Hybrid Approach for Electrocardiogram Image Classification: Residual Network and Random Forest

Noor Qaqos[1], Ekenedirichukwu N Obianom[2], Chintan Patel[1], Jingzhi Gong[1], Shamsu Idris Abdullahi[1], Fan Feng[1], Abdulhamed M Jasim[1], Abdulmalik Koya[2,4], Zaid A Abod[3], G Andre Ng[2,4], Xin Li[1,4]

[1]University of Leicester, School of Engineering, Leicester, UK
[2]University of Leicester, Department of Cardiovascular Sciences, Leicester, UK
[3]University of Leicester, School of Computing and Mathematical Sciences, Leicester, UK
[4]National Institute for Health Research Leicester, UK

Abstract

This paper presents a hybrid approach for classifying electrocardiogram (ECG) papers in the George B. Moody PhysioNet Challenge 2024 by integrating Deep Residual Networks (ResNet) and Random Forest (RF) algorithms. Two distinct pipelines were developed to enhance classification accuracy and efficiency. The first pipeline (RRF) utilizes ResNet50 to directly extract features from ECG paper, which are subsequently classified using an RF model. The second pipeline (YRRF) employs the YOLO object detection model to segment ECG papers into 12 individual leads before feature extraction with ResNet50, followed by classification with RF. The accuracy, F1 score, and area under the receiver operating characteristic curve (AUROC) were 79.8%, 78.26%, and 0.89, and 73.74%, 71.11% and 0.81 for the RRF and YRRF models respectively. These were trained on two classes (normal and abnormal). This demonstrated that the lead-specific feature extraction approach was not helpful in this study because it caused a reduction in the performance of the hybrid model. Our team, Leicester Fox had an F_measure score of 0.068 for an official phase using the RRF classification model when trained on 12 classes. This study underscores the effectiveness of hybrid modelling techniques in medical image analysis and sets the stage for future research in automated cardiovascular diagnostics.

1. Introduction

The electrocardiogram (ECG) is a non-invasive way of measuring the electrical activity of the human heart and represents an essential pre-testing tool that helps in the diagnosis of different cardiovascular diseases (CVDs). Since Willem Einthoven invented it in 1895, ECG devices have developed from big equipment to portable ones, which were introduced by General Electric in 1927[1]. By 1948, these ECG devices were able to print ECG waveforms on paper [1]. Recently, researchers have developed several algorithms which can interpret the ECG waveforms, and different companies have introduced digital ECG devices that can record the digital forms of the ECG signals. These developments have helped to enhance and improve the possibility of diagnosing of CVDs based on ECG waveforms [2]. Despite the advent of these digital systems, paper ECG has been a staple to care for patients in the world. This has been around for a century and it is still common, with billions existing today in the Global South. These ECG records reflect the extent of the diversity and development of CVDs across populations over time [3]. However, organisations with large digital data banks that provide accessibility to their data require technologies or specific licences to access and process this data. This creates a barrier for researchers or health institutions, who may lack the resources to participate which may lead to difficulties in the ability to use these vital data and share it for medical search and innovations. Several algorithms have been suggested for ECG paper classification, including those that work with grayscale and colour images [4]. Classical image processing techniques, however, are often sensitive to low-quality or distorted ECG papers. Recently, deep learning-based algorithms have been proposed to improve the performance of ECG paper classification models [3, 5].

Our team, Leicester Fox, proposed work aims to build and compare two approaches for classifying ECG papers in the George B. Moody PhysioNet Challenge 2024 [6-9]. The first approach involves directly extracting features from the ECG paper, while the second involves cropping the ECG images into 12 individual segments (representing the 12 leads) before feature extraction.

 ISSN: 2325-887X DOI: 10.22489/CinC.2024.343

2. Materials and Methods

2.1. ECG Papers Dataset

The PTB-XL ECG dataset was used for this analysis. It's comprised of a substantial database containing 21,799 clinical 12-lead ECG obtained from 18,869 patients. Each ECG recording has a duration of 10 seconds [6, 9]. The unprocessed signal data was captured and saved in a compressed format exclusive to the organization. It offered 12 leads (I, II, III, AVL, AVR, AVF, V1, V2, V3, V4, V5 and V6) for all signals, with the reference electrode on the right arm. The relevant demographic information such as (age, sex, weight, and height) was stored in a database. The challenge data was annotated. For the analysis, each ECG record was converted into a paper image for training and testing the model [3, 10].

