Developing a Convolutional Neural Network Classifier to Identify Atrial Fibrillation Rhythms from Electrocardiogram Recordings

Iva and Ella Chee

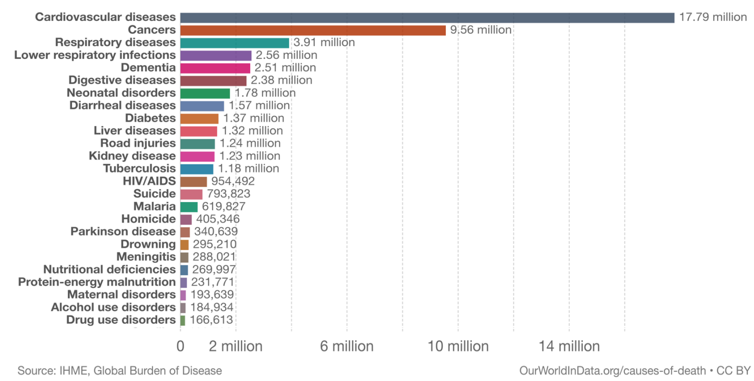
Yorktown High School

INTRODUCTION

Overview of Research Problem

Cardiovascular diseases are the main cause of death across the globe, as seen in Figure 1, with more than 17 million people dying each year (Ritchie, 2018). Furthermore, the prevalence and mortality of cardiovascular diseases continue to grow, underscoring the importance and necessity of regular heart rhythm monitoring to manage and curb cardiovascular diseases (World Health Organization, 2017). Electrocardiograms or ECGs record the electrical activity of the heart, allowing for the discovery of abnormalities in the heart's rhythm and making it the central method of heart disease diagnosis.

Figure 1. Number of Deaths by Cause, World, 2017



Cardiovascular diseases, such as arrhythmias, are the leading cause of death worldwide (Ritchie, 2018).

Cardiac arrhythmia is one common cardiovascular disease, characterized by an irregular rhythm of the heartbeat when electrical impulses in the heart fail to work properly. Moreover, atrial fibrillation is a major cardiac arrhythmia, shown through fast irregular heartbeats of the atria, and though common, the condition can go misdiagnosed and is associated with a greater risk for stroke, cardiac arrest, and death. About 2.3 million people in the U.S. have atrial fibrillation and the number is expected to increase 2.5 times by the year 2050 (Lainscak et al., 2008). Given these predictions, atrial fibrillation presents a critical issue, and diagnosis methods from ECGs require more research.

Review of Literature

Currently, screening methods for atrial fibrillation are challenging—the ability to diagnose momentary arrhythmia from a single electrocardiogram (diagnostic yield) is low, and current prolonged monitoring methods are incommodious (Attia et al., 2019). Several algorithms have been developed utilizing deep learning to process ECGs and classify the conditions or diseases they present (Krittanawong et al., 2019). Recently, deep learning techniques have accumulated much attention in biomedical engineering, in specific, the convolutional neural network (CNN), due to its variety of successful applications. Several studies have applied convolutional neural networks to detect, predict, or classify arrhythmia from electrocardiograms.

2D-CNN-LSTM Technology for Arrhythmia Diagnosis via ECG

In a recent study for arrhythmia diagnosis via electrocardiogram, Zheng et al. (2020) used a two-dimensional convolutional neural network with long short-term memory (LSTM) which allowed for greater capability of retaining and feeding back information from selectively stored information. In classifying eight ECG beat signals and normal sinus rhythm, they were able to achieve 99.01% accuracy, 99.57% specificity, and 97.67% sensitivity, confirming the efficacy of 2D-CNN-LSTM technology in helping doctors diagnose arrhythmias with ease (Zheng et al., 2020). While this model classified with very high performance, it classified beat type arrhythmias (i.e.: atrial premature beat, premature ventricular contraction, paced beat, ventricular escape beat), rather than rhythm types (i.e.: sinus rhythm, atrial fibrillation, ventricular fibrillation).

1D-CNN for Prediction of AF from ECG

Erdenebayar et al. (2019) designed a one-dimensional CNN model to predict atrial fibrillation using a short-term ECG signal (less than 60 seconds). Between training and test sets consisting of ECGs displaying atrial fibrillation and normal sinus rhythm, the convolutional neural network predicted atrial fibrillation with high performance: 98.6% sensitivity, 98.7% specificity, and 98.7% accuracy. In the study, results supported the possibility of predicting atrial fibrillation from a short-term normal ECG signal (Erdenebayar et al., 2019). However, one of the shortcomings of this study was that atrial fibrillation was predicted from datasets containing only that and normal sinus rhythm, whereas a more realistic dataset would include several other types of arrhythmias.

