The Impact of Class Weight Optimization on Improving Machine Learning Outcomes in Identifying COVID-19 Specific ECG Patterns

Sara Khan[1], Walaa N. Ismail[2][a], Shada Alsalamah[3][b], Ebtesam Mohamed[4] and Hessah A. Alsalamah[3,5][c]

[1]*Department of Information Systems and Technology Management, George Washington University, Washington, DC, U.S.A.*
[2]*Management Information Systems Department, College of Business, Al Yamamah University, Riyadh, K.S.A.*
[3]*Information Systems Department, King Saud University, Riyadh, K.S.A.*
[4]*Faculty of Computer Science, Minia University, Minia, Egypt*
[5]*Computer Engineering Department, College of Engineering and Architecture, Al Yamamah University, Riyadh, K.S.A.*

Keywords: Convolutional Neural Network (CNN), Covid-19, Data Imbalance, Electrocardiogram (ECG), Class Weights, VGG16.

Abstract: The Covid-19 pandemic has resulted in 550 million cases and 6.3 million fatalities, with the virus severely affecting the lungs and cardiovascular system. A study utilizes a VGG16 model adapted for a 12-Lead ECG Image database to assess the disease's impact on cardiovascular health. The research addresses the challenge of data imbalance by experimenting with different training approaches: using balanced datasets, imbalanced datasets, and class weight adjustments for imbalanced datasets. These models are designed for a three-class multiclass classification of ECG images: Abnormal, Covid-19, and Normal categories. Performance evaluations, including accuracy scores, confusion matrices, and classification reports, show promising results. The model trained on a balanced dataset achieved a 90% accuracy rate. When trained on an imbalanced dataset, the accuracy dropped to 82%. However, with class weight adjustments, the accuracy rebounded to 87%. The study proves that the adapted VGG16 model can effectively handle both balanced and imbalanced datasets. Further testing and enhancements can be carried out using additional datasets, making it a valuable tool for understanding the cardiovascular implications of Covid-19.

1 INTRODUCTION

Training the models with an imbalanced dataset gives rise to a class imbalance problem, which is strongly discouraged in supervised machine learning. This is because during training the model becomes biased towards the class that is present in the majority. The model may achieve a high accuracy score, the reason being it over-classifies the majority class and fails to identify the minority class which ends up being misclassified. One of the ways to overcome data imbalance is by balancing class weights in an imbalanced dataset. In other words, the class weight of the category having less data is increased. During training, the machine learning algorithm by default assumes every category to be of equal weight. The learning can be influenced during training by passing customized class weight values. The minority class is given a higher-class weight, as the model trains every point, the error is multiplied by the weight of the point. The model attempts to minimize this error for categories having higher class weights.

2 RELATED WORK

In the ever-changing field of healthcare technology, the application of machine learning, especially deep learning, has shown significant promise. Numerous recent research efforts have covered various

[a] https://orcid.org/ 0000-0002-1499-438X
[b] https://orcid.org/ 0000-0002-3054-5015
[c] https://orcid.org/ 0000-0002-4761-0864

Khan, S., Ismail, W., Alsalamah, S., Mohamed, E. and Alsalamah, H.
The Impact of Class Weight Optimization on Improving Machine Learning Outcomes in Identifying COVID-19 Specific ECG Patterns.
DOI: 10.5220/0012413100003657

In *Proceedings of the 17th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2024) - Volume 2*, pages 562-567
ISBN: 978-989-758-688-0; ISSN: 2184-4305

dimensions of its capability, from monitoring cardiovascular health amid the COVID-19 pandemic to detecting heart abnormalities.

An innovative study integrated 5G technology into a real-time cardiovascular monitoring system tailored for COVID-19 patients (Tan, 2021). Utilizing a combination of convolutional neural networks (CNNs) and long short-term memory (LSTM) networks, the research achieved a prediction accuracy of 99.29%, illustrating the potential of real-time monitoring and deep learning in COVID-19 patient care.

Another study used a one-dimensional CNN (1D-CNN) for classifying various types of ECG rhythms and beats (Darmawahyuni , 2022). The model, trained on multiple databases, boasted an impressive accuracy of 99.98%, thereby demonstrating the power of deep learning in diagnosing complex heart abnormalities.

