Multi-Class ECG Classification using Continuous Wavelet Transforms

Reuben Kouidri

Under the Supervision of *Henggui Zhang*

School of Physics and Astronomy

University of Manchester

January 2022

Abstract

A 42-layer deep 2D-CNN with 2 channel attention modules was designed for a 9-class ECG classification task. A separate model was trained on leads 4 and 7 using a Mexican hat wavelet to transform the raw data into a 2D scalogram, and on lead-4 using a complex-Morlet wavelet. The models were trained for 50 epochs on a single NVIDIA Tesla P100 GPU using mini-batch SGD to update the weights, and cross-entropy was used as the loss criterion. The dataset used was from the 2018 China Physiological Signal Challenge. The model trained on lead-7 outperformed lead-4 in all aspects, and the Mexican hat wavelet outperformed the complex-Morlet wavelet on lead-4, again, in all aspects. Lead-7 produced F1 scores of 0.945 and 0.851 for the classifications of atrial fibrillation and premature contraction respectively, beating all other competitors in the Challenge. Different results on leads 4 and 7 suggest that more testing needs to be carried out on all leads to determine whether or not an ensemble model could be deployed to yield higher overall performance.

1 Introduction

Cardiovascular diseases (CVDs) are the leading cause of mortality worldwide totalling 17.9m, or an estimated 32% in 2019 [55]. Cardiac Arrhythmia (CA) refers the abnormal rhythm or beating of the heart, and is present in 12.6% of the US population over 65 [45] [30]. The most common CA is atrial fibrillation (AF) with an estimated world prevalence of 0.51% - an increase of 33% over the last 20 years [43]. Even though AF is not always malignant, it is a serious heart disease and is one of the main precursors of cardioembolic stroke [30]. Cardioembolic stroke accounts for $14-30\%$ of all ischemic strokes, and results in high morbidity and mortality [9], as well as being a significant burden on health systems worldwide [43]. If identified early, proceeding diseases can be treated by medicines such as beta blockers, surgery such as ablation or having a pacemaker fitted if required [29].

ECG signals provide invaluable voltage-time information about the heart that, if interpreted correctly, can be used to diagnose CAs. However, false-negatives and misdiagnoses are very common. A meta analysis [18] found that before specific training, physicians correctly interpreted only 54% of ECG signals (median score), increasing to 67% post-training, and cardiologists correctly interpreted 74.9% of the ECG signals and with large variation across individual studies highlighting a wide-spread deficiency in interpretation ability that needs to be addressed. A typical ECG diagnosis is also very time consuming. Combined with an increase in demand for healthcare and a shortage of healthcare professionals - especially physicians [12] - there is a need for some level of automation that can aid physicians and cardiologists carry out more accurate diagnostics, reduce waiting times, reduce avoidable deaths and reduce the economic burden on healthcare systems around the world [60] [56].

Machine learning (ML) is a type of artificial intelligence (AI) that derives knowledge from historical data in order to make predictions, without being explicitly programmed to do so [48] [36]. Deep learning (DL) is a subfield of ML that uses self-learning algorithms as artificial neural networks [36] [49]. ML has seen a boom in recent years thanks to the availability of big data and increased computing power [61]. ML algorithms have already been successfully applied in a large breadth of fields, from voice and text translators, automatic facial recognition (e.g. airport gates), self-driving cars, trading bots and market analysis. More recently, ML has been used to make groundbreaking achievements in science such as solving the 50-year-old 'Protein Folding' problem [37], as well as equalling or surpassing human ability in a variety of tasks such as image classification [31], gaming [68] and skin cancer detection [24] [47]. In 2017, Stanford ML Group training a 34-layer deep convolutional neural network (CNN) on a single-lead ECG that surpassed the accuracy of board-certified cardiologists [58].

Applying AI to healthcare will allow earlier, more accurate detection and diagnostics, reduce waiting times, hospital visits, and take a lot of pressure off already struggling healthcare systems [12].

2 Background

2.1 Electrical Decomposition of the Heartbeat

An electrical signal is first initiated by a cluster of pacemaker cells which form the sinoatrial node (SAN) above the right atrium. The signal spreads out over the left and right atria causing them to contract, moving blood from the atria into the ventricles. This is seen as the P-wave of an ECG [3].

After a short delay of approximately $80ms$ allowing for optimal ventricular preload [33], the signal is regenerated by a second group of pacemaker cells in the atrioventricular node (AVN) and relayed through the Bundle of His. The Bundle of His bifurcates into right and left bundle branches that take the signal through the septum to the apex of the ventricles and their corresponding ventricle walls [64]. The bundle branches are covered in an insulating sheath that terminates only at specific locations that connect the Purkinje fibres, with the goal of creating coordinated ventricular contraction [10]. Depolarisation of the ventricles is observed as the QRS complex of an ECG signal, and repolarisation of the ventricles as the T-wave. There is another short pause before the process repeats [25] [4]. Fig.1 and fig.**??** show how the different electrical components sum together to produce an ECG signal at lead-I, and how the the ECG

morphology differs between leads respectively.

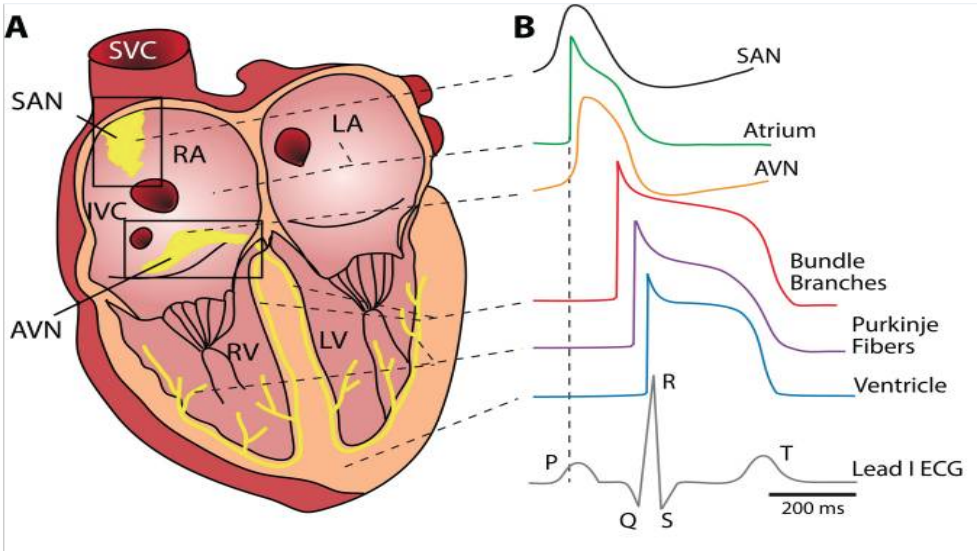

Figure 1: Lead-I ECG decomposition. ECG signals are made up of different components with distinct origins. Image taken from [10]