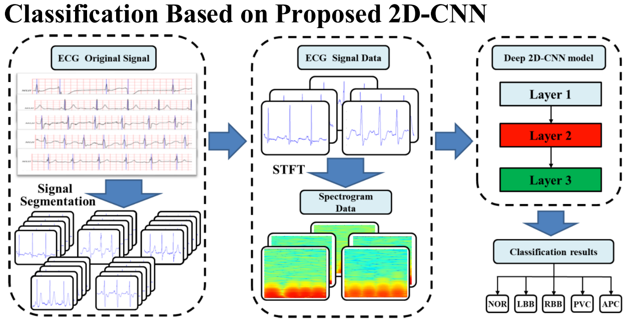
2D-CNN for Arrhythmia Classification via ECG

In a 2017 study, a dataset was built with over 500 times the number of patients than previous studies and was used to train a 34-layer two-dimensional convolutional neural network that maps ECG samples to a sequence of rhythm categories. The developed algorithm's performances in both sensitivity and positive predictive value were superior to that of a board of certified cardiologists in the detection of several cardiac arrhythmias from electrocardiograms (Hannun et al., 2017). Possibly due to the large dataset, the CNN struggled more in terms of sensitivity and F1 score for its performance in classifying arrhythmia rhythms, and in specific, atrial fibrillation.

2D-CNN for Arrhythmia Classification via ECG Spectrograms

A paper by Huang et al. (2019) offered a method for ECG arrhythmia classification utilizing a two-dimensional deep convolutional neural network, as seen in Figure 2. Short-time Fourier transform was used to convert time-domain signals of the ECGs into time-

Figure 2. Overall Procedures in ECG Arrhythmia Classification Based on Proposed 2D-CNN



One method for arrhythmia detection, where the ECG signal is segmented, transformed into spectrogram data, input into a 2D-CNN model, and classified (Huang et al., 2019).

frequency spectrograms. Then, spectrograms of the five arrhythmia beat types were used as input for arrhythmia identification and classification by the 2D-CNN. In this study, it was found that the two-dimensional model could reach a 99.00% averaged accuracy. These results validated the improved classification accuracy from a CNN classifier utilizing ECG spectrograms as input, although it only identified by beat type, not rhythm (Huang et al., 2019).

Between the use of long short-term memory technology, one-dimensional convolutional neural networks with short-term ECG data, deep CNNs with large datasets compared to expert human cardiologists, and spectrogram representations of ECGs with two-dimensional CNNs, there any many different ways convolutional neural networks can be applied in tandem with electrocardiogram data. Despite the extensive research involving CNN applications with ECGs, few CNNs have classified arrhythmias by rhythm type (i.e.: sinus rhythm, atrial fibrillation, ventricular fibrillation) rather than beat type, and those that did were not as accurate or used unrealistic datasets.

Problem Statements

- Screening methods for atrial fibrillation from ECGs are lacking; low diagnostic yield of a single ECG to detect momentary arrhythmia, prolonged monitoring methods
- GAP: few CNNs classify arrhythmias by rhythm type; past studies focus on beat type and those that did use rhythm type were not as accurate or used less realistic data

Objective

- Create and evaluate the effectiveness of a deep neural network that automatically detects and identifies atrial fibrillation rhythms from short-term electrocardiogram recordings, in order to advance diagnostic procedures

- Classify arrhythmias from electrocardiograms by rhythm type: atrial fibrillation or not AF (normal sinus rhythm or other arrhythmias), from ECGs containing various types of rhythms

Hypothesis

- A convolutional neural network can be trained to identify atrial fibrillation present in a short-term 30-second 12-lead ECG from a dataset of patients with various arrhythmias and healthy sinus rhythms
- ECGs from patients with various arrhythmias can be used to distinguish atrial fibrillation rhythm arrhythmia from normal sinus rhythm or other arrhythmias

METHODOLOGY

Role of Student vs. Mentor

Over the course of six weeks during the past summer (over 90 hours), we were responsible for preprocessing the ECG signals, constructing the convolutional neural network, training/testing the model, and completing data analysis to evaluate its performance. Our mentor supervised the processes above and provided guidance and assistance for more specific programming, as well as troubleshooting errors.