Arrhythmia Classification Focusing on the classification of arrhythmias into five categories, a particular study employed deep convolutional neural networks and used a well-established arrhythmia database for training (Raza, 2022). The model attained an accuracy of up to 98.9% with clean data, emphasizing the effectiveness and reliability of machine learning in heart disease diagnosis.

COVID-19 Detection Based on ECG Two studies specifically tackled the early detection of COVID-19 through ECG trace images (Shahin, 2021) (Attallah, 2022). One study tested multiple CNN architectures and found one model to outperform the others with an 89.64% accuracy rate. Another study examined a broader array of deep learning algorithms and achieved an accuracy rate of 98.8% in binary classification scenarios.

Beyond ECG: Other Applications in COVID-19 Detection Research has also extended into other diagnostic methods for COVID-19, particularly focusing on chest X-ray images (El-Rashidy, 2020) (Ozturk, 2020). High levels of accuracy, surpassing 97%, were achieved using various machine learning models, with one study notably demonstrating consistent training and testing accuracy, which speaks to the model's robustness.

In conclusion, these studies set robust benchmarks and provide a solid foundation in healthcare applications involving machine learning. The current study aims to contribute to this body of work by introducing a technique for optimizing class weights in imbalanced datasets to improve machine learning model performance.

3 METHODOLOGY

The work completed can be divided into four sections: Dataset Gathering, Pre-processing Dataset, Building and Training model, and Evaluating Results.

3.1 Data Gathering

The VGG16 model is trained using a publicly available ECG image database (Khan, 2021). This database was created by collecting 12-lead ECG images using the "EDAN SERIES-3" ECG device, with a sampling rate of 500 Hz. The device was installed in the Cardiac Care and Isolation units of various healthcare institutes across Pakistan. Initially, the database contained the following numbers of images: 250 for COVID-19 patients, 859 for normal individuals, 77 for myocardial infarction patients, 203 for patients with a previous history of myocardial infarction, and 548 for patients with abnormal heartbeats. For the purpose of three-class multiclass classification, images belonging to the abnormal, COVID-19, and normal categories were selected.

To create a balanced dataset, a total of 750 images were used, with each category containing 250 images. For an imbalanced dataset, 1470 images were utilized: 380 images from patients with abnormal heartbeats, 250 from COVID-19 patients, and 840 from normal individuals.

3.2 Pre-Processing Dataset

The methods used for processing the images are important for the machine to learn the necessary features to classify the images accurately. The images are processed in MATLAB using the following three steps including gamma correction (Fig. 1B), grayscaling (Fig. 1C), and cropping (Fig. 1D). For this specific problem, color is not an essential feature, hence the images are grayscaled. Grayscaling will reduce the computational power required and increase training speed simplifying the learning process. It also consumes less space which should be taken into consideration when dealing with large datasets. Gamma correction however helps in the brightness and contrast adjustments, The gamma value used is 0.6. The gamma is set < 1 to get the desired effect that is the image is brightened and darker regions are enhanced, decreasing sensitivity in difference of lighting and making relevant patterns easier to learn for the model.

3.3 Building and Training Model

Using Python 3.7, TensorFlow 2.9.2 library, and Keras interface, the VGG16 model is built, trained, and tested. The model accepts a pre-processed 12-Lead ECG Image as the input and undergoes three experiments. The images were categorized into the following types: Abnormal, Covid-19, and Normal. The detailed experiment architecture is depicted in Fig 2.

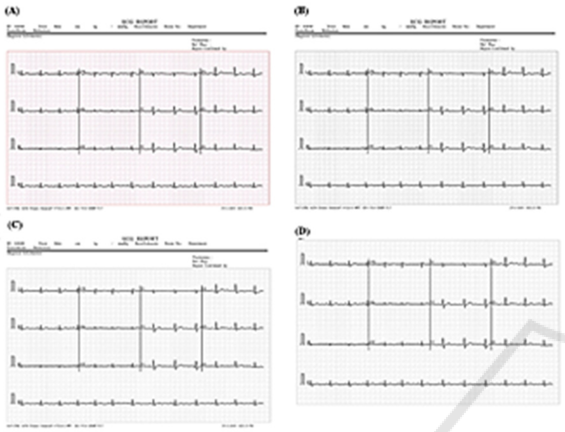

Figure 1: (A) Original Image, (B) Image after grayscaling, (C) Image after gamma correction, (D) Image after cropping.