2.2 Types of CA

Cardiac arrhythmia (CA) is defined as the abnormal beating of the heart due to the disrupted propagation of the action potential [16] [8] and falls into 2 categories: those consisting of a single irregular heartbeat, defined as 'morphological arrhythmia', and those consisting of a set of irregular heartbeats, defined as 'rhythmic arrhythmia' [46] [10]. There are many different types of CA each with their different morphological appearances. The types of CA relevant to this study are: AF - this is where the atria contract rapidly and out of sync with the rest of the heart, and can be seen in abnormal QRS complexes as well as the absence of P waves [10] [29];
first-degree AV block (I-AVB) is defined as an elongated PR interval of more than 0.2s, usually asymptomatic [54]; premature ventricular contractions (PVC) occur when the heartbeat is initiated by the purkinje fibres instead of the SA node [27] and is seen as an enlarged QRS complex; premature atrial contraction (PAC) is defined as early depolarisation of the atrial myocardium [21] and presents itself as a non-sinus p-wave; left- and right-bundle branch blocks (LBBB and RBBB) are both associated with widened QRS complexes [6] [65], with RBBB also diagnosed by the 'rsR' pattern at leads V1-3 [15]; ST-elevation and ST-depression (STE and STD) refer to elevation or reduction from a baseline value by at least $0.1mV$ [38]. The absence of CA is defined as a sinus (or normal) rhythm (SR).

CAs are mainly associated with genetic mutations effecting ion channels, either inherited or due to other causes such as ageing [10]. For example, SAN dysfunction is a pathology of the SAN due to ageing or inherited factors and manifests itself as sinus bradycardia (slow beating of the heart), sinus pause/arrest (defined as atrial inactivity lasting longer than 3s), and tachy-brady syndrome (most frequently seen as AF) [19]. Heart blocks are caused by blockages in the conduction system between the atria and ventricles, caused by diseases such as AVN dysfunction which also effects ion channels [10] [28].
In addition to genetic and disease factors, CAs can also be caused by cardiotoxicity as a result of chemotherapeutic drugs [72], radiotherapy [11] [26], and have been observed in long-term breast-cancer and Hodgkin Lymphoma patients [57].

While some CAs are benign, others, such as those caused by dysfunction of ion channels, may be more serious. Both elongation or shortening of the QT interval caused by ion-channel dysfunctions can lead to ventricular fibrillation and sudden death if left untreated [10] [8].

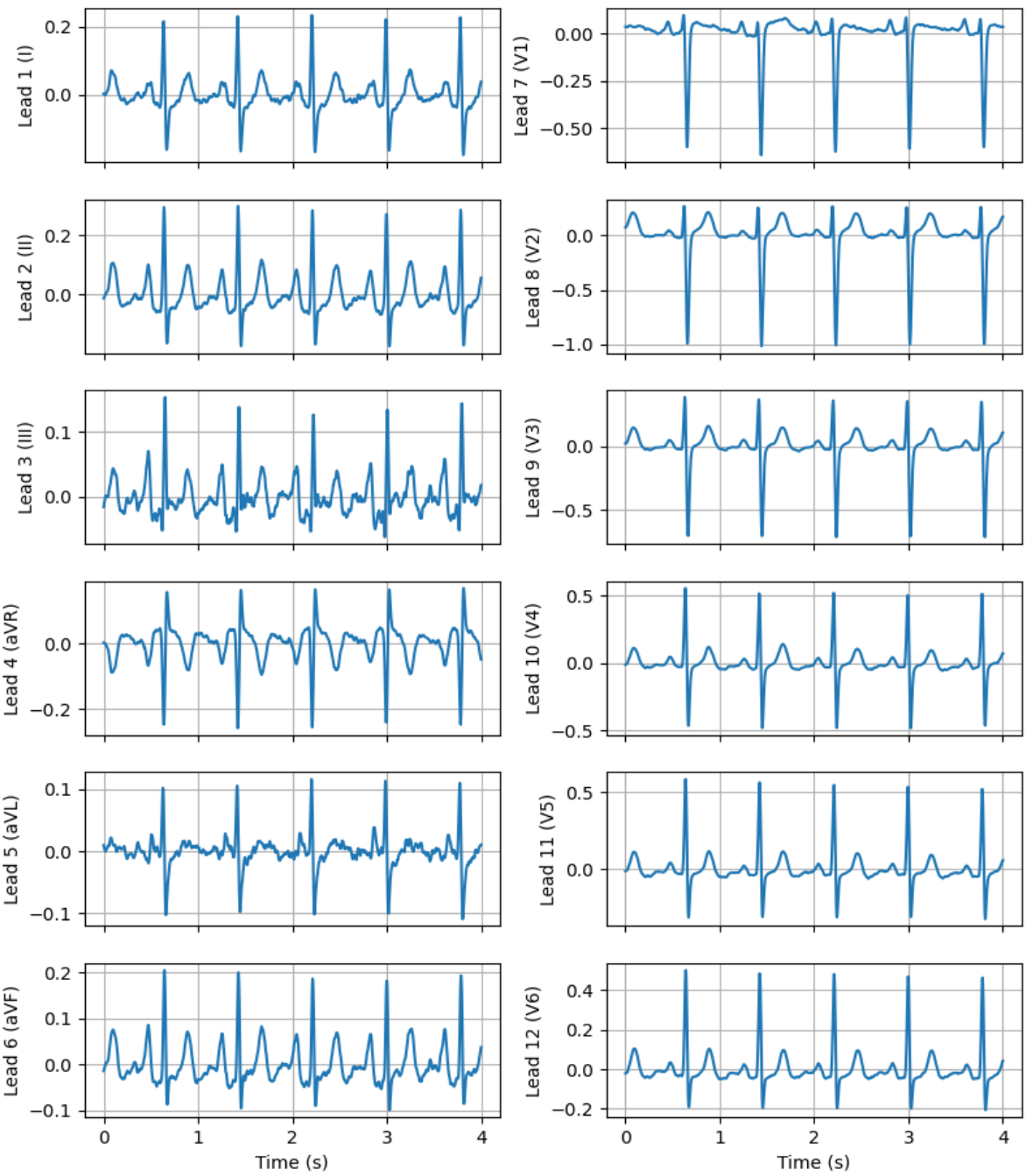

Figure 2: ECG signals obtained from each of the 12 leads for a patient with sinus rhythm.

2.3 Machine Learning

Machine learning (ML) is a type of artificial intelligence (AI) that derives knowledge from historical data in order to make predictions, without being explicitly programmed to do so [48] [36]. There are 3 types of machine learning: supervised-, unsupervised-, and reinforcement-learning. With supervised learning, the training data is labelled, and the objective of the model is to learn the features and predict on unseen data [48]. A 'classification problem' is a type of supervised learning problem where the class labels are discrete and finite, such as the classification of ECG signals considered in this paper. With unsupervised learning, labels and data structure are not known. The objective is, therefore, to extract and learn unique features from the data, and cluster items with similar features [23]. With reinforcement learning, the objective is to continuously improve upon a system, often called an 'agent', in response to interactions with the environment. Feedback is provided to the agent as a 'reward signal' - a measure of how well the agent performed with respect to a reward function, and actions are taken to maximise the long-term reward [13] [52].