Equipment & Materials

Data Collection

The short-term ECG signal was used as the input for the CNN model. The ECG samples were taken from two PhysioNet databases: the MIT-BIH Arrhythmia Database and the MIT-BIH Atrial Fibrillation Database. These databases, containing the first widely available ECG material, were obtained from labs at Boston’s Beth Israel Hospital and MIT between 1975 and 1980. That said, many studies have used these data for arrhythmia detection via ECG, and methods for identification have been greatly improved. The MIT-BIH Arrhythmia Database consisted of 48 annotated, half-hour excerpts of 12-lead, two-channel ambulatory ECG recordings, taken from 47 subjects at 360 Hz. The MIT-BIH Atrial Fibrillation Database contained 25 annotated, 10-hour ECG records from atrial fibrillation patients taken with a Holter monitor at a sampling frequency of 250 Hz. Due to a lack of information and annotations for two records of the Atrial Fibrillation Database, they were excluded. A combination of data from the two databases allowed for a more realistic representation of arrhythmias—the classifier sorted between samples containing several types of arrhythmia, atrial fibrillation, and normal sinus rhythm.

CNN Implementation

The convolutional neural network was developed in Python using the Keras framework and Tensorflow backend. Python waveform-database, NumPy, and SciPy packages were also utilized for data analysis.

Procedures

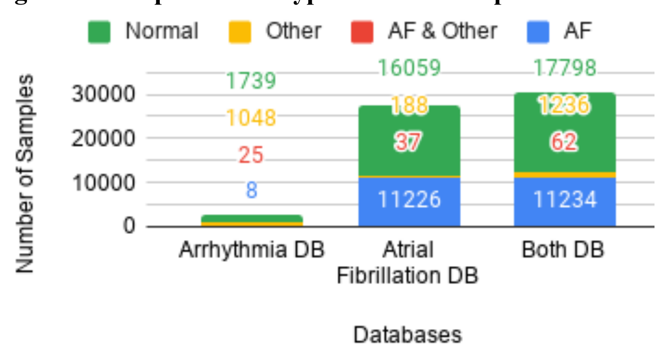
Data Processing and Segmentation

Since Atrial Fibrillation records were at a sampling frequency of 250 Hz, they were resampled to 360 Hz using fast Fourier transform, in order to match the frequency of the Arrhythmia ECG records. ECG records contained two channels: the upper signal being a modified limb lead II (MLII) and the lower being a modified lead V1 (as well as V2, V5, and V4). As a result, the upper signal, MLII, was extracted for use while the lower signal was excluded, due to the fact that normal beats are often difficult to detect in the lower signal, and to simplify classification in the CNN (rather than having to extract information from two sets of data) (Moody & Mark, 2001). The ECG records were then segmented into samples of 30 seconds, resulting in a total of 30,480 short-term samples, which were used as the input for the neural network model. Of those samples, 60 contained noise, a change in signal quality, missed beats, pauses, or tape slippage, and 90 contained no data. Those defective samples were therefore removed from the dataset, resulting in 30,330 remaining segments. The samples were then organized into the following groups based on annotations: atrial fibrillation ('A'), atrial fibrillation and other arrhythmias ('B'), other arrhythmias ('O'), and normal sinus rhythm ('N'), as seen in Table 1 and Figure 3. The group, 'other arrhythmias,' included atrial bigeminy, atrial flutter, ventricular bigeminy, nodal (AV junctional) rhythm, and ventricular flutter, among other arrhythmias. 5-fold cross-validation (CV) was carried out in order to obtain training and testing sets with an 80% to 20% proportion, respectively. Overall, compared to previous studies, we conducted very little preprocessing; we believed that the convolutional neural network should be able to distinguish AF rhythms from other potential noise and artifacts within the signal, which would hopefully improve its efficiency.

Table 1. Types of ECG Samples from MIT-BIH Databases
Samples containing AF, AF and other arrhythmias, other arrhythmias, and sinus rhythm found in the MIT-BIH databases, and the combined databases (Source: Student Researchers).

Type of Sample/Database	AF	AF & Other	Other	Normal
Arrhythmia DB	8	25	1048	1739
Atrial Fibrillation DB	11226	37	188	16059
Both DB	11234	62	1236	17798

Figure 3. Comparison of Types of ECG Samples

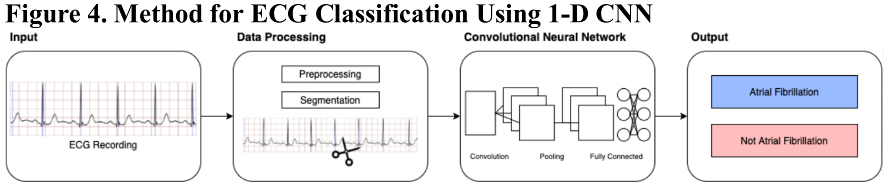


Comparison of ECGs from MIT-BIH Arrhythmia and AF databases, and the combination of both. Sample types include sinus rhythm, other arrhythmias, AF and other arrhythmias, and AF (Source: Student Researchers).

