprehensive analysis is performed on how resampling techniques(SMOTE and Random Under sampler) affect the model's predictions considering this as an imbalance problem. The results show that VGG-16 delivers consistent performance with and without resampling compared to CNN and DenseNet121. Furthermore, we analyse the explainability of the model predictions using SHAP and perform a comparative analysis on how the critical regions identified by SHAP align with the actual decisions made by expert radiologists in identifying infections from radiographic images.**

***Keywords*—Deep Learning, COVID-19, Biomedical Imaging, CNN, Imbalance data**

I. Introduction

The COVID-19 outbreak had a colossal impact on the health and well-being of the global population [1]. Although the virus has potential multi-systemic involvement, the lung is the most common organ affected by SARS CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2) infection [2]. Reverse transcriptase-polymerase chain reaction (RT-PCR) testing was identified as an effective method to screen infected patients by detecting SARS-CoV-2 ribonucleic acid (RNA) from respiratory specimens (collected through a variety of means such as nasopharyngeal or oropharyngeal swabs). Although RT-PCR testing was found to be the best method to screen infected patients, it is time-consuming and laborious [2]. Chest radiography(chest X-ray (CXR) or computed tomography (CT) imaging) was chosen as an alternative screening method where the radiologists look for visual indicators of the virus infection[3, 4, 5]. Radiographic abnormalities seen in CXR and CT images such as ground glass opacity, consolidation, bilateral and interstitial can be identified as COVID-19 positive cases [5]. Although there are several studies revolving around CT imaging due to greater image details, there are numerous advantages to using CXR imaging to screen for COVID-19, particularly in areas with resource constraints and heavy infections. CXR imaging is advantageous over CT imaging because of its high availability, accessibility, portability, and support for rapid triaging in heavily infected population areas [6]. However, since the visual indicators are subtle in CXR imaging, there is a need for expert radiologists to interpret the images, which is considered the most significant bottleneck faced for COVID-19 screening.

Many research studies were performed to detect the virus and hence to support the healthcare system. This research explores the efficacy of neural networks by leveraging the readily available and accessible CXR imaging modality and performing a comparative study of a custom CNN model and transfer learning models (VGG-16 and DenseNet-121) to detect COVID-19 from a set of chest X-ray images comprising healthy, viral pneumonia affected and COVID-19 affected cases. Additionally, this study investigates this as an imbalance problem and evaluate the effect of resampling techniques on the model predictions. Furthermore, this study uses SHAP to assist the explainability of the predictions, which can help not only to gain deeper insights on critical factors but also help clinicians to perform better screening of the patients.

II. Related Work

A. *COVID-19 detection from chest X-ray images using neural networks*

With COVID-19 spreading rapidly, many research studies were performed to detect the virus and to help the healthcare system. Wang et al. [9] discuss the pattern of lung abnormalities and manifestations in chest X-rays for COVID -19 and suggest that chest X-rays can be a reliable means to identify COVID-19 cases. Some research studies used several deep learning techniques on COVID-19 x-ray images for detection [7, 8, 9]. However, few researchers approached the study as a classification problem, where they considered a set of x-ray images with different chest infections, including pneumonia and COVID-19, as input [7, 8, 9]. They used DNN models to classify COVID-19 from other diseases. ResNet50 produced an accuracy of 88.92% [7], and the VGG-19 model produced an overall accuracy of 93.3% [7] in detecting COVID-19, while DenseNet showed a prediction accuracy of 93.26% in detecting pneumonia[8]. In[8], the authors examined the influence of augmentation concerning detection accuracy, dataset diversity, augmentation methodology, and network size. They compared the performance of 17 deep learning algorithms with and without different geometric augmentations on three data sets. The authors made performance comparisons based on accuracy and MCC (Matthews Correlation Coefficient) on 17 pre-trained neural networks, including AlexNet, SqueezeNet, GoogleNet, and ResNet-50. However, the authors concluded that the accuracy and MCC for models which use augmentation were lower than that of the model which did not use augmentation. To fine-tune the existing deep learning networks for the medical imaging process from overfitting and low transfer efficiency, Shuaijing Xu et al. in [11] designed a hierarchical convolutional neural network (CNN) structure for ChestXray14. Using a sinloss function in CXNet-m1, the author claims to achieve better performance in terms of accuracy, recall, F1-score, and AUC than the best performing deep network model, ResNet-50-DCNN.

B. *Imbalance approach on image classification*

To balance class distribution in imbalanced datasets, Sun et al. [12] discuss two techniques - undersampling and oversampling. Under-sampling randomly removes the majority class records, whereas over-sampling increases the records in the minority class to get a balanced class distribution. Chawla et al. [13] proposed a combination of oversampling and undersampling, which claim to achieve better performance in terms of the ROC curve. They use the Synthetic Minority Oversampling Technique (SMOTE) rather than oversampling with replacement. SMOTE has garnered praise in several real-world applications [13] and in biomedical research [15]. Since Baseline SMOTE [12] gained much attention, several extensions of SMOTE like ADASYN, MWMOTE, WSMOTE, and k-means SMOTE are also popular now [14].

