Automated ECG Image Classification with InceptionV3

Antony M Gitau[1], Victor Ruto[1], Yuri Njathi[1], Lorna Mugambi[1], Victory A Sitati[1], Austin Kaburia[1]

[1]Centre for Data Science and Artificial Intelligence,
Dedan Kimathi University of Technology, Nyeri, Kenya

Abstract

Despite the rise of digital electrocardiogram (ECG) technology, paper-based ECGs continue to be prevalent, especially in underrepresented and underserved communities. This paper presents the DSAIL team's participation in the George B. Moody PhysioNet Challenge 2024 to develop an open-source algorithm for classifying ECG images. We fine-tuned a pre-trained InceptionV3 model on the PTB-XL dataset, comprising 21,799 12-lead ECG recordings, supplemented with synthetic ECG images from the ECG-Image-Kit. The model was trained using 80% of these images, reserving 20% for validation. Our choice of the InceptionV3 architecture leverages its capability to effectively capture local and global features, which is crucial for the inherent variability in ECG image patterns. The model achieved a validation macro F-measure score of 0.429 on a dataset accessible only to the organizers, securing 6th place on the official classification leaderboard. However, the algorithm struggled with mobile phone images of stained, deteriorated, and cleaned ECGs, yielding a low F-score of 0.08. In contrast, it performed significantly better on color scans of clean and deteriorated paper ECGs, achieving an F-score of 0.5. Although further improvements are necessary, neural network-based algorithms demonstrate promising potential for enhancing access to ECG-based diagnosis and cardiac care.

1. Introduction

Cardiovascular diseases (CVDs) are the most common cause of mortality globally, with an estimated 17.9 million deaths in 2019. CVDs account for 32% of all global deaths, of which 85% were due to heart attacks and strokes. Over three-quarters of these deaths occurred in low and middle-income countries, underscoring a stark disparity in health outcomes and access to care [1]. This prevalence highlights a critical need for scalable and efficient diagnostic tools, particularly in regions with limited healthcare resources. Electrocardiograms (ECGs) are crucial for CVD diagnosis, but traditional manual analysis is time-consuming and resource-intensive [2–5]. However, the recent advancements in artificial intelligence, particularly deep learning, have shown promise in automating ECG analysis, offering potential solutions for improving efficiency and accuracy.

Researchers have explored various methods for detecting atrial fibrillation (AF), a common arrhythmia, from ECG recordings. These include improved neural networks and classifiers, such as random forest classifiers and decision tree ensembles. The use of techniques such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), deep belief networks, and stacked autoencoders for patient-specific ECG classification has also been investigated, emphasizing the importance of multilevel features and advanced algorithms for accurate detection [5–9].

Studies have demonstrated the effectiveness of deep learning in ECG classification for diagnosing various cardiovascular conditions [10]. Preserving temporal information in ECG signals has been found crucial for accurate diagnosis. Specifically, long short-term memory (LSTM) networks, have achieved high accuracy in classifying different heartbeat types, showcasing the potential of automated feature extraction without the need for extensive preprocessing [11].

This paper builds upon this literature, presenting our contribution to the George B. Moody PhysioNet Challenge 2024 [12]. We leverage InceptionV3, a state-of-the-art deep learning architecture, for the classification of ECG images. The focus is on creating an open-source algorithm that can classify ECGs from both conventional paper printouts and digital platforms. Through the integration of CNN-based technologies, this research seeks to provide robust, scalable, and precise ECG analysis tools, bridging the gap between traditional analysis techniques and modern digital solutions. These innovations hold particular promise for enhancing cardiac care in resource-limited environments, making accurate diagnostics more accessible.

2. Method

We fine-tuned an Inception V3 [13] model, a convolutional neural network architecture, to classify ECG images

 ISSN: 2325-887X DOI: 10.22489/CinC.2024.498

into 11 classes. This work was done in PyTorch and the code is open-sourced and published on GitHub [1].

2.1. Data

We utilized waveform and label data from the publicly available PTB-XL dataset [14–17], which comprises 21,799 12-lead ECG recordings.

To augment the training dataset, we employed a synthetic ECG image generator from ECG-Image-Kit [18,19], to generate synthetic ECG images from time-series ECG data, including various artifacts such as one shown in figure 1. The generated images were then used to fine-tune the InceptionV3 model and the labels.