CNN Model Architecture

The convolutional neural network model was designed to identify AF from a short-term ECG (Figure 4). The CNN comprised three primary layers—a convolutional layer, a pooling layer, and fully-connected layers. First, the convolutional layer created feature maps by merging the whole input signal using several kernels. Since the input ECG signal was a 1D time series, a one-dimensional convolution was used. The proposed CNN model contained four 1D convolutional layers with a kernel size of 36, 100 filters, a stride of 1, an input size of 10,800 x 1, and batch size of 24, as seen in Table 2. The kernel size and input sizes were chosen based on the sampling frequency: because it was 360 Hz, the input size had dimensions of 10,800 (10,800 samples in a 30s ECG at 360 Hz). The kernel size and batch size were chosen based on the hyperparameter tuning experiments. Then, the pooling layer chose representative features by decreasing the size of the feature maps with max-pooling. The 1D max-pooling layers contained a kernel size of 36 and had a pool size of 3. Two more convolutional layers were added, followed by a global average pooling layer with a kernel size of 36. A dropout layer was also included to prevent overfitting, with a rate of 0.5. Finally, the fully-connected layer contained a dense layer and sigmoid activation function for the concluding differentiation of the model (binary classification between ‘not atrial fibrillation’ and ‘atrial

Figure 4. Method for ECG Classification Using 1-D CNN



The proposed method for ECG classification. Input will be a segmented 30-second ECG recording, and the output of the 1D-CNN will be a classification of either AF or not AF (Source: Student Researchers).

Table 2. Overview of Proposed 1D-CNN Model
Types of layers and parameters used for the classifier (Source: Student Researchers).

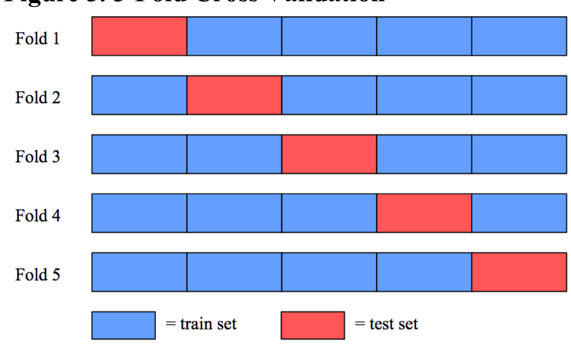
Layer	Type	Kernel Size	Stride	Filters	Input Size
1	Conv1D	36	1	50	10800 x 1
2	Conv1D	36	1	50	10800 x 1
3	MaxPooling1D	36	1		
4	Conv1D	36	1	50	10800 x 1
5	Conv1D	36	1	50	10800 x 1
6	GlobalAveragePooling1D	36			
7	Dropout	36			
8	Fully-connected	36			

fibrillation'). The rectified linear unit (ReLU) activation function was used for the convolutional layers, as it is more computationally efficient, simplistic, and shows better convergence performance than other activation functions (Brownlee, 2019). The 1D-CNN model contained a total of 8 layers and 272,051 total trainable parameters.

Training and Testing

The training of the CNN model was fully-supervised (learning on a labeled dataset rather than finding patterns independently) and model hyperparameters were optimized using the Adam optimizer with binary cross-entropy as the loss function. We manually altered the batch size, kernel size, and the number of filters to achieve the highest accuracy, specificity, recall, and precision. 5-fold cross-validation was completed to evaluate the performance of the CNN model (Figure 5).

Figure 5. 5-Fold Cross Validation



Process of 5-fold CV: with 5 folds, data is split into fifths, and the fifth used for testing is alternated each fold, maximizes data used for training and testing (Source: Student Researchers).

Data Analysis

Accuracy, precision, sensitivity (or recall), specificity, and the F1 score were calculated to determine the performance of the CNN model. Table 3 shows the confusion matrix for the true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). Accuracy refers to (TP + TN)/(TP + TN + FP + FN); precision refers to TP/(TP + FP); recall refers to TP/(TP + FN); specificity refers to TN/(TN + FP). The accuracy represents the overall correctness of the classifications, while precision shows the fraction of the positive

Table 3. Confusion Matrix for ECG Classification
Confusion matrix for true positives, true negatives, false positives, and false negatives, and their corresponding meanings, for the metrics (Source: Student Researchers).

	Positive	**Negative**
True	AF classified as AF	normal classified as normal
False	normal classified as AF	AF classified as normal

predictions that are actually positive. Recall demonstrates the proportion of normal ECG data detected by the model to the total normal data (true positive rate) and specificity demonstrates the ratio of abnormal ECG data to the total abnormal data (false positive rate). AUC curves were used to plot the TP rate against the FP rate in order to show the relationship between recall and specificity, as well as evaluate the performance of the CNN. These C-statistics, ROC curve, recall, and specificity are a standard for model evaluation with cardiovascular medicine (Krittanawong et al., 2019). The performance metrics were then compared to those of previous studies identifying AF from ECG recordings using a convolutional neural network.