2.4 Convolutional Neural Networks

CNNs refer to a subclass of NN algorithms that use a feed-forward method, first introduced by Yann LeCun et al. in 1989 [41] and are now ubiquitous in computer vision tasks [70]. They take direct inspiration from the visual cortex of mammalian brains [34], using multiple layers to extract progressively finer details of an image. CNNs utilise multiple types of layer such as convolutional layers, pooling layers, and dense layers, along with various non-linear activation functions and are trained by the back-propagation algorithm [70] [42]. CNNs perform exceptionally well on computer vision tasks due to sparse connectivity and parameter sharing [22]. Sparse connectivity means that each element of a given feature map is connected to only a small subset of pixels, or elements of the previous feature map, or in other terms, each neuron is connected to a small number of other neurons [66]. Parameter sharing means that the same weights are used to connect all patches of neurons in the input image (or feature map) to the output [59].

2.4.1 Convolutional Layers

In the language of ML, images are tensors and, henceforth, I shall use this terminology. Convolutional layers use a set of kernels (filters) to perform discrete convolution on input tensors, and each kernel will produce its own unique output tensor (c.f. Fig.3). Convolutional layers are the components of the CNN that perform the feature extraction, and each feature map extracts its own unique features.

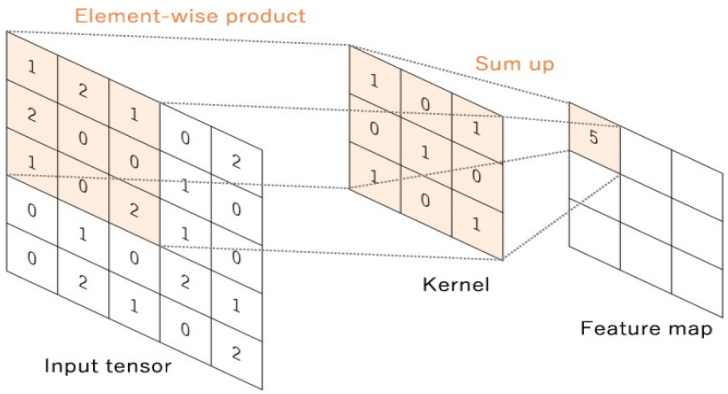

Figure 3: 2D convolution performed between a kernel and an input tensor to produce the first element of the output tensor. Since there is no activation function, this will also be the 'feature map'. No padding has been used resulting in dimensionality reduction. Different kernels have different weights and will therefore produce a different output tensor. Image taken from [70].

For a square input tensor $\mathbf{T} \in Mat(n \times n)$ and square kernel $\mathbf{K} \in Mat(m \times m)$, the mathematical operation of discrete 2D convolution $\tilde{\mathbf{T}} = \mathbf{T} * \mathbf{K}$ is given by

$$\tilde{T}_{i,j} = \sum_{p,q=1}^{m} k_{pq} T_{p+i-1,q+j-1}. \tag{1}$$

A feature map is generated by acting on the output of this convolution (with optional addition of a bias, b) with an activation function, ϕ, such as ReLU, defined as $\phi(x) = max(0, x)$, giving $\mathbf{A} = \phi(\mathbf{Z})$ [51].

Each convolutional layer has a number of hyperparameters. These are the kernel size, number of kernels, stride, padding, activation function, and need to be tuned during training [70].

2.4.2 Subsampling

Subsampling is used to decrease the size of feature maps, reducing computational cost and help reduce overfitting [62] while maintaining the most relevant features. Two common types of subsampling are are max-, and average-pooling [70]. These layers have no learnable parameters or biases. Common applications of max-pooling use a kernel of size $m \times m$ with stride-length $s = m$ as in fig.4 [50] [70]. Another form of subsampling used to reduce overtraining is dropout. This is where neurons are dropped at random during the training process. The effect of this is to force the network to learn from a reduced representation of the data, hence preventing it from relying on specific activations alone. However, during testing the full model is used with all neurons active [7].

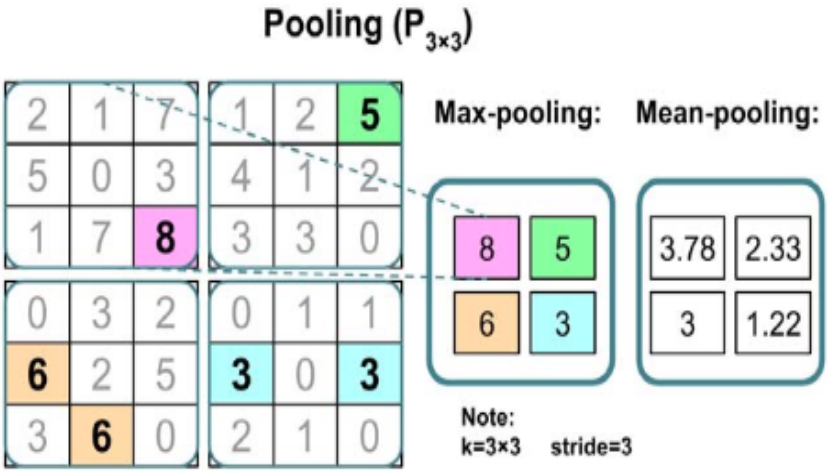

Figure 4: The max pooling operation with kernel size 3x3, and stride length 3. Image taken from [50]

2.5 Training a CNN

In the context of a CNN, training refers to the process of finding kernels and weights that minimise the difference between the model's predicted value and the ground truth value [70]. To generate a prediction, an input tensor is propagated forwards through the network. Using this prediction, we calculate the error using a designated loss function. Based on this loss function, model parameters are updated using a selected criterion and the back-propagation algorithm. The loss function used in this paper is cross-entropy loss, as is frequently used when training a multi-class classification problem [2] [70] [7]. Entropy in ML is a measure of disorder of information. The lower the entropy, the more accurate the prediction. Cross-entropy refers to the entropy difference between two probability distributions, and is defined as the number of bits required to represent of transmit an average event from one distribution to another [40]. The cross-entropy loss function is the composition of the Softmax activation function (eqn. 2)

which converts a vector of numbers into a vector of probabilities —where the components sum to 1— and the cross-entropy function (eqn. 4) which calculates the cross-entropy between the target distribution (ground-truth) and this vector of probabilities.

$$S : (v_1, ..., v_C) \mapsto (p_1, ..., p_C), \qquad p_j \equiv (S(\boldsymbol{v}))_j = \frac{e^{v_j}}{\sum_c e^{v_c}}, \tag{2}$$

For a single training image, the loss is calculated as

$$\mathcal{L} = -\sum_{j=1}^{C} w_j t_j ln(p_j) \tag{3}$$

where w_j and t_j are the j^{th} components of the **class-weight** and ground-truth (target) vectors respectively, and C is the number of classes. Class weights are used to account for class imbalance in a dataset [14]. For a batch of N images, the loss is calculated as the mean of all the individual losses

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N} \mathcal{L}_i \tag{4}$$

Batch normalisation is used to normalised the batch of output tensors before they are passed to the activation function. Batch normalisation reduces 'internal covariate shift', speeding-up the training of the network, as well as increasing accuracy and time to converge [35].

