Scanned ECG Arrhythmia Classification Using a Pre-trained Convolutional Neural Network as a Feature Extractor

Hanadi Aldosari[1,4], Frans Coenen[1], Gregory Y. H. Lip[2], and Yalin Zheng[2,3]

[1] Department of Computer Science, University of Liverpool, Liverpool, UK;
[2] Liverpool Centre for Cardiovascular Science, University of Liverpool and Liverpool Heart & Chest Hospital, Liverpool, UK;
[3] Department of Eye and Vision Science, University of Liverpool, Liverpool, UK
[4] College of Computer Science and Engineering, Taibah University, Madinah, Saudi Arabia
{H.A.Aldosari, Coenen, Gregory.Lip, Yalin.Zheng}@liverpool.ac.uk,
hdosari@taibahu.edu.sa

Abstract. The classification of cardiovascular diseases using ECG data is considered. It is argued that to obtain a satisfactory classification features should be extracted from ECG images in their entirety, instead of translating the image into a 1D time series and only considering a small number of features as is the current common practise. The presented approach used a pre-trained Convolutional Neural Network (CNN) as a features extractor, followed by the application of T-distributed Stochastic Neighbour Embedding (T-SNE) to find the best discriminant features to perform ECG classification. The motivation using a pre-trained CNN model is that available ECG data sets tend to be limited in size; typically insufficient for training a bespoke deep learning model for feature extraction. Using a pre-trained CNN this challenge can be addressed. The features were extracted from the fully connected layers immediately preceding the softmax layer. The use of several pre-trained CNNs is reported on: VGG16, InceptionV3, and ResNet50. The operation of the proposed approach was also compared with recent relevant published approaches. A best AUC value of 0.960 was produced using the proposed approach; while the best alternative approach, out of those considered, produced an AUC of 0.932.

Keywords: ECG Classification, Convolutional Neural Networks, SVM Classifier, KNN Classifier

1 Introduction

Cardiovascular disease (CVD) generates a huge amount of biomedical and clinical data as a part of patient care. That data requires automated abstraction and manual analysis to be easily utilized by cardiovascular researchers, practitioners and doctors. Strategies and techniques that employ Machine Learning (ML) and Deep Learning (DL) have become essential for improving the cardiologists' work processes and performance, including making early diagnoses of diseases [14]. In recent decades, there has been a significant growth of interest in the automated classification and prediction of various cardiovascular diseases based on ECG data [5, 17, 35]. Much of the available ECG data tends to be in paper "print out" format; this is especially the case with respect to longitudinal studies [23]. It is only in recent times that ECG machines have been able to output digital format ECG data.

The typical process for processing “paper” ECG data is to first scan the paper format data so that a digital image is available; although, depending on the nature of the scanning, the image quality may be affected. The digitised ECG images can be used directly as the input to a selected ML or DL algorithm. However, this entails a high computational cost, in many cases making this approach prohibitive. To reduce this computational cost a frequent practice is to translate the 2D ECG image data into a 1D time series format, and then select features from the resulting time series [26, 30, 36, 39, 49].

The adopted feature extraction process is usually founded on a small number of global characteristics of ECG data; notably the amplitude and interval values of what are referred to as the P, QRS and T (P-QRS-T) waves[1] [20, 29, 38, 46]. These are the same key points that practitioners consider when conducting visual analysis of ECG signals.

The translation from 2D to 1D tends to introduces further irregularities and information loss. However, the main disadvantage of P-QRS-T style techniques is that they only focus on a small set of features. The alternative, and that considered in this paper, is to extract discriminatory features from the 2D scanned ECG image data without transformation to a 1D format. Thus maintaining the computational advantage of feature-based ECG classification, whilst circumventing the disadvantages associated with the P-QRS-T style 1D techniques.

The question is then how best to extract the desired features from the raw ECG data. The Convolutional Neural Network (CNN) is the state of art DL technique for working with image data. However, to train a CNN requires a considerable amount of data. Generally speaking the amount of ECG data available for CVD classification is insufficient to train a CNN for feature extraction, especially where practitioners are interested in a particular form of CVD. The solution here is to use a pre-trained CNN model [9, 16, 21, 31, 33, 40, 44, 52]. In this paper three different pre-trained CNN models, to extract features from ECG scanned image data, are considered: (i) ResNet50, (ii) VGG16 and (iii) InceptionV3. These models were selected because of their high robustness and proven efficiency with respect to ECG data applications. However, using these techniques a large number of features will be identified. If all the extracted features were to be considered any computational advantage that might have been gained would be lost. A feature selection process should therefore be applied so as to identify features that are the best discriminators of class. With respect to the work presented in this paper use of the T-distributed Stochastic Neighbour Embedding (T-SNE) dimension reduction technique is advocated. Once an appropriate subset of features has been identified a classification model can be built. In this paper the use of Support Vector Machine (SVM) and k-Nearest Neighbour (KNN) classification were considered.

A further advantage of using features, regardless of the nature of these features, is that they can be augmented with additional features obtained from elsewhere, for example features extracted from Electronic Patient Record (EPR) data. As a consequence it may be the case that a ”better” classification results. This idea is also explored in this paper.

The application focus for the work presented in this paper is the detection of abnormal ECGs. ECGs that feature some form of CVD indicated by an irregular, and often an unusually fast, heart rate. The remainder of this paper is structured as follows. A review of relevant existing work is presented, in Section 2. The proposed approach is then described in Section 3, and the associated evaluation in Section 4. The paper is concluded, in Section 5, with a summary of the main findings and some ideas for future work.