RESULTS

Data from the MIT-BIH Arrhythmia Database and MIT-BIH Atrial Fibrillation Database was used in this study. For the experiment, the data was divided into 80% for training and 20% for testing sets. Initially, training and testing of the model was done with 10 epochs, dropout regularization at a rate of 0.5, and an early stopping function. Early stopping was utilized with a patience argument set to 1 to end training once the model performance stopped improving over 1 epoch and thus prevent overfitting. Table 4 displays the performance metrics found for the initial

Table 4. Performance Metrics for 5-Fold Cross-Validation (10 epochs)
The metrics for 5-fold CV with Early Stopping at 10 epochs with a patience of 1 (Source: Student Researchers).

Fold	Loss	Accuracy	Precision	Recall	Specificity	F1 Score	AUC
1	0.0458	98.83%	98.60%	98.20%	99.19%	98.40%	0.9970
2	0.0532	98.57%	97.86%	98.34%	98.70%	98.10%	0.9972
3	0.0506	98.71%	98.54%	98.02%	99.13%	98.28%	0.9969
4	0.0566	98.29%	99.03%	96.29%	99.45%	97.64%	0.9967
5	0.0507	98.65%	99.02%	97.38%	99.42%	98.19%	0.9963
Average	**0.0514**	**98.61%**	**98.61%**	**97.64%**	**99.18%**	**98.12%**	**0.9969**

5-fold cross-validation. Although the metrics were high, further training with 20 epochs, patience set to 2, and a same dropout rate of 0.5 was conducted to achieve improved results. The second run saw improvement, mostly due to the increased patience and number of epochs that provided more time to train the model. Several performance metrics were utilized in testing to measure the ability of the model. The accuracy and loss curves of the training model are shown in Figure 6, and with the early stopping function, training ended early at 14 epochs. The model demonstrated a very high average accuracy (98.89%) and low average loss (0.04), indicating that the model performed well and accurately after each iteration of optimization. Additionally, precision, recall, and specificity were used as evaluation metrics for the performance of the model. The performance metrics for each fold of the cross-validation are shown in Table 5. The model achieved a 98.89% average accuracy (as mentioned before), 99.34% average precision, 97.66% average recall, and 99.62% average specificity. The F1 score was also calculated to be 98.50%, indicating an excellent balance between the precision and recall of the model. Area under the curve (AUC) measures the sensitivity and specificity tradeoff, showing the model's classification ability, and was found to be .9972, indicating an outstanding classification capability, as seen in Figure 7. Overall, the performance metrics indicate successful performance of the model.

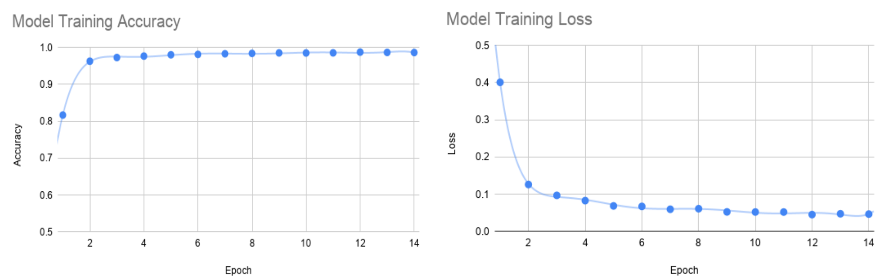


Figure 6. Accuracy and loss of training model with 5-fold CV, using 20 epochs and patience of 2 (Source: Student Researchers).

Table 5. Performance Metrics for 5-Fold Cross-Validation (20 epochs)
The metrics for 5-fold CV with Early Stopping at 20 epochs with a patience of 2 (Source: Student Researchers).

Fold	Loss	Accuracy	Precision	Recall	Specificity	F1 Score	AUC
1	0.0403	98.96%	99.10%	98.09%	99.48%	98.59%	0.9967
2	0.0422	98.96%	99.56%	97.69%	99.73%	98.62%	0.9968
3	0.0389	98.88%	99.59%	97.40%	99.76%	98.49%	0.9975
4	0.0388	98.81%	99.23%	97.55%	99.56%	98.38%	0.9984
5	0.0450	98.83%	99.23%	97.58%	99.56%	98.40%	0.9964
Average	**0.0410**	**98.89%**	**99.34%**	**97.66%**	**99.62%**	**98.50%**	**0.9972**

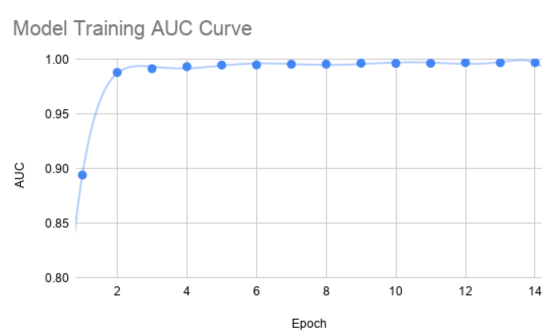


Figure 7. Area under the curve over 14 epochs, Early Stopping at 20 epochs, patience of 2 (Source: Student Researchers).