C. *Explainability using Shapley values*

Explainable Artificial Intelligence (XAI) is a promising research area where researchers are interested in interpreting how a machine learning algorithm makes a decision [16]. According to [19], two categories of explainability models - proxy and direct strategies. In proxy methods [17,18], the deep learning model is approximated by a proxy model and interpreted by querying the proxy model. In the direct category [20,21], the interpretation is made by studying the internal behaviour of the DNN and using that information to explain the DNN. In their study, Marco et al. [18] discuss LIME (Local Interpretable Model-agnostic Explanations), an example of a proxy method that interprets the model predictions by locally approximating the model around a given prediction. Expected gradients [20] and smoothGrad [21] are gradient-based direct explainability models that help measure each input's importance in decision-making. SHAP (SHapley Additive exPlanations), introduced by Lundberg et al. [17], is another proxy method that provides a unified framework for interpreting predictions. SHAP interprets a model using additive feature attribution, i.e., an output model defined as a linear combination of input features. As per the author, SHAP uses a concept from game theory that resolves the problem of computing the contribution of every set of features in a dataset with m features to a model's prediction. Magalathu et al. [22], in their study, point out global as well as local interpretability as a unique advantage in SHAP. Furthermore, the author highlights that SHAP not only interprets the feature model but also helps showcase whether each input feature's contribution is positive or negative. Marcilio et al., in [23] used SHAP to order the features according to their importance and used them as a feature selection strategy. In their study, SHAP gave more promising feature selection results than feature selection methods like mutual information, recursive feature elimination, and ANOVA. Singh et al., in [24] used SHAP to explain the prediction of the Inception-v3 based model to detect choroidal neovascularization (CNV), diabetic macular edema (DME), and drusen from a dataset of over 80,000 Optical coherence tomography (OCT) images. Their results showed successful attribution of specific pathological regions in the OCT images responsible for a given condition. Young et al.[25], in their study used GradCAM and KernelSHAP to perform three sanity checks - reproducibility, model dependence, and sensitivity on their model prediction of skin lesions using dermoscopy images. Additionally, the author claims that using GradCAM and Kernel SHAP gave them potential sources of bias in the input. In this study, we perform a comparative analysis of the performance of the CNN model and transfer learning models - VGG-16 and DenseNet-121 with imbalance input data and data after applying resampling techniques- SMOTE (oversampling) and Random Undersampler (undersampling). Furthermore, we investigate the explainability of these models using SHAP. Finally, based on the SHAP results, we conduct a comprehensive analysis of the models in terms of the critical regions identified by SHAP and the actual decision-making process of an expert radiologist.

III. Methodology

A. *Dataset*

This research uses chest X-ray images retrieved from Kaggle [38]. It has 5144 images (460- COVID-19, 1266- normal and 3418- pneumonia) for training data and 1288 images(116- COVID-19, 317- normal and 855- pneumonia) for test data. Shown in Figure 1 are the representative images from the dataset. The dataset was chosen considering the investigation of the imbalance problem in this research. For this research, it is assumed that all the radiographs are Posterior-Anterior (PA). Chest x-ray images are the most used imaging technique, and the data is widely available even though it has challenges compared to other imaging modalities like CT imaging [12].

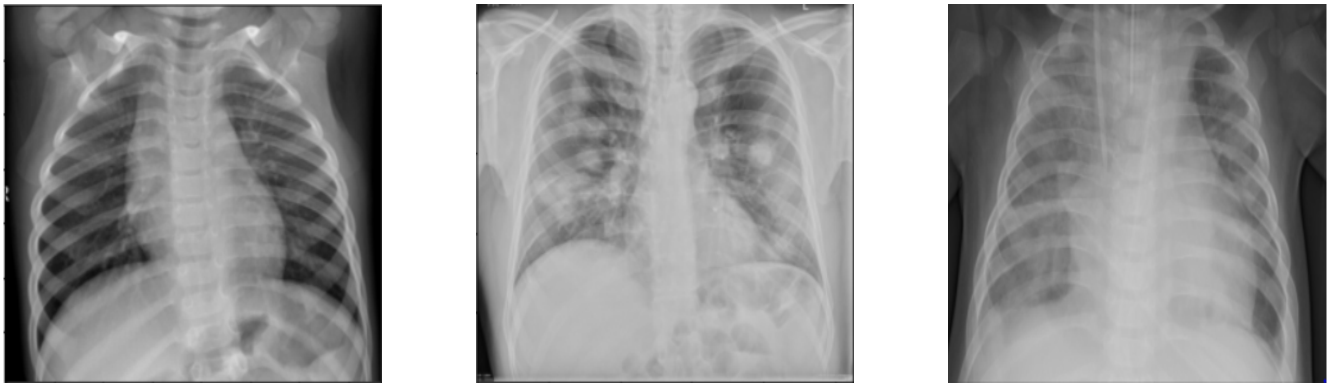

Fig. 1: Sample images from the dataset for normal (left), COVID-19 (middle) and pneumonia (right)

In the scope of this research, only chest x-rays are considered due to their high data availability and accessibility.

B. *Image preprocessing*

Images are described in three dimensions- height, width, and channel [13]. The shape of a sample COVID-19 x-ray image(figure1) is (609, 605, 3). Using Preprocessed images in neural networks shows an increase in the accuracy of predictions [12]. Image resizing is applied to elevate the prediction accuracy, and the images used in the neural networks should be of the appropriate size. A small-sized image will impact Keras's fine-tuning, and too large an image size will impact the model's accuracy as there are limited resources to add more layers to the network [7]. The input image size used in this research is (70,70,3). To normalise the brightness of the images and to improve the contrast, histogram equalisation was used [12]. Figure 2 shows the effect of applying histogram equalisation on a sample image.