[1] The P wave indicates atrial depolarization, the QRS wave ventricular depolarization and the T wave ventricular re-polarization

2 Previous Work

The analysis and interpretation of cardiovascular activity using ECG data is acknowledged, by medical practitioners, to be a challenging task. According to [41], the interpretation of ECGs can only be done by practitioners with extensive prior knowledge and skills. It is also very time-consuming. Consequently the automated interpretation of ECG data using the tools and techniques of ML and DL is seen as desirable. Examples, with respect to the classification of a variety of CVDs using ML and DL technology, can be found in [5, 17, 19, 24, 35]. Many of these methods report good result.

As noted in the introduction to this paper ECG data, traditionally, comes in a "paper print out" format. For ML and DL algorithms to be applied the paper format data needs to be scanned. A process that entails the introduction of: (i) irregularities of various kinds, (ii) the inclusion of spurious information and noise, and (iii) information loss. The quality of the scanned images is thus, in many cases, in question; however this is accepted as an unavoidable side effect if ML and DL techniques to existing is to be applied to, paper format, ECG data. The work presented in this paper is directed at the use of scanned ECG image data. The normal process is to convert to a 1D time series conceptualisation so that, typically, P-QRS-T wave form-based ML and DL techniques can be applied [26, 30, 36, 39, 49]. A range of ECG digitisation tools have been used covert 2D ECG data into time series formats [7, 8, 13, 18, 30, 36, 43]. However, as already noted, the P-QRS-T approach features the disadvantage that only a small part of the available data is utilised.

Although the work presented in this paper is directed at paper format ECG data it is worth noting in this literature review that in more recent times ECG machines that produce digitised output have become increasingly available, although much existing data is still in the paper format. This avoids many of the issues associated with the scanning of paper ECG data. The digital output can be in two forms 2D digital image output or 1D time series output. The 1D time series format offers the advantage that P-QRS-T style 1D techniques can be applied directly. In the case of 2D output the practice is to first translate into a 1D format, and then apply the tried and tested P-QRS-T waveform approach to feature selection. However, as noted in the introduction to this paper, P-QRS-T style techniques feature the disadvantage that much of the data is ignored. For any kind of longitudinal study paper format ECG data remains the dominant format.

In this paper the focus is on using state-of-the-art technique to extract features from 2D ECG data using Convolution Neural Networks (CNNs), and to then we used the power of DL and ML classification algorithms to classify the ECG data. CNNs have been successfully applied to classify ECG data in the 1D, time series, context [1, 32, 42, 53] (although RNNs are more common). Although CNNs provide promising results for 1D time series data, better results have been reported when CNN are applied to 2D medical image data [34, 45]. Accordingly,it is noteworthy that some researchers working with 1D time series ECG data, directly produced by a ECG machine, have considered converting this time series data into a 2D image format so the CNNs can be applied [10, 11, 25, 27, 28, 37, 50].

However, as noted in the introduction to this paper, training CNNs for ECG classification require large amounts of data to achieve a desirable performance. The state-of-art solution is to use some form of Transfer Learning (TL) where a pre-trained model is adopted and fine tuned using additional data. The utilization of TL allows Knowledge learnt from patterns in one domain to be applied in another domain; for example to enable classification with respect to the other domain. In the case of CNNs pre-trained image recognition models, such as ResNet and Inception v3, are available which can be fine tuned using a limited amount of ECG image data [9, 16, 21, 31, 33, 40, 44, 52]. For example in [9] the authors fine tuned Inception V3 with real scanned images data to produced good results (AUC of 0.935). The work presented in this paper used the same data set as used in [9]. In this paper three pre-trained CNN model were fine tuned to extract features from 2D ECG image data, which were then used to construct SVM and KNN classification models. We compared the results with [9] model and [16].

3 Proposed Approach

This section presents the proposed approach to feature extraction from ECG raw image data. The approach comprises five stages:

1. ECG Image Pre-processing.
2. Features Extraction.
3. Dimension Reduction
4. Data Augmentation
5. Feature Vector Generation

Detail concerning each of these stages is presented in the following five sub-sections, Sub-sections 3.1, 3.2, 3.3, 3.4 and 3.5.

3.1 ECG Image Pre-processing

So that the ECG image data could be used with an appropriate pre-trained CNN model the images needed to be dimensioned so as to be compatible with nature of the the adopted architecture for the CNN model in question. For example ResNet50 requires that the input images size is a multiple of 32. Therefore, with respect to the work presented here, all images were first resized; 299×299 pixels for InceptionV3, and 224×224 pixels for ResNet50 and VGG16

3.2 Feature Extraction

The idea presented in this paper is to extract features from ECG image data and then to use the extracted features to build a classification model, thus avoiding the disadvantages of 1D techniques as considered earlier in this paper. The proposed method is to extract the desired features using pre-trained CNNs. Usually, initial layers of a CNN capture basic input image features, such as boundaries and colour patterns. Then the deeper hidden layers capture the complex higher-level feature patterns [6]. The most discriminating features are thus held in the Fully-Connected Layers (FCLs) before the final output classification layer (the softmax layer). The features in these FCLs were the features used with respect to the work presented here. Three pre-trained CNN image recognition models were considered: (i) VGG16, (ii) ResNet50 and (iii) InceptionV3. Details concerning these pre-trained models are presented below.