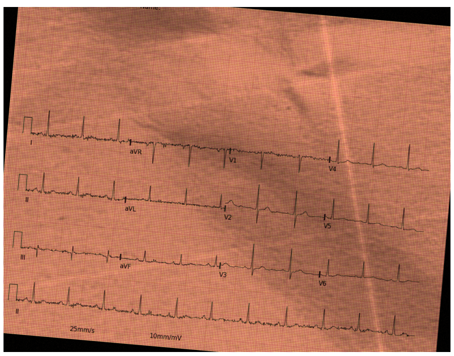

Figure 1. An example of an ECG image generated using ECG-Image-Kit

2.1.1. Data pre-processing

All input images were resized to uniform dimensions of 425x550 pixels to standardize the input size across the dataset. Additionally, we converted all images to the RGB colour space, ensuring consistent colour representation. The dataset was divided into distinct training and validation sets, with the training data comprising 17,418 samples and the validation data divided into two folds: fold 9 with 2,183 samples and fold 10 with 6,198 samples. Each sample within these folders was structured to include a header, a PNG image file, and an associated data file.

2.2. Classification model development

We employed the InceptionV3 architecture as illustrated on figure 2. The InceptionV3 model is well-regarded for its optimization of depth and width, making it particularly effective for image recognition tasks. It was originally developed to address challenges where computational resources are limited. InceptionV3 incorporates multiple kernel sizes within its convolutional layers, enabling it to capture spatial hierarchies in images across different scales. This architecture enhances its ability to recognize complex patterns with improved efficiency.

For the ECG image classification task, we adapted InceptionV3 by adding another linear layer on top of the model to allow for the classification of the 11 types of cardiac conditions represented in our dataset. We conducted a fine-tuning process of the pre-trained model using the PTB-XL dataset, enhanced with synthetic images generated by the ECG-Image-Kit to represent a wider variety of ECG signal conditions, including artifacts.

The training was conducted over 20 epochs using the Adam optimizer with a weight decay of 0.0001 and binary cross-entropy loss to handle the multi-label classification task. The model's efficacy was validated through a stratified 10-fold cross-validation process, allowing us to assess its performance across different subsets of data, optimizing generalizability and reducing overfitting risks. The experiment was run on a Kaggle notebook with an accelerator on for 4 hours.

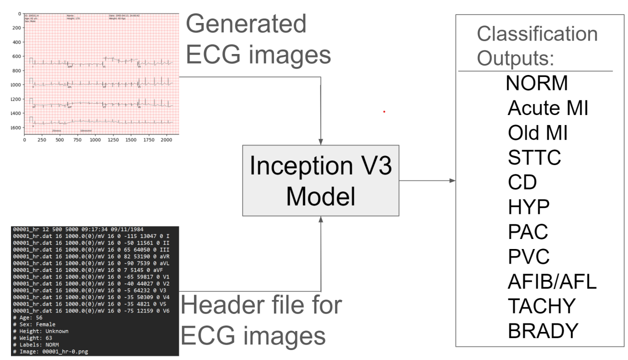

Figure 2. Workflow of ECG Image Classification Using InceptionV3.

2.2.1. Testing other architectures

Furthermore, we finetuned additional architectures: EfficientNet [20], EfficientNetV2 [21], Swin Transformer V2 [22], and ResNeXt [23] chosen based on their documented performance metrics and parameter efficiency as reported in PyTorch's documentation. These models were assessed under similar training conditions as InceptionV3 described above to understand their respective efficacies in classifying ECG images. The variations in architecture offered insights into the trade-offs between computational demand and diagnostic performance, crucial for deploying these models in resource-constrained environments. Comparative performance metrics of these models are discussed in the results section.

[1] GitHub link of the code `https://github.com/vickruto/physionet-challenge-2024-dsail-full-dataset.git`

3. Results

The performance of the evaluated models is summarized in Table 1, which includes F-scores from our local validation folds, batch size, and GPU time on a P100 GPU in Kaggle. Table 2 shows a hidden dataset's official evaluation from PhysioNet organizers.

EfficientNetB0 achieved the highest F-score in our local validation set. However, it did not surpass InceptionV3 on the organizers' training set: An F score of 0.359 against Inception's of 0.429. Other variants of EfficientNet architecture performed relatively in the same range.

Conversely, Swin Transformer V2 recorded the lowest F-score. This underperformance could be attributed to the model's complexity and the need for more extensive fine-tuning to adapt to the specific challenges presented by ECG image classification.