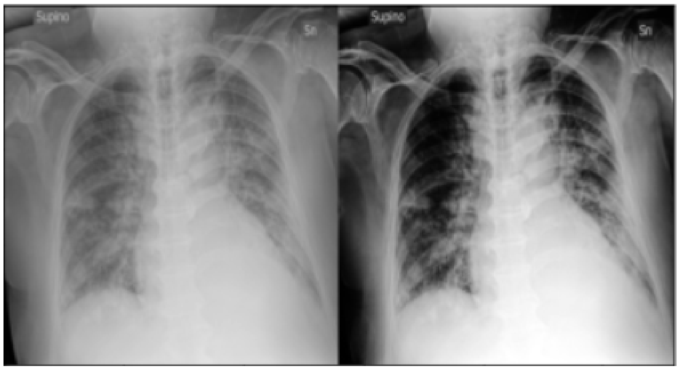

Fig. 2: The image on the left side is the original image and on the right side is the image after applying historgram equalisation

C. *Resampling Methods*

In general, imbalance problems in machine learning are dealt with by resampling [12]. Imbalance in the classes will lead the machine learning classifiers to bias toward the majority class. The resampling method either increases the number of instances in the minority class or eliminates some instances of the majority class, creating a balance between the classes.

*1) **Oversampling**:* This technique increases the number of instances in the minority class and is used extensively for imbalance problems in image datasets. Synthetic Oversampling Technique (SMOTE) is among many sought methods for oversampling [13]. Many extensions of SMOTE like ADASYN and SVMSMOTE are also widely popular. SMOTE works by drawing a line between two close features in the feature space and generating a new example somewhere along this line. Figure 3 shows input image distribution before and after performing SMOTE on the input data. Since oversampling does not consider majority classes there are possibilities of generating ambiguous examples. Studies also show that applying oversampling alone leads to over-fitting [12].

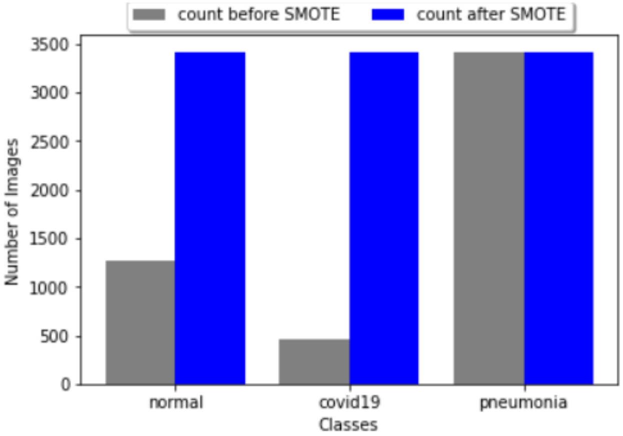

Fig. 3: The graph shows class wise image distribution before and after applying SMOTE

*2) **Undersampling**:* This technique randomly removes the instances from the majority class to gain class balance [13]. Random UnderSampling is a method that can be used for this purpose. However, since this method involves removing data, there is a considerable risk of data loss with undersampling [12]. Figure 4 shows input image distribution before and after performing Random Undersampling(RUS). This research

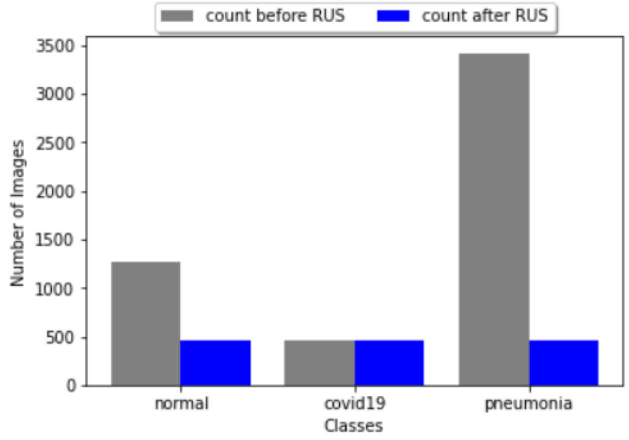

Fig. 4: The graph shows class-wise image distribution before and after applying Randum UnderSampler (RUS)

investigates the performance of the models when resampling methods- oversampling, and undersampling are used [8].

D. *Classification Models*

1) ***Convolutional Neural Network****:* In this study, a convolutional neural network was constructed to detect COVID-19 from chest x-ray images. The CNN model is designed with 10 layers- 3 convolutional layers, 2 pooling layers, 3 dropout layers, one fully connected layer, and one output layer. A 3x3 kernel is used for feature extraction as it works well with image size less than 128x128. Each convolutional block consists of a 2D Convolutional Network and a pooling layer followed by a dropout layer that randomly drops 30%, 50%, and 20% of the output of the applied layer. The pooling and dropout layer avoids the chance of overfitting in this model. Since multi- class classifications work well with softmax activation[11], the same is used in this study. Early stopping callback from Keras is used to stop training when the metric-validation loss shows no further improvement [37]. Keras ”Adam” was used as an optimiser as it is the most popular optimiser which produces better results, has faster computation time, and lesser parameters for tuning [26]. The model is fitted with 50 epochs and a batch size of 256.