VGG16 The VGG network architecture was introduced by Simonyan and Zisserman in [47]. The acronym VGG stands for Visual Geometry Group, a group within the Department of Science and Engineering at the University of Oxford. The group has released a series of CNN models beginning with VGG, VGG16 to VGG19, the number denotes the number of layers. VGG16 was used with respect to the work presented in this paper. The VGG16 architecture requires a $224 \times 224 \times 3$ image size as an input, and generates an output feature vector size of 4096. Details of the adopted VGG16 architecture are presented in Table 1. Features were extracted from the last two FCLs prior to the SoftMax layer.

Table 1. The network structure of VGG16 convolutional neural network used in this paper

Layer type	Kernel size	Output size
Conv 1	3x3,64 3x3,64	224x224x64
Max pool		112x112x64
Conv 2	3x3,128 3x3,128	112x112x128
Max pool		56x56x128
Conv 3	3x3,256 3x3,256 3x3,256	56x56x256
Max pool		28x28x256
Conv 4	3x3,512 3x3,512 3x3,512	28x28x512
Max pool		14x14x512
Conv 5	3x3,512 3x3,512 3x3,512	14x14x512
Max pool		7x7x512
FC-4096,FC-4096,FC-1000,softmax		

Table 2. The network structure of ResNet50 convolutional neural network used in this paper

Layer type	Kernel size	Output size
Conv 1	7x7x3	112x112x64
Max pool	3x3	56x56
Conv 2_x	$\begin{Bmatrix} 1x1, 64 \\ 3x3, 64 \\ 1x1, 256 \end{Bmatrix}$ x 3	56x56
Conv 3_x	$\begin{Bmatrix} 1x1, 128 \\ 3x3, 128 \\ 1x1, 512 \end{Bmatrix}$ x 4	28x28
Conv 4_x	$\begin{Bmatrix} 1x1, 256 \\ 3x3, 256 \\ 1x1, 1024 \end{Bmatrix}$ x 6	14x14
Conv 5_x	$\begin{Bmatrix} 21x1, 512 \\ 3x3, 512 \\ 1x1, 2048 \end{Bmatrix}$ x 3	7x7
Average pool		1x1
1000-d fc, softmax		

ResNet50 The ResNet network architecture was introduced by He et al. [22]. It is considered by some to be the state-of-the-art for CNN-based image recognition [4]. There are multiple versions of ResNetXX where 'XX' indicates the number of layers. The most commonly used, and that used with respect to the work presented in this paper, is ResNet50. The ResNet-50 architecture, as in the case of VGG16, requires a $224 \times 224 \times 3$ pixel images as an input, and outputs a feature vector of size is 2048. The details of the adopted ResNet-50 architecture are given in Table 2, features were extracted from the last FCL prior to the SoftMax layer.

Inception-V3 The "Inception" micro-architecture was introduced by Szegedy et al. [48]. The original name of this architecture was GoogLeNet, but subsequent manifestations have simply been called Inception VN where N denotes the version number. Inception-V3 is that used in the context of this paper. The Inception-V3 architecture require $224 \times 224 \times 3$ pixel images. The output feature vector is of size 2048. The details of the adopted Inception-V3 architecture are given in Table 3. Features were extracted from the last FCL prior to the SoftMax layer.

Table 3. The network structure of the Inception-v3 CNN used in this paper

Layer name	Patch size	Output size
Conv	3×3/2	149×149×32
Conv	3×3/1	147×147×32
Conv padded	3×3/1	147×147×64
Max Pool	3×3/2	73×73×64
Conv	3×3/1	71×71×80
Conv	3×3/2	35×35×192
Conv	3×3/1	35×35×288
3 × Inception	Module 1	17×17×768
5 × Inception	Module 2	8×8×1280
2 × Inception	Module 3	8 × 8 × 2048
Max Pool	8 × 8	1 × 1 × 2048
Linear, Logits, Softmax		

3.3 Dimensionality Reduction

Dimensionality Reduction (DR) is pre-processing step aimed at either reducing the number of features, thus reducing the resources required for classification model generation, or to aid in visualising the data before any analysis is performed. In the case of classification model generation DR is applied before model training is commenced [51]. From the previous section, 4096 features were extracted using VGG16, and 2048 using ResNet50 and Inception-V3. Thus DR was adopted in order to reduce the number of features, while attempting to keep as much of the variation in the original features set as possible. There are many algorithms that can be used for DR that can be categorised into two groups: linear algebra and manifold learning. In linear algebra the methods used examine the linear relationship between the variables, while in manifold learning non-linear approaches are used to capture more complex relationships between variables. For the work presented in this paper three methods were considered: (i) Principal Components Analysis (PCA) and (ii) Singular Value Decomposition (SVD) and T-distributed Stochastic Neighbor Embedding (T-SNE). The first two are linear algebra approaches and the third a manifold approach. In more detail:

- PCA conducts a linear combination of an existing large set of features so as to create a new set of features. These new features are referred to as “Principal Components”. The aim is to capture as much information as possible in the smallest number of principal components.
- SVD decomposes the original features into three constituent matrices to remove redundant features. The twin concepts of Eigenvalues and Eigenvectors are used to calculate these matrices.
- T-SNE reduces the number of features by creating two or three new features. It calculates the probability similarity of points in a high dimensional space and uses this define a low dimensional space. Nearby points in the high dimensional space are then mapped to the nearest points in the low dimensional space.