Table 1. Different models and validation results and the various hyperparameters we tested

Model	Batch size	F-score (fold 9/10)	GPU time (hrs)
InceptionV3	32	0.610, 0.598	4.8
EfficientNetB0	32	0.630, 0.622	4.7
EfficientNetB5	8	0.565, 0.582	6.1
EfficientNetV2s	24	0.601, 0.582	5.3
EfficientNetV2m	8	0.555, 0.545	6.1
Swin	8	0.056, 0.056	7.9
ResNeXt	8	0.533, 0.526	5.0

From the submissions to the organizers, table 2 is the performance of the InceptionV3 against different types of ECG images.

Table 2. F-measure scores of the InceptionV3 for different types of ECG images

ECG Image Type	**F-measure**
Leaderboard F-measure	0.429
Color scans of clean papers	0.502
Black-and-white scans of clean papers	0.118
Mobile phone photos of clean papers	0.072
Mobile phone photos of stained papers	0.080
Mobile phone photos of deteriorated papers	0.078
Color scans of deteriorated papers	0.463
Black-and-white scans of deteriorated papers	0.093
Screenshots of computer monitor	0.104
Official Rank	6 out of 16

4. Discussion and Conclusion

This study illustrates the potential challenges of utilizing deep learning models for classification of ECG images. InceptionV3, among other tested models, has shown promising results, suggesting that deep learning could significantly enhance diagnostic accuracies in cardiac care, especially in resource-limited settings. The observed variability in model performance across different architectures underscores the absence of a universal solution, necessitating careful selection of model architecture tailored to specific clinical needs and dataset characteristics.

Overall, integrating AI into healthcare, through projects like this one offers a promising avenue to address some of the pressing challenges in global health, particularly in enhancing the accessibility and quality of diagnostic services.

5. Future Work

Future work could focus on exploring the integration of ensemble techniques and expanding the dataset variability to include more diverse and larger datasets to enhance the generalizability of the models. Additionally, we plan to explore more advanced pre-processing techniques for ECG images. For instance, we experimented with detecting all leads using YOLOv8 [24] with oriented bounding boxes (OBB). After cropping the detected leads, we applied Otsu's method to binarize the images, and then combined the binarized leads into a new composite image, as shown in Figure 3. While we ran out of time to fully implement and test this approach before the challenge submission, we believe that this data pre-processing step could significantly enhance the performance of fine-tuned models.

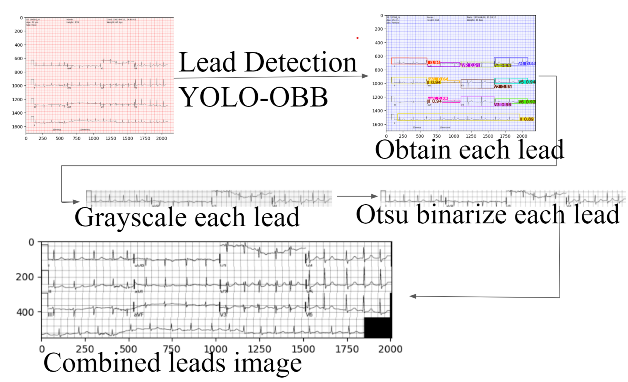

Figure 3. Future pre-processing steps

Acknowledgments

We are grateful to the director of the Centre for Data Science and Artificial Intelligence (DSAIL) at Dedan Kimathi University of Technology (DeKUT), Professor Ciira wa Maina for his support during our participation in the PhysioNet Challenge 2024.

References

[1] World Health Organization. Cardiovascular diseases (CVDs), 2021. 28th August 2024.

[2] Fatma Murat and Ozal Yildirim and Muhammed Talo and U. Baloglu and Y. Demir and U. R. Acharya and U. R. Acharya and U. R. Acharya. Application of deep learning techniques for heartbeats detection using ECG signals-analysis and review. Computers in biology and medicine 2020;120:103726.

[3] Singstad BJ, Gitau AM, Johnsen MK, Ravn J, Bongo LA, Schirmer H. Phonocardiogram classification using 1-dimensional inception time convolutional neural networks. In 2022 Computing in Cardiology (CinC), volume 498. 2022; 1–4.

[4] Tison GH, Zhang J, Delling FN, Deo RC. Automated and interpretable patient ECG profiles for disease detection, tracking, and discovery. Circulation Cardiovascular Quality and Outcomes 2019;12(9):e005289.

[5] Mewada H. D-wavelet encoded deep CNN for image-based ECG classification. Multimed Tools Appl 2023; 82:20553–20569. URL https://doi.org/10.1007/s11042-022-14302-z.