DISCUSSION

Support for Hypotheses

A model was successfully created to effectively automatically detect and identify atrial fibrillation rhythms from short-term electrocardiogram recordings, as indicated by the excellent performance metrics. The hypothesis that a convolutional neural network could be trained to identify atrial fibrillation present in short-term 30-second 12-lead ECGs from a dataset of patients with various arrhythmias and healthy sinus rhythms was supported by this study. Additionally, this study substantiated the hypothesis that ECGs from patients with various arrhythmias could be used to distinguish atrial fibrillation rhythm arrhythmia from normal sinus rhythm or other arrhythmias.

Potential Limitations

Although the results of the study indicate a success, several limitations exist. One potential oversight was that when hyperparameters were originally being altered, multiple parameters were being changed at once, therefore complicating the determination of how each specific parameter was a factor of the model's performance, as seen in Table 6. Additionally, due

Table 6. Hyperparameter Tuning Experiments
Several experiments were conducted adjusting parameters to determine optimal hyperparameters for the model before 5-fold CV (Source: Student Researchers).

Experiment	1	2	3	4	5	6	7
Filters	100	80	20	100	100	100	**100**
Batch Size	32	24	24	24	32	16	**24**
Kernel Size	18	18	360	9	36	36	**36**
Loss	0.5089	0.5129	0.6566	0.5158	0.4071	0.4852	**0.353**
Accuracy	0.7579	0.7574	0.6307	0.7511	0.8067	0.7796	**0.8452**

to the extended duration of completing 5-fold cross validation training and testing, experiments were run with only one epoch potentially resulting in miscalculations in accuracies and losses for adjusted hyperparameters. Though the final 1D-CNN performed extremely well, other hyperparameters may have provided for an even more accurate model. Additionally, much of the original hyperparameter tuning was conducted using the softmax activation function, but later it was found that a sigmoid function operated more effectively for binary classification. Perhaps, more research could be done to determine the efficacy of the softmax function over the sigmoid function. Moreover, a key limitation in the utilization of neural networks is the lack of explainability. In using convolutional neural networks, the specific features that the network identifies and detects are hidden, which if found, could provide new therapeutic targets or create more certainty in what features contribute to model performance. Another limitation may have been sample size. Numerous previous studies have utilized the MIT-BIH Arrhythmia Databases, but perhaps newer information with a larger more diverse dataset would train the model more effectively. Also, while the 1D-CNN identified only AF or not AF, a more useful neural network would also differentiate the 'not AF' category into either normal sinus rhythm or other arrhythmias (and could go further to identify those 'other arrhythmias').

Comparison to Past Studies

As previously mentioned, several studies have utilized the MIT-BIH Arrhythmia database for various applications with different neural networks. The performance of our 1D-CNN model was compared to that of four previous studies, placing fairly well amongst the others. The performance comparison is not exact, due to the fact that all the CNNs are slightly different. While the studies by Zheng et al. (2020) and Huang et al. (2019) both produced CNNs, they were varied: a 2D-CNN LSTM and a STFT 2D-CNN, respectively. Also, though both identified different arrhythmias, they classified several beat types, among normal sinus rhythm (not atrial fibrillation), but the same MIT-BIH Database was used. On the contrary, research by Erdenebayar et al. (2019) and Hannun et al. (2017) both distinguished atrial fibrillation from

ECGs using 1D-CNNs, and though rhythm classifications were made from 30s segments for both studies, Erdenebayar et al. only classified AF, while Hannun et al.'s CNN addressed 12 rhythm types. Additionally, Erdenebayar et al. used the MIT-BIH Database containing only normal sinus rhythm and atrial fibrillation, and may have achieved such successful performance due to the clear difference. Unlike Erdenebayar et al., not only did Hannun et al. classify 12 types of arrhythmia, but they also collected 64,121 ECG records from 29,163 patients for a diverse and extremely large dataset—possibly the reason for their model's slightly lower performance. Our 1D-CNN model distinguished AF from a combined MIT-BIH dataset containing 30s ECG samples of AF, several other arrhythmias, and normal sinus rhythm. As seen in Table 7, Zheng et al.'s 2D-CNN LSTM technology achieved highest accuracy (99.01%) and Erdenebayar et al.'s 1D-CNN achieved highest recall (98.60%). Our model achieved highest precision, specificity, F1 score, and lowest loss: 99.34%, 99.62%, 98.50%, and 0.0410, respectively, demonstrating its

Table 7. Comparison of Results Across Several Studies
The performance metrics of this study, in comparison to those of four previous studies (Source: Student Researchers, Zheng et al., 2020, Erdenebayar et al., 2019, Hannun et al., 2017, Huang et al., 2019).