3.4 Data Augmentation

The ECG image data set used for evaluation purposes was a binary-labelled, imbalanced, data set (see Sub-section 4.1 for more detail). To address this issue the minority class was augmented through the application of an oversampling technique. In “classic” oversampling, the minority data is simply duplicated. However, this will not add any new information. In [12] the Synthetic Minority Oversampling Technique (SMOTE) was presented, a technique which can be used to synthesize new examples from existing examples. For the work presented in this paper SMOTE was adopted. SMOTE operates by first selecting random records from the minority class and, for each selected record, the k-nearest neighbours. Synthetic data is then created using these "clusters". For the work presented in this paper three different SMOTE techniques were considered: (i) the original SMOTE, (ii) Support Vector Machine SMOTE (SVM-SMOTE) and (iii) Adaptive Synthetic Sampling (ADASYN):

- As noted above, the original SMOTE operates by first selecting random records from the minority class and, for each selected record the k-nearest neighbours. Synthetic data is then created using these "clusters".
- Using SVM-SMOTE, instead of using K-nearest neighbors, support vectors are used. Synthetic data is randomly created along the lines joining each minority class support vector with a number of its nearest neighbors.
- Using ADASYN data density is used to create synthetic data. Additional synthetic data is created in the "areas" where the minority class is less dense.

3.5 Feature Vector Generation

The last process in the proposed approach is the generation of a set of feature vectors $H = \{V_1, V_2, \dots\}$. Each $V_i \in H$ is of the form $\{v_1, v_2, \dots, c\}$ where v_i is a numerical features values extracted from an ECG scanned image. Interestingly v_i could also be a values obtained from some other source than the core ECG data. In the evaluation presented later in this paper the results of experiments are reported where age and gender are appended to H. For classification model training purposes a final element c, a class label taken from a set of classes C, is added to each $V_i \in H$. For the evaluation presented in Section 4, $|C| = 2$ was used. A previously unseen record will have a null value for the variable c as this is the value we wish to predict.

4 Evaluation

The evaluation of the proposed mechanism is presented in this section. For the evaluation the Guangzhou Heart Study data set was used [15]. Some detail concerning this data set is provided in Sub-section 4.1. Both SVM and KNN classification models were used for the evaluation with Grid Search to choose the best parameters: SVM (C, gamma and kernels) and KNN (neighbors, weights and p). The evaluation metrics used were accuracy, F1 and AUC; Ten-fold cross-validation was used throughout. The Friedman Test was used to determine whether or not there was a statistically significant difference between the performance of the models. Where a statistically significant difference was found, the Nemenyi post-hoc test was used to identify the distinctions between the performance of the mechanisms considered. The objectives of the evaluation were as follows:

1. To identify the most appropriate pre-trained CNN model, dimensionality reduction technique and data augmentation technique.
2. To compare the operation of the proposed approach when the feature set is appended with additional data.
3. To compare the operation of the proposed approach with other published approaches.

Each is discussed in further detail below in Sub-sections 4.2, 4.3 and 4.4.

4.1 Data Set

As we require a scanned image dataset in this paper, The Guangzhou Heart Study data set is used, comprised 1172 patients (399 males, and 773 females) with a mean age of 71.4 years; each patient was associated with a 12-leads ECG scanned image and patient attributes, including age and gender. Each patient record had been labelled according to arrhythmia type, either sinus arrhythmia (normal) or abnormal. The abnormal category included: (i) Atrial Fibrillation (AF) and Flutter (AFL), (ii) premature atrial or ventricular contraction, (iii) Atrioventricular Block (AVB), (iv) ventricular tachycardia, (v) Supraventricular Tachycardia (SVT), (vi) Wolff-Parkinson-White syndrome (WPW), (vii) pacing rhythm and (viii) junctional rhythm. From the 1172 patients, 878 (74.9%) were classified as normal, and the remaining 294 (25.1%) as abnormal. The image resolution was 300 dpi (dots per inch) and each image was stored using JPEG compression. Ten cross validation was used throughout, thus on each run the training data comprised 1055 images, and the test data 117 images.

4.2 Best Combination of techniques

From Section 3 the proposed approach incorporated three categories of technique:

- Feature extraction.
- Dimension reduction.
- Data augmentation.

Three different techniques were considered with respect to each. Experiments were conducted to identify the best technique in each case.

For the feature extraction VGG19, ResNet50 and Inception-v3 were considered (as described previously in Sub-section 3.2). Recall that for VGG the number of features was $4,096$, and for ResNet50 and Inception-v3 the number of features was $2,048$. We also considered the combination of the extracted features, because in [40] this had demonstrated good results. We ran the experiments using each feature vector set separately and in combination: (i) all models ($4,096 + 2,048 + 2,048 = 8192$ features), (ii) ResNet50 and VGG16 ($4,096 + 2,048 = 6144$ features), (iii) ResNet50 and Inception-v3 ($2,048 + 2,048 = 4096$ features) and (iv) VGG16 and Inception-v3 ($4,096 + 2,048 = 6144$ features). In all cases t-SNE was used for feature selection and SMOTE for record augmentation (to limit the number of combinations to be considered with respect to this first set of experiments). The results obtained are given in Table 4. From, the table, it can be seen that the best accuracy and AUC values were obtained using Resnet50 feature selection coupled with SVM classification (AUC of 94.21% and accuracy of 94.12). While when using kNN classification best results were obtained when all 8192 features were used in combination (AUC of 87.32% and accuracy of 87.70%). A subsequent Friedman Test indicated that there was a statistically significant difference in operation when using SVM classification, but not when using kNN classification. Figure 1 shows the result of a Nemenyi post-hoc test using SVM classification. From the figure it can be seen that the ResNet50 model has significant differences with all other models except with ResNet50+VGG16 ($p = .25$) and VGG16 +InceptionV3 ($p = 0.20$). It was thus concluded that for feature extraction ResNet50 was the most appropriate choice.