2) ***Transfer Learning Models****:* Transfer Learning is a better approach suggested by literature to reduce the risk of overfitting or underfitting consequences on small training datasets. Transfer Learning takes advantage of the pre-trained CNNs with a larger dataset [27].

a) VGG-16: VGG16 is a convolution neural network, having convolution layers of 3x3 filter with a stride 1 and uses the same padding and maxpool layer of 2x2 filter of stride 2. In the end, it has 2 FC (fully connected layers) followed by a softmax for output. The 16 in VGG16 refers to the 16 layers that have weights [12]. VGG-16 model was provided with the input shape (70,70,3) and weights (‘ImageNet’) as arguments. To facilitate fine-tuning of the model and to save computational time only the output layer of the VGG-16 model is used in this study. As classes defined in this study are mutually exclusive, sparse categorical cross entropy was used as the loss function.

b) DenseNet-121: DenseNet-121 is a dense net model generated with 121 layers with pre-trained weights from the ImageNet database. The research conducted with DenseNet on CT images produced an accuracy of 92% and produced comparatively low accuracy for X-ray images [32]. Additionally, as compared to other models, DenseNet is less complex, and fine-tuning is easier [8]. Like the other two models used, sparse categorical cross entropy was used as the loss function, and softmax was used as the activation function. Global Average pooling was used in this model to avoid overfitting [36].

E. *Evaluation Metrics*

The performance of the CNN was assessed using the F1-score, precision, recall, ROC, and AUC Score. F1-score is used to compare the overall performance of the CNN. It combines recall and precision into a single metric [28]. The following interpretations of recall and precision were discussed in [7]:

- High Precision and High Recall: the model performs well with the classification.
- High Precision and Low Recall: the model is unable to correctly categorize the data points of a certain class or may overfit them.
- Low Precision and High Recall: the model correctly identifies data points from a certain class, but incorrectly labels a large number of data points from other classes.
- Low Precision and Low Recall: the model performed poorly in handling the classification.

The Area Under the Receiver Operating Characteristic ROC curve (AUC) examines the entire two-dimensional area underneath the ROC curve, covering from (0,0) to (1,1). This metric effectively checks the wellness and the quality of our models’ prediction performance [33].

Furthermore, this study investigates the interpretability of the classification models used. SHAP is used to explain the models and get a deeper insight into how the models perceive various image features. SHAP uses Shapley values from game theory to interpret the model decisions. By utilising the idea of game theory, SHAP assesses and identifies which features have the most significant contribution to the resulting prediction [29]. It is achieved by calculating the average marginal contribution of a feature in all possible coalitions. Each feature will be assigned a value of 0 or 1 by Coalition vector z’ to determine its presence in the coalition. These vectors are then converted to the feature space [30]. Processing in SHAP embeds the images in red and blue pixels, where red pixels represent those features that had a positive impact on classifying to a specific class, whereas blue pixels indicate the negative influence of the features on classifying to a particular class [31].

IV. Results

In this study, CNN and pre-trained models- VGG-16 and DenseNet-121 were used to perform multi-level classification on COVID-19, normal, and pneumonia cases from an input

Models	**Without Resampling**			
	F1-score(%)	***Recall (%)***	***Precision(%)***	***AUC (%)***
CNN VGG-16 DenseNet-121	92.71 93.29 92.81	92.00 92.00 95.00	91.00 91.00 90.00	98.9 97.86 98.33
	Oversampling - SMOTE			
CNN VGG-16 DenseNet121	97.03 94.20 96.83	97.00 94.00 97.00	97.00 94.00 97.00	99.65 99.18 99.71
	Undersampling - Random Undersampler			
CNN VGG-16 DenseNet121	85.87 93.84 67.39	88.00 94.00 76.00	86.00 94.00 67.00	98.47 98.19 93.55

TABLE I: Results obtained by the models with and without resampling

of preprocessed grayscale images. To analyse the effect of imbalance in the dataset, the models were tested without resampling and with resampling. The results obtained are shown in Table 1. It is seen that CNN and VGG -16 perform better than DenseNet with different resampling methods. Among these, VGG-16 outperforms the other two models for different