2.2. Pipeline Architecture

Two pipelines were adopted for the classification process. These pipelines included the image cropping capability of You Only Look Once (YOLO) model, the feature extracting capabilities of Deep Residual Network (ResNet), and the classifying efficacy of Random Forest (RF). YOLO is a real-time object detection system [11]. It is a single neural network that divides an image into sections and predicts the bounding boxes and probabilities of each section. This assists in recognising which section represents which leads. ResNet is a convolutional neural network architecture that utilises residual connections to skip certain layers to bypass the potentially diminishing effect of deep layers [12]. Particularly, ResNet50 which is composed of 50 layers including convolution, pooling, and fully connected layers, pre-trained on a large-scale image dataset was used for this work. Lastly, the RF classifier is an ensemble learning method that utilizes multiple decision trees to classify data [13]. It leverages the power of averaging and feature randomness and can handle high-dimensional data. This makes it able to capture complex relationships between features and class labels.

The first pipeline (denoted by ResNet50-RF (RRF)) proposed was to directly extract features from the ECG images using ResNet50. The ResNet50 was re-trained to extract features from the ECG papers. The features extracted were used to train RF model to classify the papers (Figure 1A). For the second pipeline (denoted by YOLO-ResNet50-RF (YRRF)), we used the YOLO model to crop the ECG images into 12 images, representing the 12 leads extracted from the image. Features were then extracted separately. This was done to ensure that features were extracted from only the relevant parts of the images. The features extracted from each lead were then combined to form a unified feature set for an image, which could be used to train the RF model for classification (Figure 1B).

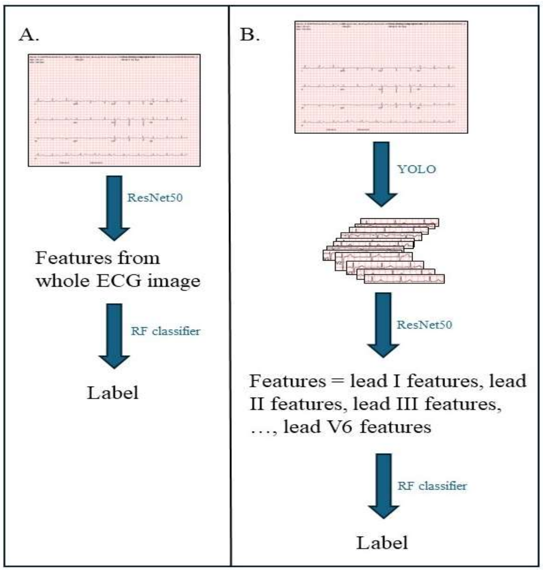


Figure 1: Pipeline for the classification of ECG images. A. Feature extraction without cropping the image (RRF). B. Feature extraction with cropping the image into various leads (YRRF).

2.3. Performance Metrics

For comparison, accuracy, F1 score and Area Under Receiver Operating Characteristic Curve (AUROC) were used. They are defined mathematically below.

$$\text{Accuracy}=\frac{TP+TN}{TP+TN+FP+FN} \quad \ldots \quad (1)$$

$$\text{Precision}=\frac{TP}{TP+FP} \quad \ldots \quad (2)$$

$$\text{Recall}=\frac{TP}{TP+FN} \quad \ldots \quad (3)$$

$$\text{FPR}=\frac{FP}{FP+TN} \quad \ldots \quad (4)$$

$$\text{F1 Score}=2*\frac{\text{Precision*Recall}}{\text{Precision+Recall}} \quad \ldots \quad (5)$$

$$\text{AUROC}=\int_0^1 \text{Recall}(t)d(\text{FPR}(t)) \quad \ldots \quad (6)$$

Where:

TP = True Positives (correctly predicted positive instances).
TN=True Negatives (correctly predicted negative instance).
FP=False Positives (incorrectly predicted positive instances).
FN=False Negatives (incorrectly predicted negative instances).

3. Results

One thousand ECG clean papers were used to train and test the RRF and YRRF models on two classes (normal and abnormal). The data was split into 90% as training set and 10% as testing set (unseen papers) to assess the performance of the models.

