

Article

Classification of Atrial Fibrillation and Congestive Heart Failure Using Convolutional Neural Network with Electrocardiogram

Yunendah Nur Fu'adah ^{1,2} and Ki Moo Lim ^{1,3,*}

¹ Computational Medicine Lab, Kumoh National Institute of Technology, Department of IT Convergence Engineering, Gumi, Gyeongbuk 39177, Korea

² School of Electrical Engineering, Telkom University, Bandung 40257, Indonesia

³ Computational Medicine Lab, Kumoh National Institute of Technology, Department of Medical IT Convergence Engineering, Gumi, Gyeongbuk 39177, Korea

* Correspondence: kmlim@kumoh.ac.kr

Abstract: Atrial fibrillation (AF) and congestive heart failure (CHF) are the most prevalent types of cardiovascular disorders as the leading cause of death due to delayed diagnosis. Early diagnosis of these cardiac conditions is possible by manually analyzing electrocardiogram (ECG) signals. However, manual diagnosis is complex, owing to the various characteristics of ECG signals. An accurate classification system for AF and CHF has the potential to save patient lives. Therefore, this study proposed an ECG signal classification system for AF and CHF using a one-dimensional convolutional neural network (1-D CNN) to provide a robust classification system performance. This study used ECG signal recording of AF, CHF, and NSR, which can be accessed on the Physionet website. A total of 5600 ECG signal segments were obtained from 56 subjects, divided into train sets from 42 subjects (N = 4200 ECG segments), and test sets from 14 subjects (N = 1400). We applied for leave-one-out cross-validation in training to select the best model. The proposed 1-D CNN algorithm successfully classified raw data of ECG signals into normal sinus rhythm (NSR), AF, and CHF by providing the highest classification accuracy of 99.643%, f1-score, recall, and precision of 0.996, respectively, with an AUC score of 0.999. The results showed that the proposed method extracted the ECG signal information directly without needing several preprocessing steps and feature extraction methods that potentially reduce the information contained in the ECG signals. Furthermore, the proposed method outperformed previous studies in classifying AF, CHF, and NSR. Therefore, this approach can be considered as an adjunct for medical personnel to diagnose AF, CHF, and NSR.

Keywords: atrial fibrillation; congestive heart failure; normal sinus rhythm; convolutional neural network



Citation: Fu'adah, Y.N.; Lim, K.M. Classification of Atrial Fibrillation and Congestive Heart Failure Using Convolutional Neural Network with Electrocardiogram. *Electronics* **2022**, *11*, 2456. <https://doi.org/10.3390/electronics11152456>

Academic Editor: Gemma Piella

Received: 6 July 2022

Accepted: 4 August 2022

Published: 7 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Atrial fibrillation (AF) and congestive heart failure (CHF) are the most common cardiac diseases and are potentially life-threatening [1]. In 2017, AF affected 37.574 million people worldwide, and its frequency has risen by 33% in the last 20 years [2]. AF cases are expected to increase by more than 60% in 2050 [2]. Similarly, the global prevalence of heart failure in 2017 was 64.34 million people [3]. The lifetime risk of developing CHF is over 20% after the age of 40 years [3]. As cardiovascular disorders affect millions of people and potentially lead to death, AF and CHF have become the main-concern public health issues all over the world. Undiagnosed AF and CHF can endanger life [3]; therefore, prompt diagnosis is a very critical issue.

Electrocardiogram (ECG) is a non-invasive method that records the heart's electrical activity, commonly measured and analyzed by researchers [4]. A single-lead ECG as basic heart monitoring is widely used in a pre-clinical setting to detect abnormalities in the heart,

such as cardiac arrhythmias or heart failure [5,6]. Atrial fibrillation is the most common sustained cardiac arrhythmia and is characterized by atrial excitation at a high frequency, resulting in desynchronized atrial contraction and erratic ventricular excitement [7,8]. In an ECG complex, the P wave measures electrical activity in the atrium. A flat P wave is associated with an increased rate of AF due to atrial muscle loss and prolonged conduction time [9]. Meanwhile, CHF is a chronic cardiac disease due to a malfunction of the left ventricle (LV), the main contractile chamber that pumps blood throughout the body. The LV ejection fraction (EF), defined as the ratio of LV stroke and end-diastolic volumes, is used to measure the LV's systolic contractile performance, with a normal LVEF of 50% or higher [10,11]. In terms of ECG signal morphology, the value of the QRS complex can be used to assess the left ventricular ejection fraction, as proposed by J Askenazi et al. [12]. Moreover, in patients with heart failure, QRS prolongation (≥ 120 ms) is a strong predictor of LV systolic dysfunction [13].