Model	Study	Rhythm Types	Accuracy	Precision	Recall	Specificity	F1 Score	Loss
2D-CNN LSTM	Zheng et al.	8	**0.9901**		0.9767	0.9957		
1D-CNN	Erdenebayar	2	0.9870		**0.9860**	0.9870		
1D-CNN	Hannun et al.	12		0.8090	0.8270		0.8090	
STFT 2D-CNN	Huang et al.	5	0.9900					0.0414
1D-CNN	**This Study**	2	0.9889	**0.9934**	0.9766	**0.9962**	**0.9850**	**0.0410**

strong performance. Those successful metrics, however, may be accredited to the fact that our model distinguished between only two classes (AF and not AF), rather than 8 beat types, 12 different rhythms, or 5 beat types. In addition, some of the metrics from other studies were not available and may have been better than those of our study. Nonetheless, our 1D-CNN measured up fairly well in comparison to previous studies.

Implications and Applications

During this study, it was found that our 1D-CNN could effectively classify atrial fibrillation from short-term ECG recordings containing various arrhythmias and normal sinus rhythms. This finding indicates that deep learning methods could be applied to diagnose cardiovascular disease that manifest themselves in electrocardiograms. As current screening methods are challenging, due to issues with diagnosing momentary arrhythmia from a single electrocardiogram and inconvenient prolonged monitoring methods, this study could present a potential solution that isn't so intrusive, uncomfortable, or inefficient. Neural networks could potentially be implemented into wearables, such as wrist watches, to comfortably and effectively alert individuals of arrhythmias. Additionally, as we reduced preprocessing and still achieved outstanding performance, these findings could work to streamline diagnosis for arrhythmias. As demonstrated by the findings in this study, neural networks have great potential to advance diagnostic methods for cardiovascular diseases.

FUTURE RESEARCH

Further research in utilizing convolutional neural networks for classifying arrhythmias may provide even more effective methods of diagnosis. With our study, in the preprocessing stage, resampling ECGs to 250 Hz rather than 360 Hz may provide a varied result. Additionally, future studies could filter out or mark samples containing AF near the bounds of the sample (towards the beginning or end of the 30 seconds) as it would be unreasonable to identify atrial fibrillation based on such a small portion of the sample. Also, conducting hyperparameter tuning with full 5-fold cross-validation (rather than single epoch experiments) would more accurately provide optimal parameters for the final model, and thus greatly improve the performance. Studies should be done focusing on classifying multiple arrhythmias or identifying developing symptoms of cardiovascular diseases before they manifest. More studies should be conducted not only with automatically classifying AF like our study, but with several different arrhythmia rhythms including sinus rhythm, atrial fibrillation, and ventricular and supraventricular arrhythmias. Additionally, studies should also be conducted with various other metrics besides C-statistics and an AUC curve, which measure only discrimination (Krittanawong et al., 2019). Though much research has already been conducted with ECGs and CNN applications, there is always more to improve upon—ranging from the datasets used to preprocessing methods to the actual model itself, along with the metrics used for evaluation.

CONCLUSION

Cardiovascular diseases, such as arrhythmias like atrial fibrillation are extremely prevalent today, and will continue to become more common. This study hoped to create an effective convolutional network that automatically classifies atrial fibrillation rhythms from short-term ECG recordings. ECG data was taken from the MIT-BIH Arrhythmia and Atrial Fibrillation databases, and underwent preprocessing and segmentation, and was split into training and testing sets conducted with 5-fold cross-validation. The 1D-CNN 8-layer model was evaluated with several performance metrics and was found to have: 98.89% average accuracy, 99.34% average precision, 97.66% average recall, 99.62% average specificity, an F1 score of 98.50%, an AUC of .9972, and loss of .041, indicating its outstanding efficacy. The findings of this study, in tandem with previous research on the use of deep learning, could potentially make diagnostic methods of arrhythmias via ECG more comfortable and efficient.