3.1. Performance Evaluation

To evaluate the performance of the proposed approaches, five metrics including accuracy, sensitivity, specificity, precision, and F1_score were used. The accuracy for RRF and YRRF was 79.8% and 73.74%, respectively. The other performance metrics ranged between (70-90%) and (62-86%) for RRF and YRRF models, respectively. Figure 2 shows the performance assessment of the methods. The official phase results for the RRF classification model are shown in the table 1.

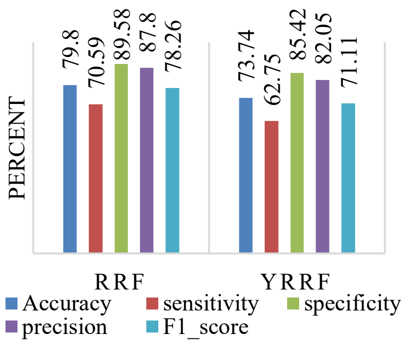


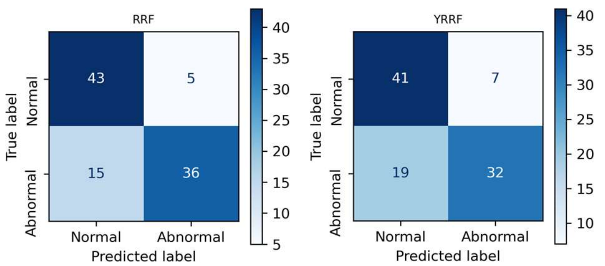
Figure 2: Performance evaluations for the models

Table 1. Official results for the classification RRF model.

Task	Score	Rank
Classification	F_measure: 0.068	15/16

3.2. CM and the ROC

The confusion matrix (CM) is the fundamental tool for visualizing and evaluating the classification model's performance. A high performance was obtained from the RRF in comparison to the YRRF approach (Figure 2). In addition, RRF method made the balance between the normal and abnormal subjects. Figure 3 shows the CMs for the proposed methods. A receiver operating characteristic (ROC) was also used to evaluate the model performance by representing the trade-off between TPR and FPR at various decision thresholds. Figure 4 shows the ROCs and the area under curves (AUCs) for both classifiers. The AUC values for RRF and YRRF models were 0.89 and 0.81, respectively. It can be seen the AUROC for the RRF model was higher than YRRF by around 0.08. Moreover, the processing time of RRF model was shorter than YRRF for training the results which took around thrice the amount of time while running on the same computing system.



(A) (B)

Figure 3: CMs for (A) RRF and (B) YRRF model

4. Discussions

The low performance of the YRRF model might be related to the high number of features that resulted from cropping the ECG papers to the 12 small segments (12 leads). Therefore, the number of features extracted using the YRRF was 12 times more than the RRF. The RF classifier is prone to overfitting when the number of features exceeds the number of samples (ECG papers) [14]. This might be the main reason for the higher performance for the RRF over the YRRF model.

The official score (for the George Moody Physionet Challenge), which was trained to classify the 12 classes, had an F1 score of 0.068 for the submitted RRF classification model. This lower score might be due to the larger number of classes being predicted and the variation in noise level of the ECG paper images.

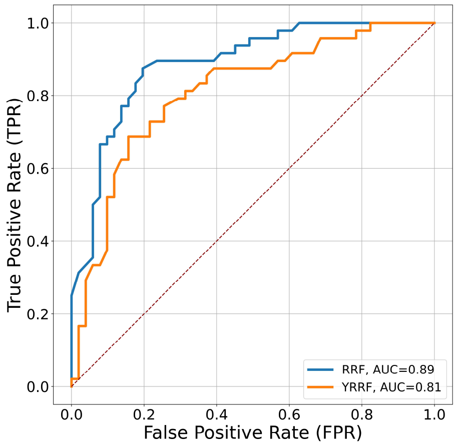


Figure 4: The ROCs for the proposed models.