Mini-batch stochastic gradient decent (SGD) is a criterion that is frequently used to update a network's parameters, and was chosen for this model. After each pass, t, of a mini-batch, SGD tries to minimise the loss with respect to a subset of randomly-sampled weights and parameters [7] by updating them as

$$W_t = W_{t-1} - \Delta W_{t-1} \tag{5}$$

with

$$\Delta W_t = \eta \frac{\partial \mathcal{L}}{\partial W_t} + \mu \Delta W_{t-1} \tag{6}$$

where η is the step-size/learning rate, μ is the momentum (when using the Nesterov algorithm), and W_t is the selected weight at pass t. The derivative of the loss function $\mathcal{L}$ with respect to the weight is calculated using the back-propagation algorithm which uses the chain-rule to connect derivatives in earlier parts of the network to those in later parts of the network [20].

2.5.1 Channel Attention Module

Like CNNs, channel attention is inspired by the human ability to selectively focus on important information [69]. In humans and other animals, this is achieved using a retina and fixating on relevant locations to the task at hand [1]. Image resolution decreases rapidly with eccentricity, and performance in a human visual task is higher when the target objects are located around the fovea - a small pit located in the retina that provides us with the sharpest vision [53] [1] [63].

In computer vision, attention mechanisms are lightweight add-ons that boost training speeds and accuracy by 'noise reduction of irrelevant clusters' [69], i.e. by focusing on the most 'meaningful' parts of a given image and ignoring redundant information. In channel attention, input feature maps undergo both global-max and global-average pooling to generate two new feature maps. They are then passed through a shared multi-layer perceptron (MLP) with a single hidden layer, producing channel attention maps, C_m and C_a respectively. The hidden layer's size is reduced by a reduction factor, r, in comparison with the input and output layer of this MLP to increase speed whilst maintaining functionality. Element-wise summation is then applied between C_m and C_a before a sigmoid activation function gives the final output. During the training process, the weights of the MLP are updated to learn which channels/feature maps contain the most important information, and therefore which should be ignored and which should be amplified [71] [69].

I use a 'soft attention' [71] method which is differentiable, and has learnable parameters that are updated by the back propagation algorithm, as opposed the 'hard attention' as sometimes used in reinforcement learning that is not differentiable.

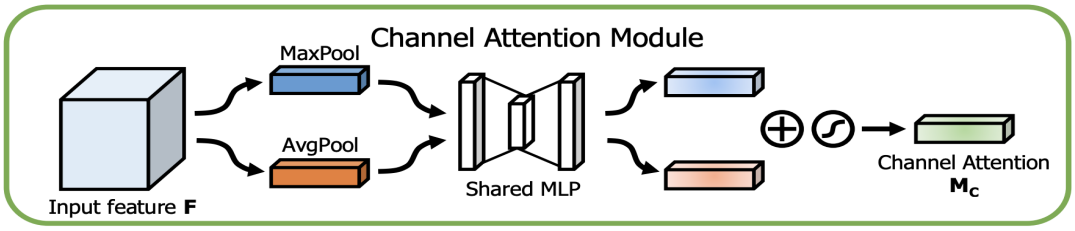

Figure 5: Channel attention module - a single component of the Convulutional Block Attention Module (CBAM). Image taken from [69].

2.5.2 Continuous Wavelet Transforms - CWT

In 1984, Jean Morlet modified Dennis Gabor's (1946) work on Gaussian-enveloped sine waves and applied it to the field of geophysics to analyse seismic waves, coining the term "wavelet". Fourier integrals have a fixed wavelet —the sinusoid— and are integrated over all time giving only global features [3]. Local information may be lost or obscured. Short-time FT (STFT) introduces a sliding time window of fixed width which partly resolves this issue [5]. Wavelet transforms (WTs), however, are not slaves to a single wavelet type; there exist many different types that can be chosen at will, such as Morlet, and Complex Morlet, Gaussian- and Complex Gaussian-derivative wavelets. I shall focus on the Complex Morlet and Mexican Hat wavelets (second Gaussian derivative). Unlike STFTs, the sliding window deployed by WTs is of variable width. This produces a time-frequency distribution where the individual signal components are more effectively separated, offering both high temporal and spatial resolution [17]. A wavelet transform is defined as

$$T(a,b)(x(t)) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)\psi^*(\frac{t-b}{a})dt, \tag{7}$$

where a is the dilation parameter, b the location of the wavelet, and ψ^* is the complex conjugate of a given wavelet, ψ, satisfying the following properties

1. has finite total energy:
$$E = \int_{-\infty}^{\infty} |\psi(t)|^2 dt < \infty,$$
2. satisfies the 'admissibility condition'
$$C_g = \int_0^{\infty} \frac{|\hat{\psi}(f)|^2}{f} df < \infty \implies \hat{\psi}(0) = 0,$$
where $\hat{\psi}(f)$ is the Fourier transform of $\psi(t)$, and
3. the FT of complex wavelets must be real and vanish for negative frequencies [5].

The Complex Morlet wavelet is

$$\psi(t) = \frac{1}{\sqrt[4]{\pi}}(e^{i\omega_0 t} - e^{\frac{\omega_0^2}{2}})e^{-\frac{t^2}{2}} \tag{8}$$

where ω_0 is the natural frequency of the wavelet, and the Mexican Hat wavelet is defined as

$$\psi(t) = \frac{2}{\sqrt{3}\sqrt[4]{\pi}} e^{-\frac{t^2}{2}(1-t^2)}, \tag{9}$$

and is just the second derivative of a Gaussian wavepacket, normalised to give unit energy. Both of these wavelets were trialled for ECG classification as they produce distinct scalograms, as can be seen in fig.6.