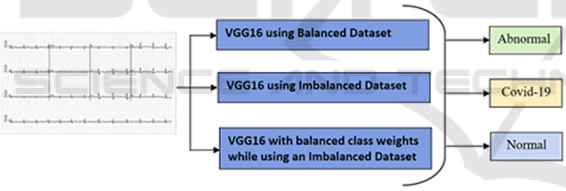

Figure 2: Experiment architecture.

Transfer learning is a deep learning approach or machine learning method where knowledge gained after developing and training a model for one task is transferred to another model for some other task. In other words, the parameters of a pre-trained model are reused to train another model that has a different task or dataset. Transfer learning helps achieve better performance even with fewer data at a much faster speed. Several pre-trained models are available for use. In this study, the VGG16 model is used. The updated CNN architecture of the VGG16 model is shown in Fig. 3.

The model is given an input image of size 500 x 700 and trained at 10 epochs. The training, testing and validation sizes are 80%, 20%, and 10% respectively. The VGG16 has a batch size of 4, an SGD optimizer with a learning rate of 0.001, and the images are in RGB color mode.

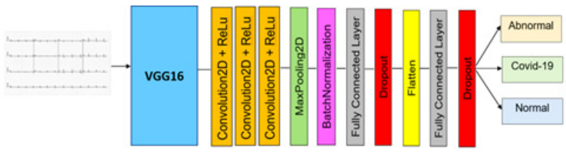

Figure 3: CNN Architecture for VGG16 network.

4 TRAINING MODEL

Three experiments were conducted as follows:

4.1 Training Model Using Balanced Dataset

The model is trained on 525 images and validated on 75 images belonging to Abnormal, Covid-19, and Normal ECG images. The accuracy and loss of the training set and validation set for each epoch are visually shown in Fig. 4 (A) and Fig. 4 (B) respectively and recorded numerically in Table 1.

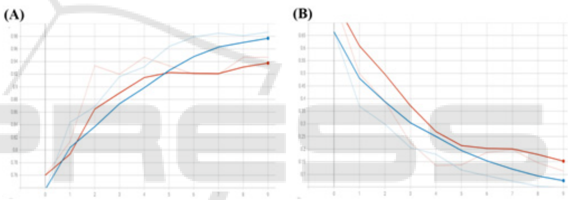

Figure 4: VGG16 Balanced Dataset: A) Training accuracy (Red) and Validation accuracy (Blue) VS Epochs. B) Training loss (Red) and Validation loss (Blue) VS Epochs.

Table 1: Balanced dataset accuracy and loss during training.

Epochs	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
1	73%	0.66	76%	0.79
2	84%	0.36	81%	0.49
3	86%	0.29	93%	0.38
4	91%	0.20	92%	0.22
5	93%	0.17	94%	0.13
6	96%	0.11	93%	0.13
7	97%	0.09	92%	0.18
8	98%	0.07	92%	0.19
9	98%	0.05	94%	0.14
10	98%	0.04	94%	0.11

4.2 Training Model Using Imbalanced Dataset

The model is trained on 233 Abnormal images, 103 Covid-19 images, and 691 Normal images and validated on 49 images from each category. The accuracy and loss of the training set and validation set for each epoch are visually shown in Fig. 5 (A) and Fig. 5 (B) respectively and recorded numerically in Table 2.

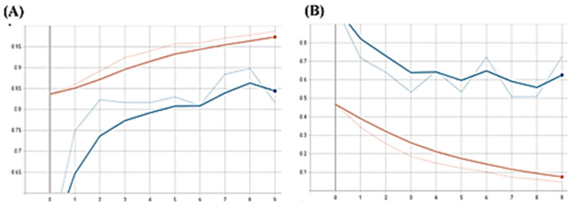

Figure 5: VGG16 Imbalanced Dataset: A) Training accuracy (Orange) and Validation accuracy (Blue) VS Epochs. B) VGG16 Imbalanced Dataset: Training loss (Orange) and Validation loss (Blue) VS Epochs.

Table 2: Imbalanced dataset accuracy and loss during training.