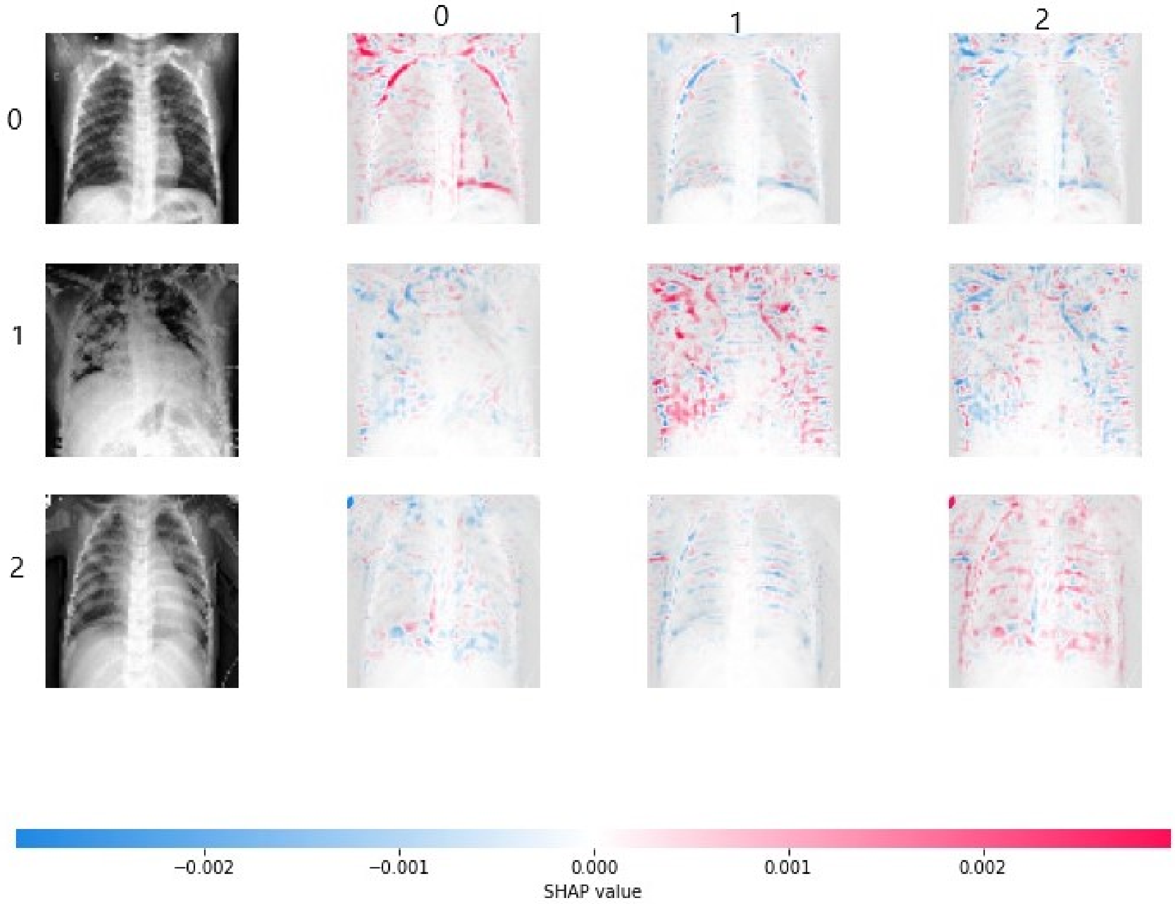

Fig. 5: SHAP interpretation on VGG-16 model predictions after oversampling. Classes represent 0- normal,1- COVID-19, 2-pneumonia.

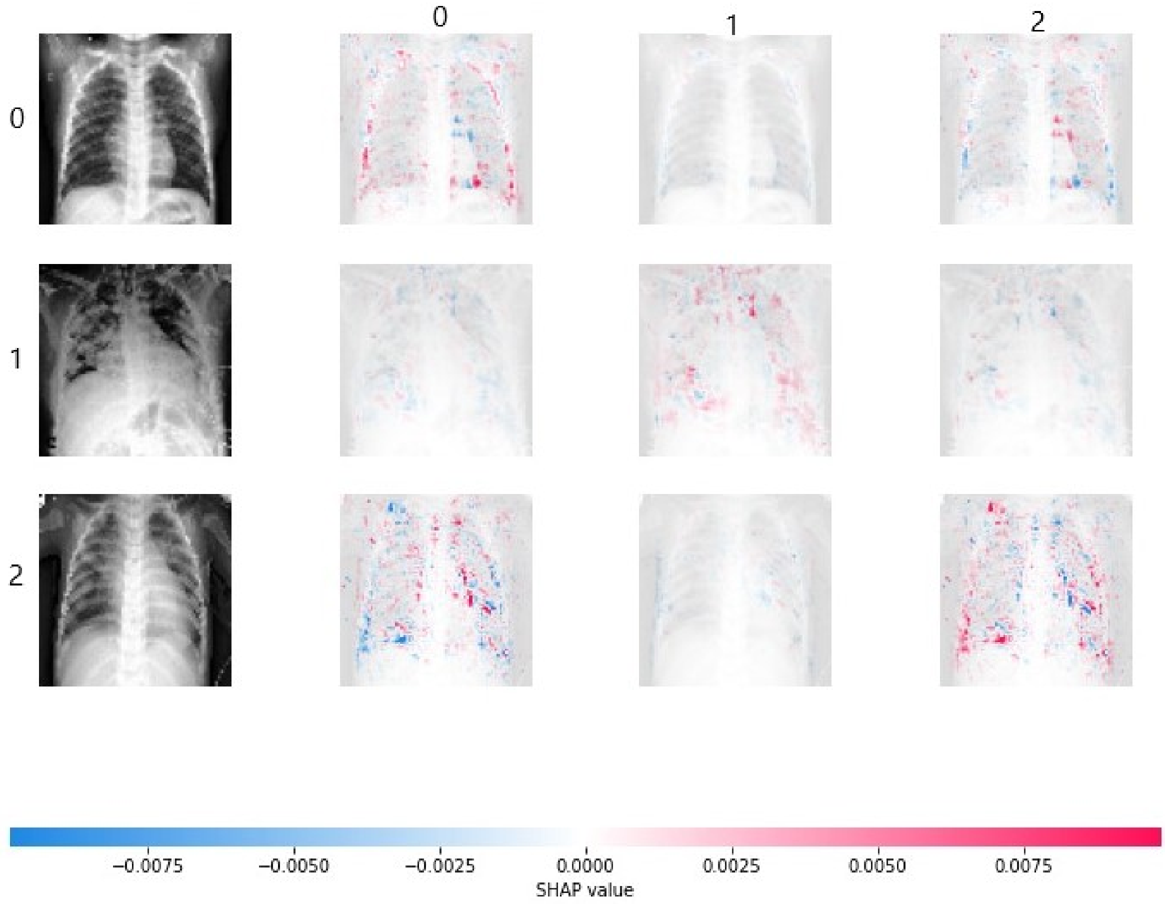

Fig. 6: SHAP interpretation on CNN model predictions after oversampling. Classes represent 0- normal,1- COVID-19, 2- pneumonia.

resampling methods. VGG-16 produced an F1 score of 93.29% without resampling, 94.20% with oversampling, and 93.84% with undersampling. The models show higher performance after SMOTE, with an F1 score of 97.03% for CNN, 94.02% for VGG-16, and 96.83% for DenseNet-121. Figure 7 and 8 shows the learning curve of VGG-16 model with SMOTE and Random Undersampling(RUS). The graphs suggest that the resampling methods work well with the model with no sign of overfitting.