ECG signal is a vital bio-signal that cardiologists widely use to diagnose AF and CHF. However, analyzing ECG signals needs a medical expert; is time-consuming, and typically involves subjectivity due to differing interpretations. Therefore, several studies have proposed algorithms to classify AF and CHF based on machine-learning and deep-learning approaches, which can reduce the subjectivity and provide the high accuracy of computer-aided diagnosis as assisting tools for a cardiologist in diagnosing AF and CHF conditions [10].

Rizal et al. analyzed ECG signals based on the Hjorth descriptor method parameters such as activity, mobility, and complexity [14,15]. A study conducted in 2015 using various classifier algorithms such as k-mean clustering, k-nearest-neighbor, and multi-layer perceptron reported accuracy of 88.67%, 99.3%, and 99.3% respectively, to classify AF, CHF, and NSR. The same researchers reported an accuracy of 94% in a study conducted in 2017 using the higher-order complexity and k-nearest-neighbor classifier [15]. Furthermore, Yingthawornsuk et al. used the primary datasets of AF, CHF, and NSR, consisting of 90 recordings, and their findings indicated that the Hjorth descriptor is capable of class separation among cardiac arrhythmia types, reporting accuracy rates of 84.89%, 88.22%, and 76% using least-squares, maximum likelihood, and support vector machine, respectively [16].

The aforementioned studies [14–16], extracted the features of ECG signals using the Hjorth descriptor method. However, owing to noise interference, some features did not perfectly represent the characteristics of ECG signals. Therefore, in our previous studies [17], we proposed discrete wavelet transform (DWT) with five-level decomposition in the preprocessing step, which decomposes the ECG signal into six sub-bands. The Hjorth descriptor method and entropy-based features (Shannon entropy, sample entropy, permutation entropy, dispersion entropy, bubble entropy, and slope entropy) were combined to extract the information from each sub-band of the ECG signal in more detail. The extensive simulation was conducted using several machine learning algorithms, including K-NN, SVM, ANN, and random forest, to classify AF, CHF, and NSR conditions based on the extracted features as input to the machine-learning algorithms. We have already obtained 100% accuracy in our previous studies. However, our proposed method in the previous studies needs preprocessing steps and feature extraction methods that manually extract the statistical information from ECG signals that possibly still did not represent the complete information in ECG signals.

Recently, convolutional neural networks (CNN) have been identified as potential approaches for ECG classification. Kamaleswaran et al. proposed 13-convolutional layers as a robust and rapid method of AF detection using single-lead ECG with variable lengths of 9–61 s. However, they identified that 2.5 s of ECG segments are enough to obtain well accuracy performance. Their study reported an average F1 score of 0.83 in classifying normal, AF, and other rhythms [18]. Ping et al. used 5 s, 10 s, and 20 s segments of ECG signals as input to the eight layers of CNN in the feature extraction layer and one LSTM layer with a fully connected layer in the classification layer. Their study reported the

highest f1-scores value of 89.55%, using the 10 s segment of ECG signals as input to CNN and LSTM to classify AF and normal conditions [19]. Sidrah et al. proposed the ten layers of CNN in the feature extraction layer and two LSTM layers with a fully connected layer and a drop layer in the classification layers. Their study reported a classification accuracy of 86.5% using a CNN and long short-term memory to classify AF and normal conditions [20]. Similarly, Georgios et al. used a short segment of ECG signals with length data of 187 samples as input to the nine layers of CNN in the feature extraction layer and one LSTM layer with a fully connected layer and a drop layer in the classification layer. Their study reported a sensitivity of 97.87% and specificity of 99.29% using a CNN and long short-term memory to classify AF and normal conditions [21].

Nurmaini et al. reported a classification accuracy of 99.17% using 9 s segment of ECG signals as input to the ten layers of CNN with two fully connected layers to classify three conditions, including normal, AF, and non-AF conditions [22]. Several studies have developed automated systems for detecting AF. Meanwhile, Wang et al. used 500, 1000, and 2000 samples of R–R Interval length as input to the four convolutional layers and one max pooling layer in the feature extraction layers, which used global average pooling layer to extract the features. Their study reported a classification accuracy of 99.85%, 99.41%, and 99.17% for 500, 1000, and 2000 samples, respectively, using a long short-term memory-convolutional neural network to classify CHF and normal conditions [23]. Ning et al. reported a classification accuracy of 98.5% using 1 min segment of ECG signals and 99.3% using 5 min segment of ECG signals as input to a recursive neural network to classify CHF and normal conditions [24]. Meanwhile, Porumb et al. used three convolutional layers, then max-pooling for each layer, and used two fully connected layers in the classification layer. Their study reported an accuracy of 100% in classifying CHF and normal condition [25].