Epochs	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
1	83%	0.46	47%	0.99
2	85%	0.34	74%	0.71
3	89%	0.25	82%	0.64
4	92%	0.18	81%	0.53
5	93%	0.14	81%	0.64
6	95%	0.12	82%	0.53
7	95%	0.10	80%	0.72
8	97%	0.07	88%	0.50
9	97%	0.06	89%	0.51
10	98%	0.04	81%	0.72

The model is trained by manually passing class weight values during training which were calculated automatically using scikit learn's utils module. The accuracy and loss of the training set and validation set for each epoch when balancing class weights in an imbalanced dataset are visually shown in Fig. 6 (A) and Fig. 6 (B) respectively and recorded numerically in Table 3.

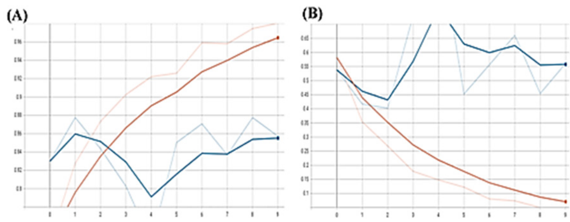

Figure 6: VGG16 balanced class weights (Imbalanced dataset): A) Training accuracy (Orange) and Validation accuracy (Blue) VS Epochs. B) VGG16 Imbalanced Dataset with Weighted Class: Training loss (Orange) and Validation loss (Blue) VS Epochs.

Table 3: Balancing class weights in imbalanced dataset.

Epochs	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
1	74%	0.58	82%	0.53
2	82%	0.35	87%	0.41
3	87%	0.26	84%	0.40
4	90%	0.17	80%	0.72
5	92%	0.14	74%	1.00
6	92%	0.12	85%	0.45
7	95%	0.08	87%	0.55
8	95%	0.07	83%	0.66
9	97%	0.05	87%	0.45
10	98%	0.04	85%	0.56

5 TESTING TRAINED MODELS

The confusion matrix in multiclass classification helps in calculating the following metrics for each category:

- True Positive (TP): The number of predictions where the classifier correctly predicts the positive class. For example, the classifier predicts an image to be of the covid-19 category which in fact was of the covid-19 category.
- False Positive (FP): The number of predictions where the classifier incorrectly predicts the negative class as positive. For example, the classifier incorrectly predicts an image to be of the covid-19 category which in fact was of the non-covid-19 category.
- False Negative (FN): The number of predictions where the classifier incorrectly predicts the positive class as negative. For example, the classifier incorrectly predicts an image to be of the non-covid-19 category which in fact was of the covid-19 category.
- True Negative (TN): The number of predictions where the classifier correctly predicts the negative class. For example, the classifier correctly predicts an image to be of the non-covid-19 category which in fact was of the non-covid-19 category.

Table 4: Classification report for all three experiments.

(A)	Balanced Dataset			
	Precision	recall	f1-score	support
Abnormal	0.86	0.84	0.85	50
Covid-19	1.00	1.00	1.00	
Normal	0.84	0.86	0.85	
Accuracy	0.90			150
(B)	Imbalanced Dataset			
	Precision	recall	f1-score	support
Abnormal	0.96	0.47	0.63	98
Covid-19	1.00	1.00	1.00	
Normal	0.65	0.98	0.78	
Accuracy	0.82			294
(C)	Balanced Class weights in Imbalanced Dataset			
	Precision	recall	f1-score	support
Abnormal	0.93	0.67	0.78	98
Covid-19	1.00	1.00	1.00	
Normal	0.74	0.95	0.83	

The confusion matrices are obtained when using the test set to make predictions on the models that were trained in Experiments 1, 2, and 3. The Experiment 1 model: trained with a balanced dataset, shown in Fig 7 A), correctly classifies 42 Abnormal images, 50 Covid-19, and 43 Normal images. It misclassifies 8 Abnromal images to be of the Normal category and 7 Normal images to be of the Abnormal category. The Experiment 2 model: trained with an imbalanced dataset, shown in Fig 7 B), correctly classifies 46 Abnormal images, 98 Covid-19 images, and 96 Normal images. It misclassifies 52 Abnormal images to be of the Normal category and 2 Normal images to be of the Abnormal category. The Experiment 3 model: trained by balancing class weights in an imbalanced dataset, shown in Fig 7 C), correctly classifies 66 Abnormal images, 98 Covid-19 images, and 93 Normal images. It misclassifies 32 Abnormal images to be of the Normal category and 5 Normal images to be of the Abnormal category.