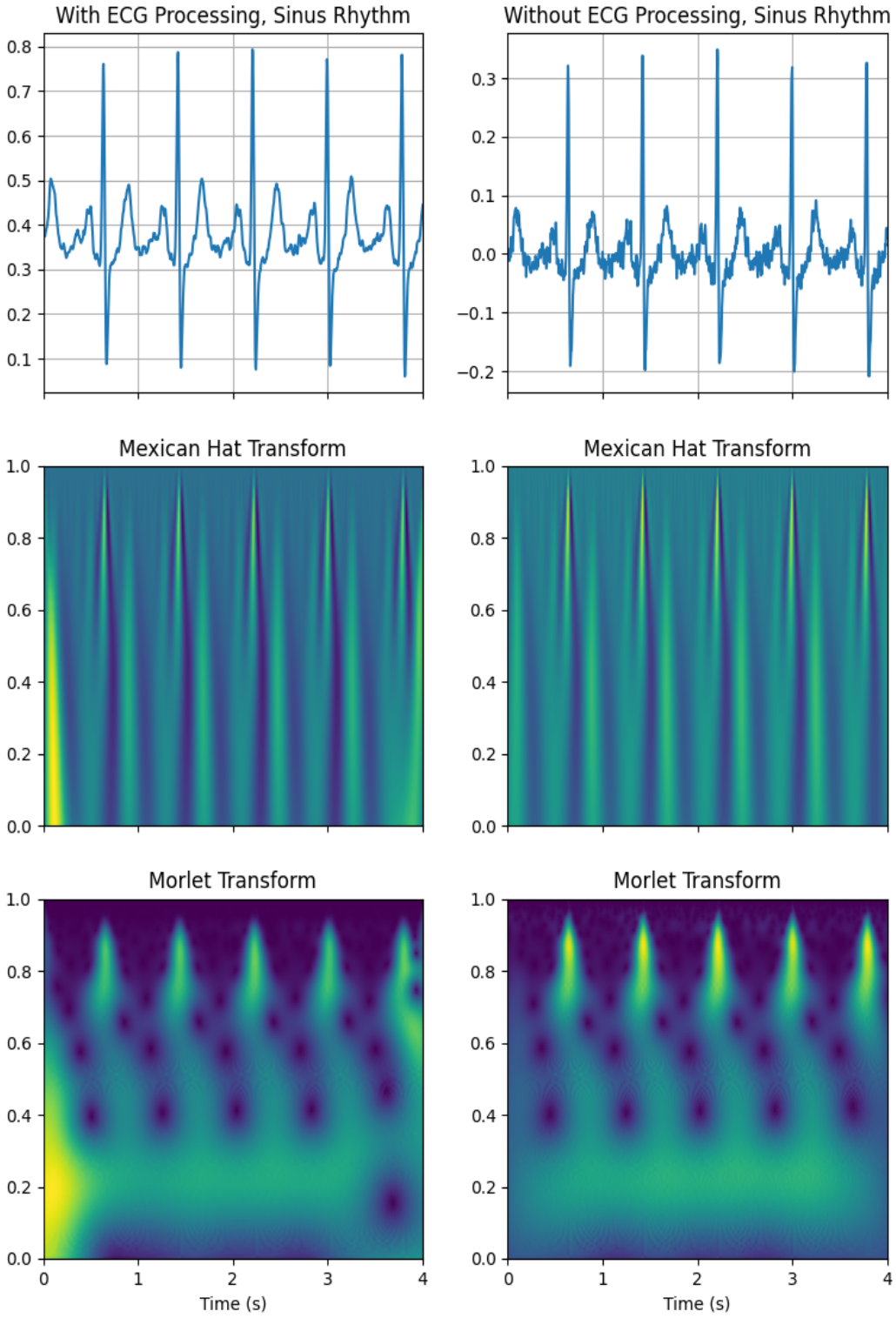

Figure 6: Processing a lead-2 ECG signal from a patient with SR

3 The Dataset

The dataset used in this work is the 2018 'China Physiological Signal Challenge' (CPSC), comprising '9,831 records from 9458 patients' lasting between 6-30s, collected from '11 hospitals', and containing one normal and eight abnormal ECG classes: atrial fibrillation (AF), type 1 atrioventricular block (I-AVB), left- and right bundle branch block (LBBB, RBBB), premature atrial contraction (PAC), premature ventricular contraction (PVC), ST-segment depression STD, and ST-segment elevation (STE). The train-test split is 70-30. ECG data is collected from 12 leads at a sampling frequency of 500Hz, and is stored in '.mat' file format [44]. This dataset was chosen due to it's relatively-large number of unique patients.

3.1 Method

A 42-layer-deep 2D-CNN with two channel attention modules is constructed that takes an ECG signal as input and outputs a vector of probabilities in the 9D output space, corresponding to the 9 CA rhythm types (including SR).

The raw ECG signals are 6-30s long and sampled at 500Hz. Before being passed to the CNN they are all attenuated to $4s$ assuming that an arrhythmia will present itself within these first 4 seconds (c.f. discussion), sampling reduced to 250Hz, and normalised before smoothing by Fast-Fourier transform. Fig.7 shows the result of this. The signal is then converted to a 128x128 image in the frequency-time domain by a CWT before being passed through the network. Both Mexican Hat and Complex-Morlet wavelets were considered and tested (c.f. fig.6). The model consists of 2 convolutional blocks, each followed by a channel-attention module. Each convolutional block contains two sub-blocks and ends with a max pooling layer. The sub-blocks in turn contain a convolution layer followed a ReLU activation function and batch normalisation. The second last layer includes a dropout layer with dropout parameter 0.5, and the final fully connected layer uses the Softmax activation function to output the vector of probabilities spanning the 9 CA classes.

The model was trained for 50 epochs on a single NVIDIA Tesla P100 GPU on Google Colab. Stochastic gradient decent (SGD) using the Nesterov algorithm was used to optimise the parameters, and the learning rate was decayed by a factor of 10 once the loss failed to reduce for more than 5 epochs, to a minimum size of 10^{-4}. Three-hundred samples were taking from the training set for testing to get preliminary test results.

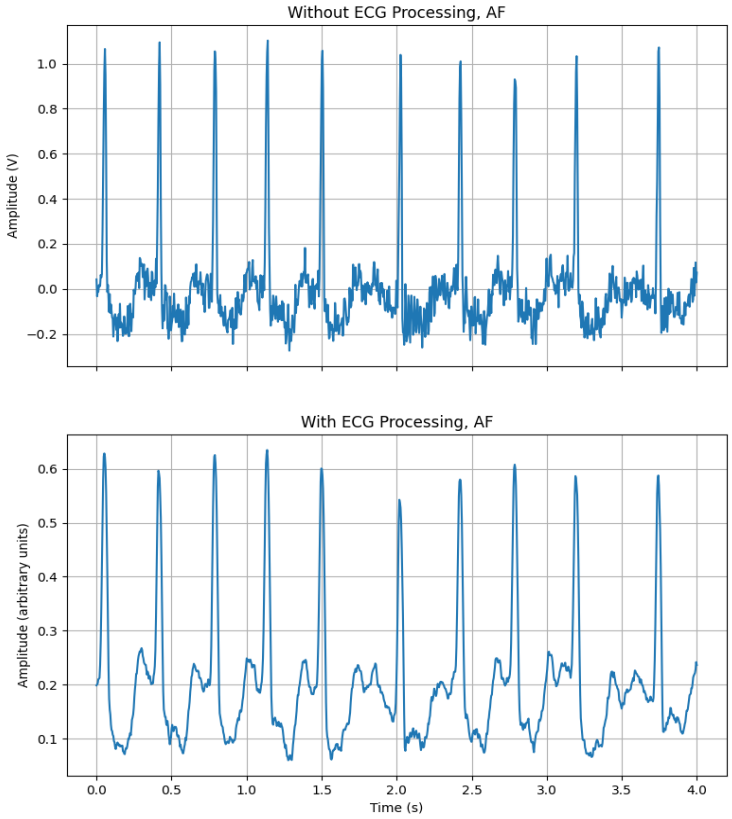

Figure 7: processing a single ECG displaying AF

3.2 Results

To draw comparisons with other groups in the study [44], results were assessed using an F1 score, defined for the i^{th} class as

$$F_{1i} = \frac{2N_{ii}}{\sum_j (N_{ij} + N_{ji})} \tag{10}$$

and an overall $F1$ score given by the average of all the $F1$ scores,

$$F_1 = \frac{1}{9} \sum_{i=1}^{9} F_{1i} \tag{11}$$

where SR, AF, I-AVB, LBBB, RBBB, PAC, PVC, STD, and STE are denoted by classes 1 to 9 respectively. In addition to this, F1 scores were given to the super-types consisting of heart blocks (block)

$$F_{block} = \frac{2(N_{33} + N_{44} + N_{55})}{\sum_j (N_{3j} + N_{j3} + N_{4j} + N_{j4} + N_{5j} + N_{j5})},$$

premature contractions (PC)

$$F_{ST} = \frac{2(N_{88} + N_{99})}{\sum_j (N_{8j} + N_{j8} + N_{9j} + N_{j9})},$$

and ST-segment changes (ST)

$$F_{PC} = \frac{2(N_{66} + N_{77})}{\sum_j (N_{6j} + N_{j6} + N_{7j} + N_{j7})}$$

Table 1 shows a comparison of the results obtained on leads 4 and 7 using a Mexican hat wavelet, and on lead-4 using a complex-Morlet wavelet. Preliminary results are promising, achieving highest scores in AF and PC detection on lead-7 using the Mexican hat wavelet when comparing with other competitors. The results also show that different leads are better than others as identifying CA, with lead-7 scoring higher than lead-4 in all super-type classifications. The complex-Morlet wavelet also performed worse than the Mexican hat wavelet when comparisons were made on lead-4, showing that wavelets with similar morphologies to the input signal perform better than those without [17].