The aforementioned studies [19–25] developed the automatic system detection for AF or CHF conditions separately. Furthermore, Padmavati et al. proposed a classification system to identify four classes of cardiac disease, including atrial fibrillation (AF), myocardial infarction (MI), congestive heart failure (CHF), and normal conditions [26]. Their study reported a classification accuracy of 80.1% to classify CHF and normal, 85.9% accuracy to classify MI and CHF, and 65.4% accuracy to classify MI and normal conditions. However, the accuracy performance of their proposed model decreased by up to 63.5% and 31.2% when extended to classify three classes (CHF, MI, and normal) and four classes (AF, CHF, MI, and normal), respectively.

Performance results of certain previous studies on the classification of ECG signals revealed that conventional machine-learning methods were sensitive to noise; therefore, data cleaning was required. In contrast to the conventional approaches that require separate dataset preprocessing, feature extraction, and classification processes, CNN can directly extract features from raw input of ECG signal. The feature extraction layer in CNN model consists of convolutional layers and commonly used discontinuous activation function such as Rel-U (Rectified Linear Unit) activation function [27]. Discontinuous activation functions in neural networks extract essential features that frequently appear in practice [28]. As a result, in sufficient training samples, the features extracted by a CNN model would be more detailed than those extracted manually.

According to the advantages of CNN, some limitations that previously existed in conventional machine learning can be overcome with CNN. A one-dimensional (1-D) CNN was proposed to design an optimal ECG classification system that can improve the performance accuracy of the previous methods. Based on the previous related studies, there are several essential parameters to classify AF and CHF based on a deep-learning approach: the amount of length data as input to the CNN model and the architecture of the CNN model. Therefore, this study aimed to investigate the performance of the 1-D CNN using several length data as input to the CNN model to classify raw ECG signals of AF, CHF, and NSR.

2. Materials and Methods

This study proposed an optimal ECG signal classification system using a 1-D CNN that directly processes the raw ECG signal data and classifies them into AF, CHF, and NSR conditions. The configuration of the proposed 1-D CNN model is shown in Figure 1.

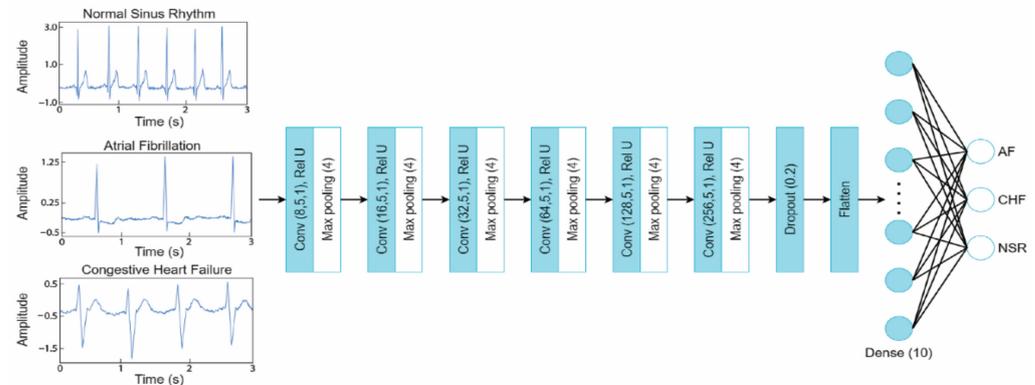


Figure 1. Proposed model of the one-dimensional convolutional neural network for NSR, AF, and CHF classification.

2.1. Dataset

ECG signal data were collected from the Physionet database, consisting of three cardiac classes, namely, AF, CHF, and NSR. AF data were provided by MIT-BIH Atrial Fibrillation database [29], CHF data were provided by BIDMC Congestive Heart Failure database [30], and NSR data were provided by MIT-BIH Normal Sinus Rhythm database [31]. The AF database consisted of 23 recordings of long-term ECG signals (mostly paroxysmal) with 10 hours (h) in each recording. Furthermore, the CHF dataset included long-term ECG recordings (20 h in each recording) from 15 subjects (11 men, aged 22 to 71, and 4 women, aged 54 to 63) with severe congestive heart failure (NYHA class 3–4). Meanwhile, for NSR dataset includes 18 long-term ECG recordings of subjects (5 men, aged 26 to 45, and 13 women, aged 20 to 50). This study divided the dataset into the training set and testing set based on the patient identity. Therefore, both training and testing sets were obtained from different subjects. Furthermore, we used 100 ECG segments for each subject. The training set consists of 18 subjects of AF conditions ($N = 1800$ ECG segments), 11 subjects of CHF conditions ($N = 1100$ ECG segments), and 13 subjects of NSR conditions ($N = 1300$ ECG segments). Meanwhile, the testing set consists of 5 subjects of AF conditions ($N = 500$ ECG segments), 4 subjects of CHF conditions ($N = 400$ ECG segments), and 5 subjects of NSR conditions ($N = 500$ ECG segments).