Ground glass opacities (GGO) and consolidation play major significance in differentiating COVID-19 from viral pneumonia in radiological findings [34]. Although both infections involve multiple lobes bilaterally, GGOs were more common in patients with COVID-19, while consolidations were more common in viral pneumonia (H1N1) [35]. Furthermore, for COVID-19, peripheral and posterior zones of the lungs show significant involvement, whereas, in viral pneumonia, the pattern of lung involvement is more diffuse with central and peripheral zone involvement [34]. Nevertheless, it is noteworthy that in the advanced stages of COVID-19 infection, radiological image findings become less specific and indistinguishable from other types of pneumonia [34].

In the second evaluation phase, the best performing models- CNN and VGG-16 were processed using SHAP Deep Explainer to compare the decision-making process of both the models for classification. The SHAP interpretations were done on three test images, one from each class. Figure 5 shows the SHAP Deep explainer visualisation on the true predictions of VGG-16 model, and figure 6 shows the interpretation for the CNN model. The classes are represented as 0 for normal, 1 for COVID-19, and 2 for pneumonia. SHAP explainer embedded the image in red and blue pixels where the red pixels represent positive SHAP values that influenced the image to be classified to the particular class. On the other hand, the blue pixels are negative SHAP values that contributed to not classifying the images into a specific class [31]. The CNN model produced fewer SHAP distribution compared to VGG-16 owing to the lesser pixel intensity in CNN compared to VGG-16. In both VGG-16 and CNN models, true predictions for a normal class involved pixels around the boundary of the lungs. For COVID-19 class, both the models embed pixels in the central and posterior regions of the lungs aligning with the literature

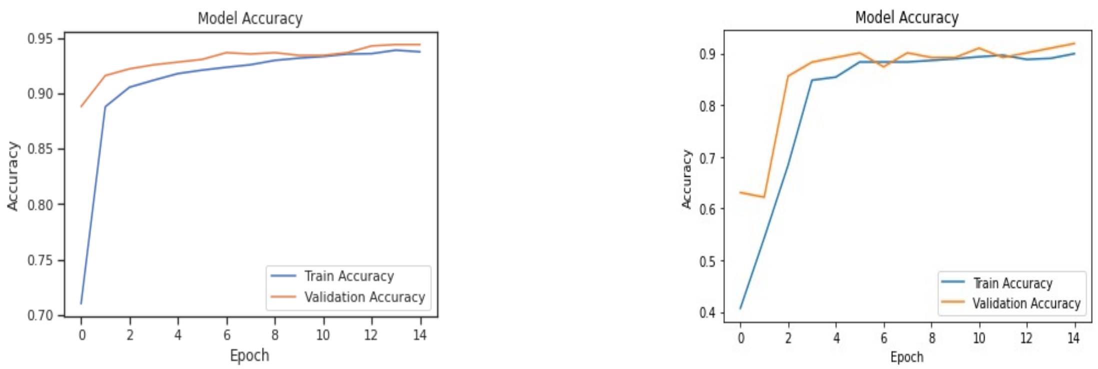

Fig. 7: Training and validation accuracy plots of VGG-16 model for SMOTE (left) and RUS (right) for 15 epochs

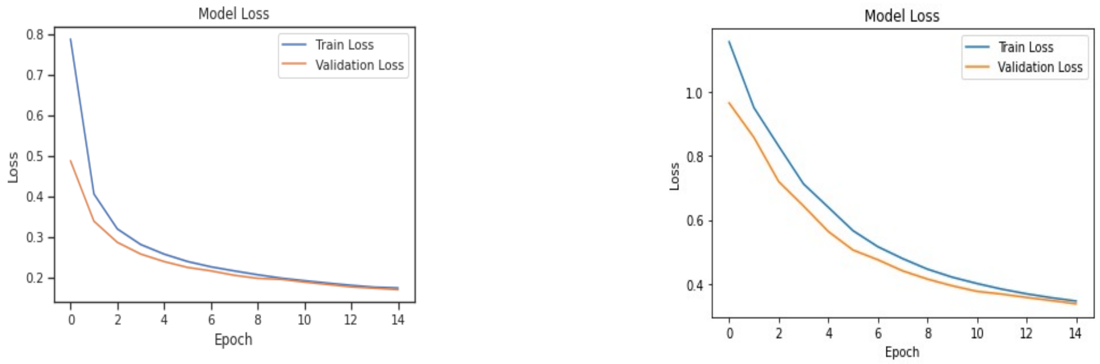

Fig. 8: Training and validation loss plots of VGG-16 model for SMOTE (left) and RUS (right) for 15 epochs

stating the peripheral and posterior zones of the lungs show significant involvement for COVID-19 [35]. It is noteworthy that VGG-16 embeds more red pixels than CNN, revealing that VGG-16 extracted image features more effectively than CNN. VGG-16 embeds red pixels in the central zone of the lungs for pneumonia classification, and the SHAP interpretation for both the models is different for the case of pneumonia. The results of the SHAP deep explainer show how the model interacts with the data making it helpful in fine-tuning the models. From the perspective of clinical end-users, the explanations build trust in the decision-making of the models and also help them identify potentially questionable decisions.