Results					
Group	F1	F_{AF}	F_{Block}	F_{pc}	F_{st}
Tsai-Min Chen et al.	0.837	0.933	0.899	0.847	0.779
Wenjie Cai et al.	0.830	0.931	0.912	0.817	0.761
Runnan He et al.	0.806	0.914	0.879	0.801	0.742
Yue Yu et al.	0.802	0.918	0.890	0.789	0.718
Yangyang Yan	0.791	0.924	0.882	0.779	0.709
Mexh lead 4	0.670	0.905	0.894	0.621	0.697
Mexh lead 7	0.778	**0.945**	0.893	**0.851**	0.747
Cmor lead 4	0.614	0.860	0.828	0.494	0.641

Table 1: Comparison of my results against the top-5 participants of the challenge

The confusion matrix shown in fig.9 for lead 4 helps illuminate some of the problems with this method, and where the model is performing well. The largest error comes from the falsely predicting SR when the correct label is STE. An additional source of error is due to the CA not being present in the first 4s of the ECG signal as is illustrated in fig.8. Fig.8 also highlights a concern with the normalisation process of the raw ECG signal. The presence of the large spike has reduced the amplitude of all other spikes, and rescaled the data in such a way as to remove the naturally-occuring ST-depression. This may also have effected the ST-elevation leading to other incorrect predictions, and tests will need to be run to verify whether or not this is the case.

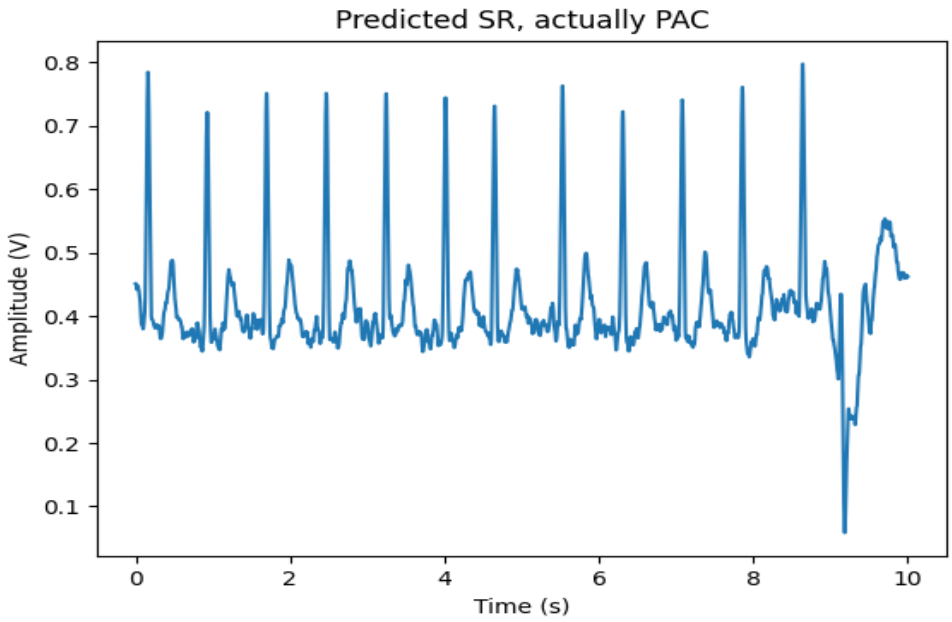

Figure 8: PAC occurring later than 4s into the ECG (lead-4) and is therefore incorrectly identified as SR. In fact, the first 4s are correctly diagnosed as SR.

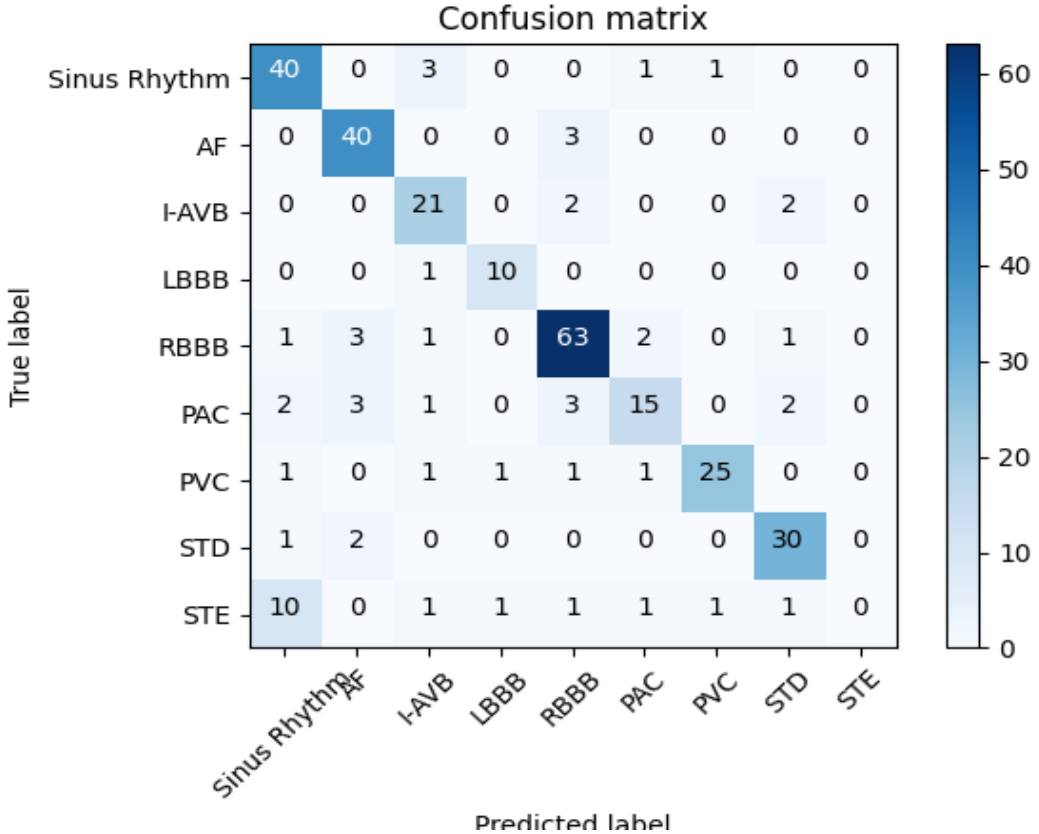

Figure 9: Confusion matrix for the 300 test samples showing the most common cause for confusion to be between STE and SR; Lead-4.