For the experiments to determine the best dimensionality reduction technique the three alternatives discussed in Sub-section 3.3 were considered: (i) T-SNE, (ii) SVD and (iii) PCA. Table 5 shows the results obtained using ResNet50 feature selection (because the experiments reported above indicated that this produced the best results) and SMOTE

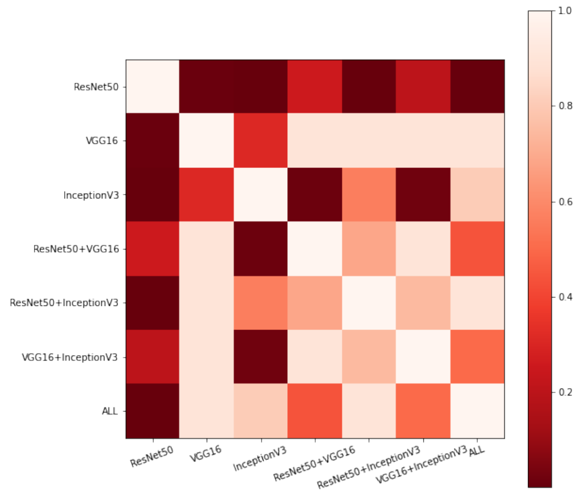

Fig. 1. Nemenyi post-hoc test for feature generation coupled with SVM classification

Table 4. SVM and KNN classification performance using a different pre-trained models features

Model/Classification	SVM			KNN		
	Accuracy %	**F1 %**	**AUC %**	**Accuracy %**	**F1 %**	**AUC %**
ResNet50	**94.21**	**93.64**	**94.12**	85.80	86.72	86.17
VGG16	89.34	89.56	89.47	85.32	86.29	84.62
InceptionV3	85.27	85.56	85.46	84.34	85.45	84.62
ResNet50+VGG16	89.97	89.89	90.06	85.38	86.37	85.75
ResNet50+InceptionV3	88.07	88.34	88.20	86.11	87.00	86.46
VGG16+InceptionV3	89.96	90.30	90.1	87.01	87.75	87.27
ResNet50+VGG16+InceptionV3	88.39	88.49	88.48	**87.32**	**87.80**	**87.70**

augmentation (to limit the number of combinations to consider). Results are given using both SVM and kNN classification. Inspection of the table indicates that best values were obtained using T-SNE, while PCA caused over-fitting regardless of whether SVM or kNN classification was used. It was thus concluded that for feature selection T-SNE was the most appropriate choice.

For the experiments to determine the most appropriate augmentation method three alternatives were considered (as discussed in Sub-section 3.4). Experiments were conducted using ResNet50 and t-SNE for SVM classification, and all features and t-SNE for KNN classification; because earlier experiments (see above) had indicated that these tended to produce a best performance. The results are presented in Table 6. From the table it can be seen that SMOTE produced the best results when using SVM classification, and ADASYN when using kNN classification. A Friedman Test indicated a statistically significant difference in the results when using SVM classification (not the case when using kNN). Figure 2, show the result from a Nemenyi post-hoc test for the results obtained using SVM classification. From the figure it can be seen that there is a statistically significant difference when using SMOTE.

Table 5. SVM and KNN classification performance using a range of features selection techniques

Technique/Classification	SVM			KNN		
	Accuracy %	F1 %	AUC %	Accuracy %	F1 %	AUC %
T-SNE	**94.21**	**94.19**	**94.12**	**87.32**	**87.80**	**87.70**
SVD	63.10	62.66	63.16	68.03	49.12	64.39
PCA	Overfitting			Overfitting		

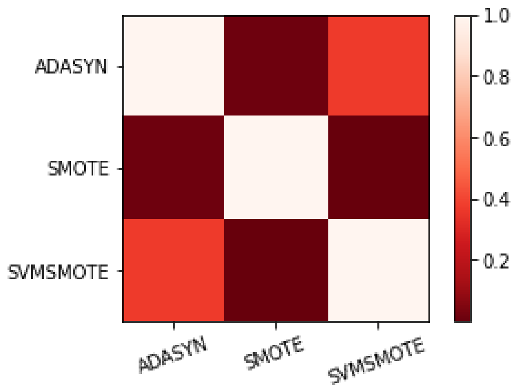

Fig. 2. Nemenyi post-hoc test for augmentation techniques and SVM classification

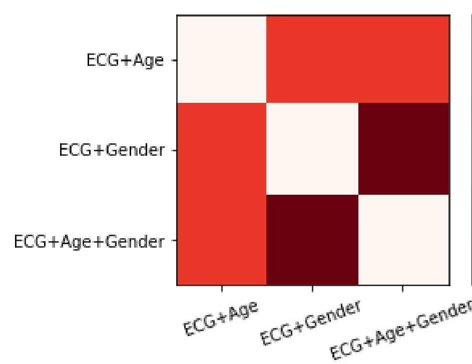

Fig. 3. Nemenyi post-hoc for all models with SVM classifier

4.3 Analysis of the Effect of Adding Additional Data

The previous sub-section described the experiments conducted to determine the best feature extraction, feature selection and augmentation techniques for processing ECG image data to support ECG classification. Best results were obtained using ResNet50 feature extraction, T-SNE feature selection and SMOTE data augmentation. Further experiments were conducted to determine whether any advantage could be gained by adding additional features from related sources. To this end age and gender were added to the generated feature vector for each ECG (note that each ECG in the data set was related to a single patient). Similar experiments were conducted in [2]; but in the 1D, time series, context.