V. Conclusion and Future Work

With COVID-19 spreading rapidly, many studies have been conducted to detect COVID-19 cases from radiographic images. Deep Learning models have produced promising results in detecting COVID-19 cases from chest x-ray images. In this research, comparative analysis is conducted to assess the performance of CNN and transfer learning models (VGG-16 and DenseNet-121) in detecting COVID-19 images from a chest X-ray set of normal, COVID-19, and pneumonia cases. Additionally, this study applied resampling techniques to solve the imbalance in the data distribution and assessed the performance of the same. It was found that oversampling produced better performance in the three models. VGG-16 outperformed CNN and DenseNet121 models with an F1 score of 93.29% without resampling, 94.20% with oversampling, and 93.84% with undersampling. Furthermore, SHAP was used to interpret the decision-making of the models and deliver local explainability using three sample images of each class.

With exponential growth in COVID-19 cases and a significant amount of data availability, further research can be conducted with larger datasets to test the robustness of the models discussed in this research. Furthermore, this research focuses on local explainability using SHAP. Further research can use SHAP for global explainability, which can interpret the misclassifications by the models and use SHAP to discover bias in the input image data.

References

[1] Wang, W. et al. Detection of SARS-CoV-2 in different types of clinical specimens. JAMA 323(18), 1843–1844 (2020).

[2] Pan A, Liu L, Wang C, et al. Association of public health interventions with the epidemiology of the COVID-19 outbreak in Wuhan, China. JAMA. 2020;323(19):1915–23. https:// doi. org/ 10. 1001/ jama. 2020. 6130.

[3] Ng, M.-Y. et al. Imaging profile of the COVID-19 infection: radiologic findings and literature review. Radiol. Cardiothorac. Imaging 2(1), e200034 (2020).

[4] Huang, C. et al. Clinical features of patients infected with 2019 Novel Coronavirus in Wuhan China. The Lancet 395, 497–506 (2020).

[5] Guan, W. J., Hu Y., & Ni Z. Y. Clinical characteristics of Coronavirus disease 2019 in China. N. Engl. J. Med. 382(18), 1708–1720 (2020).

[6] Rubin, G. D. et al. The role of chest imaging in patient management during the COVID-19 pandemic: a multinational consensus statement from the fleischner society. Radiology (2020). https ://doi.org/10.1016/j.chest .2020.04.003.

[7] Albahli S, Yar G Fast and Accurate Detection of COVID-19 Along With 14 Other Chest Pathologies Using a Multi-Level Classification: Algorithm Development and Validation Study J Med Internet Res 2021;23(2):e23693 URL: https://www.jmir.org/2021/2/e23693 DOI: 10.2196/23693 Rank-Q1

[8] Saleh Albahli, Nasir Ayub, Muhammad Shiraz, Coronavirus disease (COVID-19) detection using X-ray images and enhanced DenseNet, Applied Soft Computing, Volume 110, 2021, 107645, ISSN 1568-4946, https://doi.org/10.1016/j.asoc.2021.107645.Rank-Q1

[9] Wang, L., Lin, Z.Q. Wong, A. COVID-Net: a tailored deep convolutional neural network design for detection of COVID- 19 cases from chest X-ray images. Sci Rep 10, 19549 (2020). https://doi.org/10.1038/s41598-020-76550-zRank-Q1

[10] Elgendi M, Nasir MU, Tang Q, Smith D, Grenier J-P, Batte C, Spieler B, Leslie WD, Menon C, Fletcher RR, Howard N, Ward R, Parker W and Nicolaou S (2021) The Effectiveness of Image Augmentation in Deep Learning Networks for Detecting COVID-19: A Geometric Transformation Perspective. Front. Med. 8:629134. doi: 10.3389/fmed.2021.629134.Rank-Q1

[11] S. Xu, H. Wu and R. Bie, ”CXNet-m1: Anomaly Detection on Chest X-Rays With Image-Based Deep Learning,” in IEEE Access, vol. 7, pp. 4466-4477, 2019, doi: 10.1109/ACCESS.2018.2885997.Rank-Q1

[12] Sun, Z., Song, Q., Zhu, X., Sun, H., Xu, B. and Zhou, Y., 2015. A novel ensemble method for classifying imbalanced data. Pattern Recognition, 48(5), pp.1623-1637.Rank-Q1

[13] Chawla, N.V., Bowyer, K.W., Hall, L.O. and Kegelmeyer, W.P., 2002. SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research, 16, pp.321-357.Rank-Q2

[14] Zhu, Q., Wu, Q. and Fan, Z., 2021. A novel oversampling technique for class-imbalanced learning based on SMOTE and natural neighbors. Information Sciences, 565, pp.438-455.Rank-Q1

[15] M. Nakamura, Y. Kajiwara, A. Otsuka, H. Kimura Lvq-smote-learning vector quantization based synthetic minority over-sampling technique for biomedical dataRank-Q1

[16] Lin, Zhong Qiu, Mohammad Javad Shafiee, Stanislav Bochkarev, Michael St Jules, Xiao Yu Wang, and Alexander Wong. ”Do explanations reflect decisions? A machine-centric strategy to quantify the performance of explainability algorithms.” arXiv preprint arXiv:1910.07387 (2019).