[6] Khan F, Yu X, Yuan Z, Rehman A. ECG classification using 1-D Convolutional Deep Residual Neural Network. PLOS ONE 2023;18.

[7] Liu X, Wang H, wei Li Z, Qin L. Deep learning in ECG diagnosis: A review. Knowl Based Syst 2021;227:107187.

[8] A. Y, M. O, Q. K, S. E. Deep learning for ECG arrhythmia detection and classification: an overview of progress for period 2017–2023. Frontiers in Physiology 2023;14:1246746. URL https://doi.org/10.3389/fphys.2023.1246746.

[9] Zhang A, Yang X, Li Tea. Classification method of ECG signals based on RANet. Cardiovasc Eng Tech 2024;URL https://link.springer.com/article/10.1007/s13239-024-00730-5#citeas.

[10] Narotamo H, Dias M, Santos R, Carreiro AV, Gamboa H, Silveira M. Deep learning for ECG classification: A comparative study of 1d and 2d representations and multimodal fusion approaches. Biomedical Signal Processing and Control 2024;93:106141. ISSN 1746-8094. URL https://www.sciencedirect.com/science/article/pii/S174680942400199X.

[11] Harrane S, Belkhiri M. Classification of ECG heartbeats using deep neural networks. Res Biomed Eng 2021; 37:737–747. URL https://link.springer.com/article/10.1007/s42600-021-00176-x?fromPaywallRec=true#citeas.

[12] Reyna MA, Deepanshi, Weigle J, Koscova Z, Elola A, Seyedi S, Campbell K, Clifford GD, Sameni R. Digitization and classification of ECG images: The George B. Moody PhysioNet Challenge 2024. PhysioNet Challenge 2024;51:1–4.

[13] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016; 2818–2826.

[14] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, George B. Moody, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 2000;101(23):e215–e220. URL https://doi.org/10.1161/01.CIR.101.23.e215.

[15] Reyna MA, Deepanshi, Weigle J, Koscova Z, Campbell K, Shivashankara KK, Saghafi S, Nikookar S, Motie-Shirazi M, Kiarashi Y, Seyedi S, Clifford GD, Sameni R. ECG-image-database: A dataset of ECG images with real-world imaging and scanning artifacts; a foundation for computerized ECG image digitization and analysis, 2024. URL https://arxiv.org/abs/2409.16612.

[16] Wagner P, Strodthoff N, Bousseljot RD, Kreiseler D, Lunze FI, Samek W, Schaeffter T. PTB-XL: A large publicly available ECG dataset. Scientific Data 2020;7:154. URL https://doi.org/10.1038/s41597-020-0495-6.

[17] Wagner P, Strodthoff N, Bousseljot R, Samek W, Schaeffter T. PTB-XL, a large publicly available electrocardiography dataset (version 1.0.3), 2022. URL https://doi.org/10.13026/kfzx-aw45.

[18] Shivashankara KK, et al. ECG-image-kit: A toolkit for synthesis, analysis, and digitization of electrocardiogram images, 2024. URL https://github.com/alphanumericslab/{ECG}-image-kit. Accessed: 2024-08-22.

[19] Shivashankara KK, Deepanshi, Shervedani AM, Reyna MA, Clifford GD, Sameni R. ECG-Image-Kit: A synthetic image generation toolbox to facilitate Deep Learning-based electrocardiogram digitization. Physiological Measurement 2024;.

[20] Tan M. EfficientNet: Rethinking model scaling for Convolutional Neural Networks. arXiv preprint arXiv190511946 2019;.

[21] Tan M, Le Q. EfficientNetV2: Smaller models and faster training. arXiv preprint arXiv210400298 2021;.

[22] Liu Z, Hu H, Lin Y, Yao Z, Xie Z, Wei Y, Ning J, Cao Y, Zhang Z, Dong L, et al. Swin transformer v2: Scaling up capacity and resolution. In Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition. 2022; 12009–12019.

[23] Xie S, Girshick R, Dollár P, Tu Z, He K. Aggregated Residual transformations for Deep Neural Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017; 1492–1500.

[24] Jocher G, Chaurasia A, Qiu J. Ultralytics YOLOv8, 2023. URL https://github.com/ultralytics/ultralytics.

Address for correspondence:

Antony M. Gitau

Private Bag 10143, Dedan Kimathi University of Technology, Nyeri, Kenya

antonym.gitau9@gmail.com