The classification Report of all three experiments for the VGG16 model is summarized in Table 4. The report shows precision, recall, and f1-score for each category: Abnormal, Covid-19, and Normal. Support refers to the number of images present for each category, in the case of a balanced dataset, it is 50 images, which in total is 150 and for an imbalanced dataset, it is 98 images, which in total is 294. The classification report shows the accuracy of the model on the whole test set.

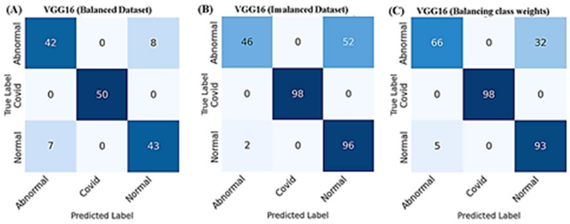

Figure 7: Confusion Matrix for VGG16: A) Using Balanced Dataset. B) Using Imbalanced Dataset. C) Balancing weights in Imbalanced Dataset.

6 CONCLUSION

In this study, we develop a VGG16 deep learning model using TensorFlow by modifying its final dense layers. The model is subsequently trained on an ECG Image Database to detect abnormalities and successfully distinguish among three categories: Abnormal heartbeat, COVID-19, and Normal ECG images. All images undergo gamma correction processing in a MATLAB environment.

We acknowledge the issue of data imbalance present within the dataset and propose a method to mitigate this problem by balancing the class weights. For experimental validation, two datasets are created: one balanced and one imbalanced. Three experiments are conducted to demonstrate differences in training, accuracy, and performance based on the distribution of image categories. Model evaluation is performed using a test set.

Results from Experiment 2 reveal that when an imbalanced dataset is used without any countermeasures, accuracy and performance decrease from 90% to 82%. Experiment 3 shows that the application of balanced class weights during training leads to a 5% increase in accuracy compared to Experiment 2, resulting in an overall accuracy of 87%.

Therefore, the method introduced is proven to significantly improve the model's performance for this specific dataset and may be applicable to similar classification problems. In conclusion, this study demonstrates that machine learning models are not only useful for image classification but also offer utility when data availability is a constraint.

ACKNOWLEDGMENT

The Authors would like to acknowledge Professor Haseeb Ahmad Khan's contribution in providing guidance to the first author. His expertise, mentorship, and constructive feedback were instrumental in shaping this research and ensuring its success.

REFERENCES

Attallah, O. (2022) "ECG-BiCoNet: An ECG-based pipeline for COVID-19 diagnosis using Bi-Layers of deep features integration", Computers in Biology and Medicine.

Darmawahyuni, A., Nurmaini, S., et al. (2022). "Deep learning-based electrocardiogram rhythm and beat features for heart abnormality classification", PeerJ Comput. Sci.

El-Rashidy, N., El-Sappagh, S. Islam, M. R. El-Bakry, H. M. and Abdelrazek, S. (2022). "End-to-end deep learning framework for Coronavirus (COVID-19) detection and monitoring," Electronics (Basel), vol. 9, no. 9, p. 1439.

Shahin, I., Bou Nassif, A. and Alsabek. M. (2021). "COVID-19 Electrocardiograms Classification using CNN Models." arXiv preprint arXiv:2112.08931.

Khan, H., Hussain, M. and Malik, M. (2021) "ECG Images dataset of Cardiac and COVID-19 Patients," Data in Brief, Vol. 34.

Ozturk, T., Talo, M. E., et al., (2020). "Automated detection of COVID-19 cases using deep neural networks with X-ray images," Comput. Biol. Med., vol. 121, no. 103792, p. 103792.

Raza, A. Tran, K. P., Koehl, L. and Li, S. "Designing ECG monitoring healthcare system with federated transfer learning and explainable AI," Knowl. Based Syst., vol. 236, no. 107763, p. 107763, 2022.

Tan, L. et al., (2021) "Toward real-time and efficient cardiovascular monitoring for COVID-19 patients by 5G-enabled wearable medical devices: a deep learning approach," Neural Comput. Appl., pp. 1–14.