[17] Lundberg, S.M. and Lee, S.I., 2017. A unified approach to interpreting model predictions. Advances in neural information processing systems, 30.

[18] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. ”Why Should I Trust You?”: Explaining the Predictions of Any Classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD ’16). Association for Computing Machinery, New York, NY, USA, 1135–1144. https://doi.org/10.1145/2939672.2939778

[19] Erico Tjoa and Cuntai Guan. A survey on explainable artificial intelligence (xai): Towards medical xai. arXiv preprint arXiv:1907.07374, 2019.

[20] Ramprasaath R Selvaraju, Abhishek Das, Ramakrishna Vedantam, Michael Cogswell, Devi Parikh, and Dhruv Batra. Grad-cam: Why did you say that? arXiv preprint arXiv:1611.07450, 2016.

[21] Daniel Smilkov, Nikhil Thorat, Been Kim, Fernanda Viégas, and Martin Wattenberg. Smoothgrad: removing noise by adding noise. arXiv preprint arXiv:1706.03825, 2017.

[22] Mangalathu, S., Hwang, S.H. and Jeon, J.S., 2020. Failure mode and effects analysis of RC members based on machine-learning-based SHapley Additive exPlanations (SHAP) approach. Engineering Structures, 219, p.110927.

[23] W. E. Marcílio and D. M. Eler, ”From explanations to feature selection: assessing SHAP values as feature selection mechanism,” 2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), 2020, pp. 340-347, doi: 10.1109/SIBGRAPI51738.2020.00053.

[24] A. Singh, A. R. Mohammed, J. Zelek, and V. Lakshminarayanan, “Interpretation of deep learning using attributions: application to ophthalmic diagnosis,” in Proc.SPIE, Aug. 2020, vol. 11511, [Online]. Available: https://doi.org/10.1117/12.2568631.

[25] K. Young, G. Booth, B. Simpson, R. Dutton, and S. Shrapnel, “Deep Neural Network or Dermatologist?,” in Interpretability of Machine Intelligence in Medical Image Computing and Multimodal Learning for Clinical Decision Support, 2019, pp. 48–55. [21] S. M. Lundberg et al., “Explainable machine-learning predictions for the prevention of hypoxaemia during surgery,” Nat. Biomed. Eng., vol. 2, no. 10, pp. 749–760, 2018.

[26] Kingma, D.P. and Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

[27] Pan, S.J. and Yang, Q., 2009. A survey on transfer learning. IEEE Transactions on knowledge and data engineering, 22(10), pp.1345-1359.H-index:22

[28] Oyelade, O.N., Ezugwu, A.E.: Deep Learning Model for Improving the Characterization of Coronavirus on Chest X-ray Images Using CNN. medRxiv, (2020)

[29] Lundberg, S.M. and Lee, S.I., 2017. A unified approach to interpreting model predictions. Advances in neural information processing systems, 30.

[30] J. H. Ong, K. M. Goh and L. L. Lim, ”Comparative Analysis of Explainable Artificial Intelligence for COVID-19 Diagnosis on CXR Image,” 2021 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), 2021, pp. 185-190, doi: 10.1109/ICSIPA52582.2021.9576766.

[31] Podgorelec, Vili, Špela Pečnik, and Grega Vrbančič. 2020. ”Classification of Similar Sports Images Using Convolutional Neural Network with Hyper-Parameter Optimization” Applied Sciences 10, no. 23: 8494. https://doi.org/10.3390/app10238494

[32] N. Hasan, Y. Bao, A. Shawon, DenseNet convolutional neural networks application for predicting COVID-19 using CT image, 2020.

[33] Oyelade, O.N., Ezugwu, A.E.: Deep Learning Model for Improving the Character- ization of Coronavirus on Chest X-ray Images Using CNN. medRxiv, (2020)

[34] Eslambolchi A, Maliglig A, Gupta A, Gholamrezanezhad A. COVID-19 or non-COVID viral pneumonia: How to differentiate based on the radiologic findings? World J Radiol. 2020 Dec 28;12(12):289-301. doi: 10.4329/wjr.v12.i12.289. PMID: 33510853; PMCID: PMC7802079.

[35] Kör, H., Erbay, H. and Yurttakal, A.H., 2022. Diagnosing and differentiating viral pneumonia and COVID-19 using X-ray images. Multimedia Tools and Applications, pp.1-17.

[36] Montaha S, Azam S, Rafid AKMRH, Ghosh P, Hasan MZ, Jonkman M, De Boer F. BreastNet18: A High Accuracy Fine-Tuned VGG16 Model Evaluated Using Ablation Study for Diagnosing Breast Cancer from Enhanced Mammography Images. Biology (Basel). 2021 Dec 17;10(12):1347. doi: 10.3390/biology10121347. PMID: 34943262; PMCID: PMC8698892.

[37] Y. U. Sinn, K. M. Hopkinson, B. J. Borghetti and B. J. Steward, ”IR Small Target Detection And Prediction With ANNs Trained Using ASSET,” 2019 IEEE Aerospace Conference, 2019, pp. 1-11, doi: 10.1109/AERO.2019.8741671.

[38] https://www.kaggle.com/datasets/prashant268/chest-xray-covid19-pneumonia

[39] https://gitlab.computing.dcu.ie/georgeb2/2022-mcm-TERMPLATE