5. Conclusions

Two proposed approaches (RRF and YRRF) were used to classify the ECG clean papers into normal and abnormal classes. The high performance was obtained from the RRF model with an accuracy, sensitivity, specificity, precision and F1_score of 79.8%, 70.59%, 89.58%, 87.8%, and 78.26%, respectively. Furthermore, the RRF took a third of the processing time compared to YRRF. Also, we obtained an AUROC of 0.89 for the RRF model, which was higher than AUROC of the YRRF model by 0.08. Our official score was 0.068 using the RRF classification model. Further study should be considered with other types of noisy ECG papers and different machine learning classifiers.

Acknowledgments

We acknowledge our sponsors (1) Higher Committee for Education Development, Iraq, (2) Petroleum Technology Development Fund (PTDF), Nigeria, (3) Future100 Scholarship Scheme, University of Leicester, UK, (4) The British Heart Foundation.

References

[1] T. Syeda-Mahmood, D. Beymer, and W. Fei, "Shape-Based Matching of ECG Recordings." pp. 2012-2018.

[2] P. W. Macfarlane, and J. Kennedy, "Automated ECG Interpretation—A Brief History from High Expectations to Deepest Networks," *Hearts (Basel, Switzerland),* vol. 2, no. 4, pp. 433-448, 2021.

[3] K. K. Shivashankara, Deepanshi, A. Mehri Shervedani, G. D. Clifford, M. A. Reyna, and R. Sameni, "ECG-Image-Kit: a Synthetic Image Generation Toolbox to Facilitate Deep Learning-Based Electrocardiogram Digitization," *Physiological measurement,* vol. 45, no. 5, pp. 55019, 2024.

[4] L. Ravichandran, C. Harless, A. J. Shah, C. A. Wick, J. H. McClellan, and S. Tridandapani, "Novel Tool for Complete Digitization of Paper Electrocardiography Data," *IEEE journal of translational engineering in health and medicine,* vol. 1, pp. 1800107-1800107, 2013.

[5] S. Mishra, G. Khatwani, R. Patil, D. Sapariya, V. Shah, D. Parmar, S. Dinesh, P. Daphal, and N. Mehendale, "ECG Paper Record Digitization and Diagnosis Using Deep Learning," *Journal of medical and biological engineering,* vol. 41, no. 4, pp. 422-432, 2021.

[6] P. Wagner, N. Strodthoff, R.-D. Bousseljot, D. Kreiseler, F. I. Lunze, W. Samek, and T. Schaeffter, "PTB-XL, a Large Publicly Available Electrocardiography Dataset," *Scientific Data,* vol. 7, no. 1, 2020.

[7] N. Strodthoff, T. Mehari, C. Nagel, P. J. Aston, A. Sundar, C. Graff, J. K. Kanters, W. Haverkamp, O. Dössel, A. Loewe, M. Bär, and T. Schaeffter, "PTB-XL+, a Comprehensive Electrocardiographic Feature Dataset," *Scientific Data,* vol. 10, no. 1, 2023.

[8] M. A. Reyna, J. Weigle, Z. Koscova, K. Campbell, S. Seyedi, A. Elola, A . Bahrami Rad, A. J. Shah, N. K. Bhatia, G. D. Clifford, R. Sameni. , "Digitization and Classification of ECG Images: The George B. Moody PhysioNet Challenge," vol. 51, pp. 1-4, 2024.

[9] M. A. Reyna, J. Weigle, Z. Koscova, K. Campbell, K. K. Shivashankara, S. Saghafi, S. Nikookar, M. Motie-Shirazi, Y. Kiarashi, and S. Seyedi, "ECG-Image-Database: A Dataset of ECG Images with Real-World Imaging and Scanning Artifacts; A Foundation for Computerized ECG Image Digitization and Analysis," *arXiv preprint arXiv:2409.16612*, 2024.

[10] "ECG-Image-Kit: A Toolkit for Synthesis, Analysis, and Digitization of Electrocardiogram Images," https://github.com/alphanumericslab/ecg-image-kit.

[11] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection." pp. 779-788.

[12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning For Image Recognition." pp. 770-778.

[13] L. Breiman, "Random Forests," *Machine Learning,* vol. 45, no. 1, pp. 5-32, 2001/10/01, 2001.

[14] L. Yu, and H. Liu, Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution, 2003.

Address of correspondences:
Dr. Xin Li, and Noor Qaqos
University of Leicester, United Kingdom.
xin.li@leicester.ac.uk, nnq2@leicester.ac.uk