3.3 Conclusion

A 42-layer deep 2D-CNN with 2 channel attention modules was designed for a 9-class ECG classification task. ECG signals were cut down to 4s in length, denoised using FFT and normalised to values between 0 and 1 before being transformed to 128x128 images in the frequency-time domain by CWT. A separate model was trained on leads 4 and 7 using the Mexican hat wavelet, and on lead-4 using the complex-Morlet wavelet. The models were trained for 50 epochs using mini-batch SGD to update the weights, and cross-entropy loss was minimised. The model trained on lead-7 outperformed that which was trained on lead-4 in all classifications, and the Mexican hat wavelet outperformed the complex-Morlet wavelet on lead-4. Lead-7 achieved a top score in the detection of atrial fibrillation and premature contractions - fig.1. However, I did not have access to the test-set that was used in CPSC-2018 and so comparisons should not be considered decisive. The fact that the models trained on different leads comparatively performed better than others at classifying certain CA types suggests that an ensemble model could be developed to yield higher classification scores overall. Due to the increased computational cost of working with complex numbers, and due to preliminary results showing less accuracy, all future use of complex-Morlet CWTs will need to be carefully evaluated.

3.4 Further Work and Discussion

During the preprocessing, ECGs are trimmed down to $4s$ in length. Fig.8 reveals that not all CAs present themselves in this window. If the dataset had individually-labelled beats or abnormal rhythm sections then care could be taken to ensure that CAs are present in every image being processed. This would certainly improve training and increase the predictive ability of the model, and selecting a dataset that satisfies these properties will be made a priority in future works. However, AF seems to consistently presents itself in the first 4s and is accurately predicted by the model, suggesting that the model itself is not redundant. Class weights were not implemented in the model, and addressing this imbalance could improve training and classification performance [14]. The data is collected from 12 ECG leads. Future work will aim to compare results from all leads and deduce the optimal lead for identifying each CA type. An ensemble model can then be constructed to utilise these advantages.

Furthermore, transforming to, and the processing of, 2D images is computationally costly, especially when involving complex numbers such as with the complex-Morlet transform. Reducing computational cost by use of a 1D model will help to speed up the training and allow for a wider range of hyperparameter tuning, as well as being able to train with more epochs. LSTM models have already been successfully applied to ECG classification tasks [32] and could be implemented. However, a more modern transformer architecture based on self-attention [67] could also be considered and this would provide a novel technique of ECG classification in addition to having the potential to outperform current state-of-the-art models while simultaneously requiring even fewer computational resources [39].

References

[1] Learning to combine foveal glimpses with a third-order boltzmann machine — proceedings of the 23rd international conference on neural information processing systems - volume 1, pages 1243-1251, 12 2010.

[2] Crossentropyloss — pytorch 1.10.1 documentation, 2019.

[3] Cardiac electrophysiology: normal and ischemic ionic currents and the ecg — advances in physiology education, 2020.

[4] How the heart works — nhlbi, nih, 04 2021.

[5] Paul S Addison. Wavelet transforms and the ecg: a review. *Physiological Measurement*, 26:R155–R199, 08 2005.

[6] A.K. Agarwal and P. Venugopalan. Right bundle branch block: varying electrocardiographic patterns. *International Journal of Cardiology*, 71:33–39, 09 1999.

[7] Laith Alzubaidi, Jinglan Zhang, Amjad J. Humaidi, Ayad Al-Dujaili, Ye Duan, Omran Al-Shamma, J. Santamaría, Mohammed A. Fadhel, Muthana Al-Amidie, and Laith Farhan. Review of deep learning: concepts, cnn architectures, challenges, applications, future directions. *Journal of Big Data*, 8, 03 2021.

[8] Charles Antzelevitch and Alexander Burashnikov. Overview of basic mechanisms of cardiac arrhythmia. *Cardiac Electrophysiology Clinics*, 3:23–45, 03 2011.

[9] Adria Arboix and Josefina Alioc. Cardioembolic stroke: Clinical features, specific cardiac disorders and prognosis. *Current Cardiology Reviews*, 6:150–161, 08 2010.

[10] Daniel C. Bartos, Eleonora Grandi, and Crystal M. Ripplinger. Ion channels in the heart. *Comprehensive Physiology*, pages 1423–1464, 06 2015.

[11] Eve Belzile-Dugas and Mark J. Eisenberg. Radiation-induced cardiovascular disease: Review of an underrecognized pathology. *Journal of the American Heart Association*, 10, 09 2021.

[12] Adam Bohr and Kaveh Memarzadeh. The rise of artificial intelligence in healthcare applications. *Artificial Intelligence in Healthcare*, pages 25–60, 2020.

[13] Matthew Botvinick, Sam Ritter, Jane X. Wang, Zeb Kurth-Nelson, Charles Blundell, and Demis Hassabis. Reinforcement learning, fast and slow. *Trends in Cognitive Sciences*, 23:408–422, 05 2019.

[14] Mateusz Buda, Atsuto Maki, and Maciej A. Mazurowski. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, 106:249–259, 10 2018.

[15] Ed Burns and Robert Buttner. Right bundle branch block (rbbb), 2021.

[16] S Chakrabarti and A G Stuart. Understanding cardiac arrhythmias. *Archives of Disease in Childhood*, 90:1086–1090, 06 2005.

[17] Gari D Clifford, Francisco Azuaje, and Patrick Mcsharry. *Advanced methods and tools for ECG data analysis*. Artech House, 2006.

[18] David A. Cook, So-Young Oh, and Martin V. Pusic. Accuracy of physicians' electrocardiogram interpretations. *JAMA Internal Medicine*, 180:1461, 11 2020.

[19] Wael Dakkak and Rami Doukky. Sick sinus syndrome, 07 2021.

[20] DeepLizard. Backpropagation explained — part 5 - what puts the "back" in backprop?

[21] Eser et al. Durmaz. The clinical significance of premature atrial contractions: How frequent should they become predictive of new-onset atrial fibrillation. *Annals of Noninvasive Electrocardiology*, 25, 10 2019.

[22] Python Machine Learning Third Edition. *PYTHON MACHINE LEARNING - THIRD EDITION : machine learning and deep learning with python, scikit... -learn, and tensorflow 2.* Packt Publishing Limited, 2019.

[23] IBM Cloud Education. What is unsupervised learning?, 09 2020.

[24] Andre Esteva, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau, and Sebastian Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542:115–118, 01 2017.

[25] Dr. Joel A. Kirsh et al. University health network and sickkids, utoronto, ecg, 2015.

[26] Katelyn M. Atkins et al. Mean heart dose is an inadequate surrogate for left anterior descending coronary artery dose and the risk of major adverse cardiac events in lung cancer radiation therapy. *International Journal of Radiation Oncology*Biology*Physics*, 110(5):1473–1479, 2021.

[27] Khashayar Farzam and John R Richards. Premature ventricular contraction, 08 2021.