2.2. Feature Extraction Using 1-D CNN

In this study, we designed several architectures with varying depths of convolution layers in order to know the influence of the number of convolutional layers on classification performance. Architecture 1 consisted of three convolutional layers in the feature extraction layers. Convolutional layers one through three had 8, 16, and 32 filters, respectively. Architecture 2 consisted of four convolutional layers with 8, 16, 32, and 64 filters. Architecture 3 consisted of five convolutional layers with 8, 16, 32, 64, and 128 filters. Architecture 4 consisted of six convolutional layers with 8, 16, 32, 64, 128, and 256 filters. We begin with a small number of filters to identify low features that combine to produce more complex shapes and increased the number of filters to help with class separation. The kernel size of five with stride one in each convolutional layer was applied for all models to assemble essential information on ECG signals in more detail.

Furthermore, rectified linear unit activation function was applied to each convolution layer. Similarly, we applied max pooling to each layer following the convolutional layer. Max-pooling layers are used to reduce the features map's dimensions and select the maximum element of the features map that is covered by the filter. Subsequently, the feature

maps were extracted from the convolutional and max-pooling layers containing the most prominent features as inputs for the classification layers.

2.3. Classification Using 1-D CNN

The classification layer of the 1-D CNN is a fully connected layer responsible for classifying the data. There is a flattening process before creating the fully connected layers, consisting of one hidden layer with ten nodes (Figure 1). Finally, the SoftMax activation function was applied to classify the signals into AF, CHF, and NSR.

This study used leave-one-out cross-validation to evaluate generalization ability of deep neural networks in predicting the unknown dataset. In leave-one-out cross-validation, the number of folds equals the number of instances in the dataset. Since the training set consisted of 42 subjects, there were 42-folds. The learning algorithm was applied once for each instance, using 41 subjects ($N = 4100$ ECG segments) as a training set and one subject ($N = 100$ ECG segments) as a validation set. Furthermore, 42 models were evaluated using a test dataset consisting of 14 subjects ($N = 1400$ ECG segments).

Furthermore, we applied regularization techniques and an optimization algorithm to improve model generalization ability. As for regularization techniques, we applied a dropout layer and callback. We used a dropout layer by deactivating neurons with probability 0.2 applied at the last feature extraction layer in the forward propagation and weight update step. This more superficial neural network resulted in less complexity to reduce overfitting. Moreover, we used a callback to monitor specific metrics, including validation loss and accuracy. The model checkpoint was implemented to save the network weights when the classification performance on the validation dataset improves. Therefore, the model in a particular epoch in each fold was automatically saved. After that, we tried the model from each fold in leave-one-out cross-validation to select the best model that provided the highest accuracy for test data.

Meanwhile, to optimize the 1-D CNN algorithm, we tried several optimizer algorithms (including Adam, Nadam, SGD, and RMSprop) and selected the optimal learning rate (0.1 to 0.0001). The optimizer algorithms, including the learning rate value, were applied while training the network. The optimizer algorithm minimizes the error related to accuracy performance.

2.4. System Performance

A confusion matrix was used to obtain the accuracy, precision, recall, and f1-score when evaluating the system performance. The following equations were used to calculate parameters to measure the system's effectiveness in diagnosing AF, CHF, and NSR conditions.

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

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

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

$$F1 \text{ score} = 2 \cdot \frac{Recall \cdot Precision}{Recall + Precision} \quad (4)$$

In Equations (1)–(3), a true positive (TP) is a result in which the model predicts the positive class correctly. A true negative (TN), on the other hand, is a result in which the model correctly predicts the negative class. A false positive (FP) is a result in which data is negative but incorrectly classified as positive. Meanwhile, a false negative (FN) is a result in which data is positive but incorrectly classified as negative [32].