WORKS CITED

Attia, Z. I., Kapa, S., Lopez-Jimenez, F., McKie, P. M., Ladewig, D. J., Satam, G., … Friedman, P. A. (2019). Screening for cardiac contractile dysfunction using an artificial intelligence–enabled electrocardiogram. *Nature Medicine*, *25*(1), 70–74. https://doi.org/10.1038/s41591-018-0240-2

Attia, Z. I., Noseworthy, P. A., Lopez-Jimenez, F., Asirvatham, S. J., Deshmukh, A. J., Gersh, B. J., … Friedman, P. A. (2019). An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *The Lancet*, *394*(10201), 861–867. https://doi.org/10.1016/s0140-6736(19)31721-0

Erdenebayar, U., Kim, H., Park, J.-U., Kang, D., & Lee, K.-J. (2019). Automatic prediction of atrial fibrillation based on convolutional neural network using a short-term normal electrocardiogram signal. *Journal of Korean Medical Science*, *34*(7). https://doi.org/10.3346/jkms.2019.34.e64

Erickson, B. J., Korfiatis, P., Akkus, Z., & Kline, T. L. (2017). Machine learning for medical imaging. *Radiographics*, *37*(2), 505–515. https://doi.org/10.1148/rg.2017160130

Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P. C., Mark, R., ... & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation [Online]. 101 (23), pp. e215–e220.

Hannun, A. Y., Rajpurkar, P., Haghpanahi, M., Tison, G. H., Bourn, C., Turakhia, M. P., & Ng, A. Y. (2017). Cardiologist-level arrhythmia detection with convolutional neural networks. *CoRR*. Retrieved from https://arxiv.org/abs/1707.01836

Huang, J., Chen, B., Yao, B., & He, W. (2019). ECG arrhythmia classification using STFT-based spectrogram and convolutional neural network. *IEEE Access*, *7*, 92871–92880. https://doi.org/10.1109/access.2019.2928017

Ko, W.-Y., Siontis, K. C., Attia, Z. I., Carter, R. E., Kapa, S., Ommen, S. R., … Noseworthy, P. A. (2020). Detection of hypertrophic cardiomyopathy using a convolutional neural network-enabled electrocardiogram. *Journal of the American College of Cardiology*, *75*(7), 722–733. https://doi.org/10.1016/j.jacc.2019.12.030

Krittanawong, C., Johnson, K. W., Rosenson, R. S., Wang, Z., Aydar, M., Baber, U., … Narayan, S. M. (2019). Deep learning for cardiovascular medicine: a practical primer. *European Heart Journal*, *40*(25), 2058–2073. https://doi.org/10.1093/eurheartj/ehz056

Lainscak, M., Dagres, N., Filippatos, G. S., Anker, S. D., & Kremastinos, D. T. (2008). Atrial fibrillation in chronic non-cardiac disease: Where do we stand? *International Journal of Cardiology*, *128*(3), 311–315. https://doi.org/10.1016/j.ijcard.2007.12.078

Luz, E. J. da S., Schwartz, W. R., Cámara-Chávez, G., & Menotti, D. (2016). ECG-based heartbeat classification for arrhythmia detection: A survey. *Computer Methods and Programs in Biomedicine*, *127*, 144–164. https://doi.org/10.1016/j.cmpb.2015.12.008

Moody GB, Mark RG. The impact of the MIT-BIH Arrhythmia Database. IEEE Eng in Med and Biol 20(3):45-50 (May-June 2001). (PMID: 11446209)

Ritchie, H., & Roser, M. (2018). Causes of death. Retrieved from Our World in Data website: https://ourworldindata.org/causes-of-death

Ullah, A., Anwar, S. M., Bilal, M., & Mehmood, R. M. (2020). Classification of arrhythmia by using deep learning with 2-D ECG spectral image representation. *Remote Sensing*, *12*(10), 1685. https://doi.org/10.3390/rs12101685

World Health Organization: WHO. (2019, June 11). Cardiovascular diseases. Retrieved from Who.int website: https://www.who.int/health-topics/cardiovascular-diseases/#tab=tab_1

Xu, X., & Liu, H. (2020). ECG heartbeat classification using convolutional neural networks. *IEEE Access*, *8*, 8614–8619. https://doi.org/10.1109/access.2020.2964749

Zheng, Z., Chen, Z., Hu, F., Zhu, J., Tang, Q., & Liang, Y. (2020). An automatic diagnosis of arrhythmias using a combination of CNN and LSTM technology. *Electronics*, *9*(1), 121. https://doi.org/10.3390/electronics9010121

Zhu, F., Ye, F., Fu, Y., Liu, Q., & Shen, B. (2019). Electrocardiogram generation with a bidirectional LSTM-CNN generative adversarial network. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-42516-z