[28] British Heart Foundation. Arrhythmias, 2021.

[29] British Heart Foundation. Atrial fibrillation (af) : causes, symptoms and treatments, 2021.

[30] Alan S. Go, Elaine M. Hylek, Kathleen A. Phillips, YuChiao Chang, Lori E. Henault, Joe V. Selby, and Daniel E. Singer. Prevalence of diagnosed atrial fibrillation in adults. *JAMA*, 285:2370, 05 2001.

[31] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification, 2015.

[32] Borui Hou, Jianyong Yang, Pu Wang, and Ruqiang Yan. Lstm-based auto-encoder model for ecg arrhythmias classification. *IEEE Transactions on Instrumentation and Measurement*, 69:1232–1240, 04 2020.

[33] Patrick Houthuizen, Frank A. L. E. Bracke, and Berry M. van Gelder. Atrioventricular and interventricular delay optimization in cardiac resynchronization therapy: physiological principles and overview of available methods. *Heart Failure Reviews*, 16:263–276, 12 2010.

[34] D. H. Hubel and T. N. Wiesel. Receptive fields of single neurones in the cat's striate cortex. *The Journal of Physiology*, 148:574–591, 10 1959.

[35] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift, 2015.

[36] Christian Janiesch, Patrick Zschech, and Kai Heinrich. Machine learning and deep learning. *Electronic Markets*, 31:685–695, 04 2021.

[37] Evans R. Pritzel A. et al. Jumper, J. Highly accurate protein structure prediction with alphafold, 07 2021.

[38] Anthony H Kashou, Hajira Basit, and Ahmad Malik. St segment, 08 2021.

[39] Salman Khan, Muzammal Naseer, Munawar Hayat, Waqas Syed, Shahbaz Zamir, Mubarak Khan, and Shah. Transformers in vision: A survey, 2021.

[40] Inara Koppert-Anisimova. Cross-entropy loss in ml - unpackai - medium, 01 2021.

[41] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1:541–551, 12 1989.

[42] Yann LeCun. Une procedure d'apprentissage pour reseau a seuil asymmetrique (a learning scheme for asymmetric threshold networks, 1985.

[43] Giuseppe Lippi, Fabian Sanchis-Gomar, and Gianfranco Cervellin. Global epidemiology of atrial fibrillation: An increasing epidemic and public health challenge. *International Journal of Stroke*, 16:217–221, 01 2020.

[44] Feifei et al. Liu. An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection. *Journal of Medical Imaging and Health Informatics*, 8:1368–1373, 09 2018.

[45] Ngai-Sang Lok and Chu-Pak Lau. Prevalence of palpitations, cardiac arrhythmias and their associated risk factors in ambulant elderly. *International Journal of Cardiology*, 54:231–236, 06 1996.

[46] Eduardo José da S. Luz, William Robson Schwartz, Guillermo Cámara-Chávez, and David Menotti. Ecg-based heartbeat classification for arrhythmia detection: A survey. *Computer Methods and Programs in Biomedicine*, 127:144–164, 04 2016.

[47] D. Douglas Miller and Eric W. Brown. Artificial intelligence in medical practice: The question to the answer? *The American Journal of Medicine*, 131:129–133, 02 2018.

[48] Sebastian Mirjalili and p.1-3 Mirjalili, Vahid. *Python Machine Learning - Third Edition*. Packt Publishing Limited, 2019.

[49] Sebastian Mirjalili and p.383 Mirjalili, Vahid. *Python Machine Learning - Third Edition*. Packt Publishing Limited, 2019.

[50] Sebastian Mirjalili and p530 Mirjalili, Vahid. *Python Machine Learning - Third Edition*. Packt Publishing Limited, 2019.

[51] Sebastian Mirjalili and p.534 Mirjalili, Vahid. *Python Machine Learning - Third Edition*. Packt Publishing Limited, 2019.

[52] Sebastian Mirjalili and p6 Mirjalili, Vahid. *Python Machine Learning - Third Edition : machine learning and deep learning with python, scikit... -learn, and tensorflow 2*. Packt Publishing Limited, 2019.

[53] American Academy of Ophthalmology. Fovea, 08 2017.

[54] Sean H Oldroyd, Bryan S Quintanilla, and Amgad N Makaryus. First degree heart block, 09 2021.

[55] World Health Organization. Cardiovascular diseases, 06 2019.

[56] World Health Organization. Cardiovascular diseases (cvds), 06 2021.

[57] Ana Pardo Sanz and José Luis Zamorano. 'cardiotoxicity': time to define new targets? *European Heart Journal*, 41:1730–1732, 02 2020.

[58] Pranav Rajpurkar, Awni Y Hannun, Masoumeh Haghpanahi, Codie Bourn, and Andrew Y Ng. Cardiologist-level arrhythmia detection with convolutional neural networks, 2017.

[59] Siamak Ravanbakhsh, Jeff Schneider, and Barnabás Póczos. Equivariance through parameter-sharing, 2017.

[60] Gregory A. et al. Roth. Global burden of cardiovascular diseases and risk factors, 1990–2019. *Journal of the American College of Cardiology*, 76:2982–3021, 12 2020.

[61] Jonathan Schmidt, Mário R. G. Marques, Silvana Botti, and Miguel A. L. Marques. Recent advances and applications of machine learning in solid-state materials science. *npj Computational Materials*, 5, 08 2019.

[62] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ruslan Salakhutdinov, and Ilya Sutskever. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15:1929–1958, 2014.

[63] Camilla Funch Staugaard, Anders Petersen, and Signe Vangkilde. Eccentricity effects in vision and attention. *Neuropsychologia*, 92:69–78, 11 2016.

[64] Faisal F. Syed, Jo Jo Hai, Nirusha Lachman, Christopher V. DeSimone, and Samuel J. Asirvatham. The infrahisian conduction system and endocavitary cardiac structures: relevance for the invasive electrophysiologist. *Journal of Interventional Cardiac Electrophysiology*, 39:45–56, 12 2013.

[65] Nicholas Y. Tan, Chance M. Witt, Jae K. Oh, and Yong-Mei Cha. Left bundle branch block. *Circulation: Arrhythmia and Electrophysiology*, 13, 04 2020.

[66] Markus Thom, Guenther Palm@uni-Ulm, and De. Sparse activity and sparse connectivity in supervised learning günther palm. *Journal of Machine Learning Research*, 14:1091–1143, 2013.

[67] Ashish Vaswani, Google Brain, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2017.

[68] Oriol et al. Vinyals. Grandmaster level in StarCraft II using multi-agent reinforcement learning. *Nature*, 575(7782):350–354, 2019.

[69] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module, 2018.

[70] Rikiya Yamashita, Mizuho Nishio, Richard Kinh Gian Do, and Kaori Togashi. Convolutional neural networks: an overview and application in radiology. *Insights into Imaging*, 9:611–629, 06 2018.

[71] Xiao Yang. An overview of the attention mechanisms in computer vision. *Journal of Physics: Conference Series*, 1693:012173, 12 2020.

[72] Xinyu et al. Yang. Anticancer therapy-induced atrial fibrillation: Electrophysiology and related mechanisms. *Frontiers in Pharmacology*, 9, 10 2018.