The results are presented in table 7. From the table, and with reference to Table 6, it can be seem that adding patient age improves the effectiveness of the classification, an AUC of 95.17 using SVM, and 89.15 using KNN (compared to 94.12 and 87.70 respectively, when not adding additional information). Gender has a less pronounced effect. Figure 3 shows the result of Nemenyi post-hoc test applied to all models for SVM classification after a Friedman Test reported that there was a statistically significant difference. From the figure it can be seen that there was a statistically significant difference between using ECG + age + gender and ECG + gender; thus adding age has prominent effect on performance. In the KNN case, the Friedman Test indicated that there was no statistically significant difference between the models ($p = 1.0$).

4.4 Comparison of Approaches

The experimental results obtained, using the proposed approach, were also compared those obtained in recently published work directed at the same data set, namely the work presented

Table 6. SVM and KNN classification performance using a range of data augmentation techniques

Technique/Classification	SVM			KNN		
	Accuracy %	F1 %	AUC %	Accuracy %	F1 %	AUC %
SMOTE	**94.21**	**94.19**	**94.12**	87.32	87.80	87.70
ADASYN	91.69	91.19	91.69	**88.46**	**81.11**	**88.82**
SVMSMOTE	88.99	85.73	88.69	87.58	80.82	88.23

Table 7. SVM and KNN classification performance when adding additional data

Technique/Classification	SVM			KNN		
	Accuracy %	F1 %	AUC %	Accuracy %	F1 %	AUC %
ECG + Age	95.19	94.98	95.17	88.97	89.13	89.15
ECG + Gender	94.45	94.06	94.38	88.49	89.13	88.86
ECG + Age + Gender	**96.06**	**95.92**	**96.07**	**89.01**	**89.46**	**89.37**

in [9] and [16]. In [9], they fine-tuned an Inception-v3 pre-trained model to extract features which were than classified using a dense, fully connected layer. To address the imbalanced nature of the data set, they used the Generalized Extreme Value (GEV) activation function as an alternative to Sigmoid activation. In [16] they converted 1D time-series ECG data into 2D colour spectrogram images and used this as the input to a pre-trained CNN model. Three pre-trained CNN models were considered: AlexNet, VGG-16, and ResNet-18. In [9] a best AUC of 93.20% was reported. Whilst the proposed approach presented in this paper produced a best AUC of 94.12% with out the inclusion pf additional features, and a best AUC of 96.07% when including age and gender. In [16] best accuracy of 83.82% was reported, obtained using AlexNet. Best accuracies of 94.21% and 96.06% were recorded with respect to the proposed approach without and with the inclusion of additional data (age and gender) respectively.

Finally, in the introduction to this paper, the disadvantages of using 1D waveform representations of ECG data was noted. It was hypothesised that using 2D image directly to extract the features would produce a better classification than that obtained using features selected from 1D transformed waveform representations of ECG data. Accordingly, the scanned images were transformed into a time series format using a recent algorithm for achieving this [18]. Once the image set had been transformed the 1D motif approach proposed in [3] was used to extract features. An accuracy of 72.35% was obtained using the 1D approach, compared that with 94.21% and 96.06% accuracies obtained using the proposed approach (without and with the inclusion of age and gender).

5 Conclusion

An approach to classifying ECG image data using a pre-trained CNN model to extract features has been presented. The motivation for the work was the observation that most ECG data is still in paper format. For classification purposes this paper ECG data is typically processed by first scanning into a digital format, from which time series are generated, from which a small number of features (P-QRS-T wave features) are extracted, which are then used to build a classification model. This chain of processes introduces a range of irregularities and noise. To reduce this chain it was suggested that 2D features, extracted directly from the scanned images, could be used; and that features be extracted using some form of CNN. It was also noted that the amount of data available is frequently insufficient to build a CNN. Hence it was proposed to use a pre-trained CNN model. Three CNN models were considered for the feature extraction, together with three feature selection mechanisms. It was also noted that the data available tends to be imbalanced; typically we have more examples of ECGs for patients with a CVD than without (individuals tend to only have ECGs taken when a CVD is suspected). A range of augmentation techniques were considered. The presented evaluation demonstrated that best results were obtained using ResNet50 feature extraction, T-SNE feature selection and SMOTE data augmentation. Experiments were also conducted to investigate the effect of adding additional features obtained from elsewhere (age and gender) and this was found to provide an improved result. A best AUC of 96.07 was obtained. Comparison was made with existing work, that presented in [9] and [16]. The results indicated that the proposed approach outperformed these existing approaches. Finally comparison was undertaken with a 1D technique indicating that the motivating hypothesis was correct.

References

1. Acharya, U.R., Oh, S.L., Hagiwara, Y., Tan, J.H., Adam, M., Gertych, A., San Tan, R.: A deep convolutional neural network model to classify heartbeats. Computers in Biology and Medicine **89**, 389–396 (2017)
2. Aldosari, H., Coenen, F., Lip, G.Y., Zheng, Y.: Addressing the challenge of data heterogeneity using a homogeneous feature vector representation: A study using time series and cardiovascular disease classification. In: International Conference on Innovative Techniques and Applications of Artificial Intelligence. pp. 254–266. Springer (2021)
3. Aldosari, H., Coenen, F., Lip, G.Y., Zheng, Y.: Motif based feature vectors: towards a homogeneous data representation for cardiovascular diseases classification. In: International Conference on Big Data Analytics and Knowledge Discovery. pp. 235–241. Springer (2021)
4. Almabdy, S., Elrefaei, L.: Deep convolutional neural network-based approaches for face recognition. Applied Sciences **9**(20), 4397 (2019)
5. Annam, J.R., Kalyanapu, S., Ch, S., Somala, J., Raju, S.B.: Classification of ECG heartbeat arrhythmia: A review. Procedia Computer Science **171**, 679–688 (2020)
6. Awais, M., Long, X., Yin, B., Chen, C., Akbarzadeh, S., Abbasi, S.F., Irfan, M., Lu, C., Wang, X., Wang, L., et al.: Can pre-trained convolutional neural networks be directly used as a feature extractor for video-based neonatal sleep and wake classification? BMC Research Notes **13**(1), 1–6 (2020)
7. Badilini, F., Erdem, T., Zareba, W., Moss, A.J.: Ecgscan: a method for conversion of paper electrocardiographic printouts to digital electrocardiographic files. Journal of Electrocardiology **38**(4), 310–318 (2005)
8. Baydoun, M., Safatly, L., Abou Hassan, O.K., Ghaziri, H., El Hajj, A., Isma'eel, H.: High precision digitization of paper-based ECG records: a step toward machine learning. IEEE Journal of Translational Engineering in Health and Medicine **7**, 1–8 (2019)
9. Bridge, J., Fu, L., Lin, W., Xue, Y., Lip, G.Y., Zheng, Y.: Artificial intelligence to detect abnormal heart rhythm from scanned electrocardiogram tracings. Journal of Arrhythmia (2022)
10. Brisk, R., Bond, R., Banks, E., Piadlo, A., Finlay, D., McLaughlin, J., David, M.: Deep learning to automatically interpret images of the electrocardiogram: Do we need the raw samples? Journal of electrocardiology **57**, S65–S69 (2019)
11. Cao, Q., Du, N., Yu, L., Zuo, M., Lin, J., Liu, N., Zhong, E., Liu, Z., Chen, Q., Shen, Y., et al.: Practical fine-grained learning based anomaly classification for ECG image. Artificial Intelligence in Medicine **119**, 102130 (2021)
12. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: Smote: synthetic minority oversampling technique. Journal of Artificial Intelligence Research **16**, 321–357 (2002)

13. Chung, D., Choi, J., Jang, J.H., Kim, T.Y., Byun, J., Park, H., Lim, H.S., Park, R.W., Yoon, D.: Construction of an electrocardiogram database including 12 lead waveforms. Healthcare Informatics Research **24**(3), 242–246 (2018)
14. Cuocolo, R., Perillo, T., De Rosa, E., Ugga, L., Petretta, M.: Current applications of big data and machine learning in cardiology. Journal of Geriatric Cardiology: JGC **16**(8), 601 (2019)
15. Deng, H., Guo, P., Zheng, M., Huang, J., Xue, Y., Zhan, X., Wang, F., Liu, Y., Fang, X., Liao, H., et al.: Epidemiological characteristics of atrial fibrillation in southern china: results from the guangzhou heart study. Scientific Reports **8**(1), 1–10 (2018)
16. Diker, A., Cömert, Z., Avcı, E., Toğaçar, M., Ergen, B.: A novel application based on spectrogram and convolutional neural network for ECG classification. In: 2019 1st International Informatics and Software Engineering Conference (UBMYK). pp. 1–6. IEEE (2019)
17. Ebrahimi, Z., Loni, M., Daneshtalab, M., Gharehbaghi, A.: A review on deep learning methods for ECG arrhythmia classification. Expert Systems with Applications: X **7**, 100033 (2020)
18. Fortune, J., Coppa, N., Haq, K.T., Patel, H., Tereshchenko, L.G.: Digitizing ECG image: new fully automated method and open-source software code. medRxiv (2021)
19. Gupta, D., Bajpai, B., Dhiman, G., Soni, M., Gomathi, S., Mane, D.: Review of ECG arrhythmia classification using deep neural network. Materials Today: Proceedings (2021)
20. Gupta, V., Mittal, M., Mittal, V., Saxena, N.K.: A critical review of feature extraction techniques for ECG signal analysis. Journal of The Institution of Engineers (India): Series B pp. 1–12 (2021)
21. Hadiyoso, S., Fahrozi, F., Hariyani, Y.S., Sulistiyo, M.D.: Image based ECG signal classification using convolutional neural network. International Journal of Online & Biomedical Engineering **16**(4) (2022)
22. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778 (2016)
23. Holkeri, A., Eranti, A., Kenttä, T.V., Noponen, K., Haukilahti, M.A.E., Seppänen, T., Junttila, M.J., Kerola, T., Rissanen, H., Heliövaara, M., et al.: Experiences in digitizing and digitally measuring a paper-based ecg archive. Journal of electrocardiology **51**(1), 74–81 (2018)
24. Houssein, E.H., Kilany, M., Hassanien, A.E.: ECG signals classification: a review. International Journal of Intelligent Engineering Informatics **5**(4), 376–396 (2017)
25. Izci, E., Ozdemir, M.A., Degirmenci, M., Akan, A.: Cardiac arrhythmia detection from 2d ECG images by using deep learning technique. In: 2019 Medical Technologies Congress (TIPTEKNO). pp. 1–4. IEEE (2019)
26. Jayaraman, S., Swamy, P., Damodaran, V., Venkatesh, N.: A novel technique for ECG morphology interpretation and arrhythmia detection based on time series signal extracted from scanned ECG record. Advances in Electrocardiograms-Methods and Analysis pp. 127–140 (2012)
27. Ji, Y., Zhang, S., Xiao, W.: Electrocardiogram classification based on faster regions with convolutional neural network. Sensors **19**(11), 2558 (2019)
28. Jun, T.J., Nguyen, H.M., Kang, D., Kim, D., Kim, D., Kim, Y.H.: ECG arrhythmia classification using a 2-D convolutional neural network. arXiv preprint arXiv:1804.06812 (2018)
29. Kar, A., Das, L.: A technical review on statistical feature extraction of ECG signal. In: IJCA Special Issue on 2nd National Conference-Computing, Communication and Sensor Network, CCSN. pp. 35–40 (2011)
30. Khleaf, H.K., Ghazali, K.H., Abdalla, A.N.: Features extraction technique for ECG recording paper. In: Proceeding of the International Conference on Artificial Intelligence in Computer Science and ICT (2013)
31. Kim, J.S., Kim, S.H., Pan, S.B.: Personal recognition using convolutional neural network with ECG coupling image. Journal of Ambient Intelligence and Humanized Computing **11**(5), 1923–1932 (2020)
32. Kiranyaz, S., Ince, T., Gabbouj, M.: Real-time patient-specific ECG classification by 1-D convolutional neural networks. IEEE Transactions on Biomedical Engineering **63**(3), 664–675 (2015)
33. Li, C., Zhao, H., Lu, W., Leng, X., Wang, L., Lin, X., Pan, Y., Jiang, W., Jiang, J., Sun, Y., et al.: DeepECG: Image-based electrocardiogram interpretation with deep convolutional neural networks. Biomedical Signal Processing and Control **69**, 102824 (2021)

34. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van Der Laak, J.A., Van Ginneken, B., Sánchez, C.I.: A survey on deep learning in medical image analysis. Medical Image Analysis **42**, 60–88 (2017)
35. Liu, X., Wang, H., Li, Z., Qin, L.: Deep learning in ECG diagnosis: A review. Knowledge-Based Systems **227**, 107187 (2021)
36. Loresco, P.J.M., Africa, A.D.: ECG print-out features extraction using spatial-oriented image processing techniques. Journal of Telecommunication, Electronic and Computer Engineering (JTEC) **10**(1-5), 15–20 (2018)
37. Mathunjwa, B.M., Lin, Y.T., Lin, C.H., Abbod, M.F., Shieh, J.S.: ECG arrhythmia classification by using a recurrence plot and convolutional neural network. Biomedical Signal Processing and Control **64**, 102262 (2021)
38. Mir, H.Y., Singh, O.: Ecg denoising and feature extraction techniques–a review. Journal of Medical Engineering & Technology **45**(8), 672–684 (2021)
39. Mishra, S., Khatwani, G., Patil, R., Sapariya, D., Shah, V., Parmar, D., Dinesh, S., Daphal, P., Mehendale, N.: ECG paper record digitization and diagnosis using deep learning. Journal of Medical and Biological Engineering pp. 1–11 (2021)
40. Naz, M., Shah, J.H., Khan, M.A., Sharif, M., Raza, M., Damaševičius, R.: From ECG signals to images: a transformation based approach for deep learning. PeerJ Computer Science **7**, e386 (2021)
41. O'Keefe, J.H., Hammill, S.C., Freed, M.S., Pogwizd, S.: The Complete guide to ECGs: A comprehensive study guide to improve ECG interpretation skills. Jones & Bartlett Learning (2017)
42. Rajpurkar, P., Hannun, A.Y., Haghpanahi, M., Bourn, C., Ng, A.Y.: Cardiologist-level arrhythmia detection with convolutional neural networks. arXiv preprint arXiv:1707.01836 (2017)
43. Ravichandran, L., Harless, C., Shah, A.J., Wick, C.A., Mcclellan, J.H., Tridandapani, S.: Novel tool for complete digitization of paper electrocardiography data. IEEE Journal of Translational Engineering in Health and Medicine **1**, 1800107–1800107 (2013)
44. Salem, M., Taheri, S., Yuan, J.S.: ECG arrhythmia classification using transfer learning from 2-dimensional deep CNN features. In: 2018 IEEE Biomedical Circuits and Systems Conference (BioCAS). pp. 1–4. IEEE (2018)
45. Sarvamangala, D., Kulkarni, R.V.: Convolutional neural networks in medical image understanding: a survey. Evolutionary intelligence pp. 1–22 (2021)
46. Seena, V., Yomas, J.: A review on feature extraction and denoising of ECG signal using wavelet transform. In: 2014 2nd International Conference on Devices, Circuits and Systems (ICDCS). pp. 1–6. IEEE (2014)
47. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
48. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–9 (2015)
49. Thanapatay, D., Suwansaroj, C., Thanawattano, C.: ECG beat classification method for ECG printout with principle components analysis and support vector machines. In: 2010 International Conference on Electronics and Information Engineering. vol. 1, pp. V1–72. IEEE (2010)
50. Ullah, A., Anwar, S.M., Bilal, M., Mehmood, R.M.: Classification of arrhythmia by using deep learning with 2-D ECG spectral image representation. Remote Sensing **12**(10), 1685 (2020)
51. Velliangiri, S., Alagumuthukrishnan, S., et al.: A review of dimensionality reduction techniques for efficient computation. Procedia Computer Science **165**, 104–111 (2019)
52. Weimann, K., Conrad, T.O.: Transfer learning for ECG classification. Scientific Reports **11**(1), 1–12 (2021)
53. Yıldırım, Ö., Pławiak, P., Tan, R.S., Acharya, U.R.: Arrhythmia detection using deep convolutional neural network with long duration ECG signals. Computers in Biology and Medicine **102**, 411–420 (2018)