3. Results

A total of 1400 test data that consisted of 500 data of AF, 400 data of CHF, and 500 data of NSR were used to evaluate the 42 models generated from the training process using leave-one-out cross-validation. The highest performance result of the selected model (out of 42 models) on test data with 3 s, 5 s, and 8 s segments of ECG signals as input to architectures 1, 2, 3, and 4 is represented in Table 1. The 8 s segment of ECG signals provided the highest classification accuracy of 99.643%, f1-score, recall, and precision of 0.996, respectively, with an AUC score of 0.999. The 5 s segment of ECG signals obtained a highest classification accuracy of 99.286%, f1-score, recall, and precision of 0.993, respectively, with an AUC score of 0.999. Meanwhile, the 3 s segment of ECG signals obtained a highest classification accuracy of 97.428%, f1-score, recall, and precision of 0.974, respectively, with an AUC score of 0.996.

Table 1. The highest classification performance of the proposed one-dimensional convolutional neural network.

Length of Duration	Selected 1-D CNN Model	Test Accuracy	F1-Score	Recall	Precision	AUC Score
3 s	Architecture 1	97.428	0.974	0.974	0.974	0.996
	Architecture 2	97.286	0.973	0.973	0.973	0.996
	Architecture 3	97.286	0.973	0.973	0.973	0.996
	Architecture 4	96.786	0.968	0.968	0.968	0.996
5 s	Architecture 1	97.857	0.979	0.979	0.979	0.997
	Architecture 2	99.286	0.993	0.993	0.993	0.999
	Architecture 3	99.214	0.992	0.992	0.992	0.999
	Architecture 4	99.143	0.991	0.991	0.991	0.999
8 s	Architecture 1	92.781	0.928	0.928	0.928	0.983
	Architecture 2	98.357	0.983	0.983	0.983	0.998
	Architecture 3	99.143	0.991	0.991	0.991	0.999
	Architecture 4	99.643	0.996	0.996	0.996	0.999

As shown in Table 1, the CNN model's architecture and the ECG segment's duration influence the classification performance. In order to determine the best architecture and the duration of the ECG signal segment, we presented a classification accuracy with a 95% confidence interval and used the ANOVA test for statistical analysis with statistical significance at $p < 0.05$. The classification accuracy of train and test data with a 95% confidence interval was obtained from the evaluation result of 42 models generated during training using leave-one-out cross-validation. As shown in Table 2, using the 8 s segment of ECG signal, architecture 4 obtained a classification train accuracy with a 95% confidence interval of 96.19–99.10% and classification test accuracy with a 95% confidence interval of 96.37–98.52%. The classification of train and test accuracy provided a similar range value to the classification accuracy performance. Therefore, the proposed model is not overfitting, which means it not only successfully classified the training dataset but also has a good generalization ability in classifying the unseen dataset. In contrast with architecture 4, the other architectures are overfitting by providing quite a significant difference between train and test accuracy. Furthermore, based on the ANOVA test, architecture 4 obtained the most statistical significance with a p -value of 9.62×10^{-7} ($p < 0.05$) compared with the other architectures. Therefore, we selected architecture 4 as the best architecture to classify AF, CHF, and NSR conditions.

Table 2. Classification accuracy performance with 95% confidence interval and *p*-value for each architecture.

Architecture	Duration	Train Accuracy	Test Accuracy	<i>p</i> -Value
Architecture 1	3 s	93.39–98.83%	91.13–93.79%	5.29×10^{-3}
	5 s	97.48–99.91%	94.21–95.44%	
	8 s	97.11–99.26%	91.66–93.90%	
Architecture 2	3 s	97.41–99.95%	94.51–95.79%	5.04×10^{-1}
	5 s	96.86–99.99%	93.75–96.48%	
	8 s	97.21–99.24%	94.92–96.88%	
Architecture 3	3 s	96.83–99.46%	94.20–95.32%	5.62×10^{-3}
	5 s	94.10–98.56%	93.02–96.06%	
	8 s	95.32–98.94%	96.05–97.51%	
Architecture 4	3 s	96.09–98.96%	92.95–94.49%	9.62×10^{-7}
	5 s	95.73–99.57%	95.03–96.80%	
	8 s	96.19–99.10%	96.37–98.52%	

Figure 2 shows the confusion matrix of the proposed CNN architecture for the 3 s, 5 s, and 8 s ECG signal segment. As shown in Figure 2a,b, using 3 s and 5 s ECG segments, several data of AF, CHF, and NSR conditions are not successfully classified according to their class. Meanwhile, using the 8 s ECG segment as input to the CNN model, the test data of the AF, CHF, and NSR are mainly classified to their class accordingly (Figure 2c).

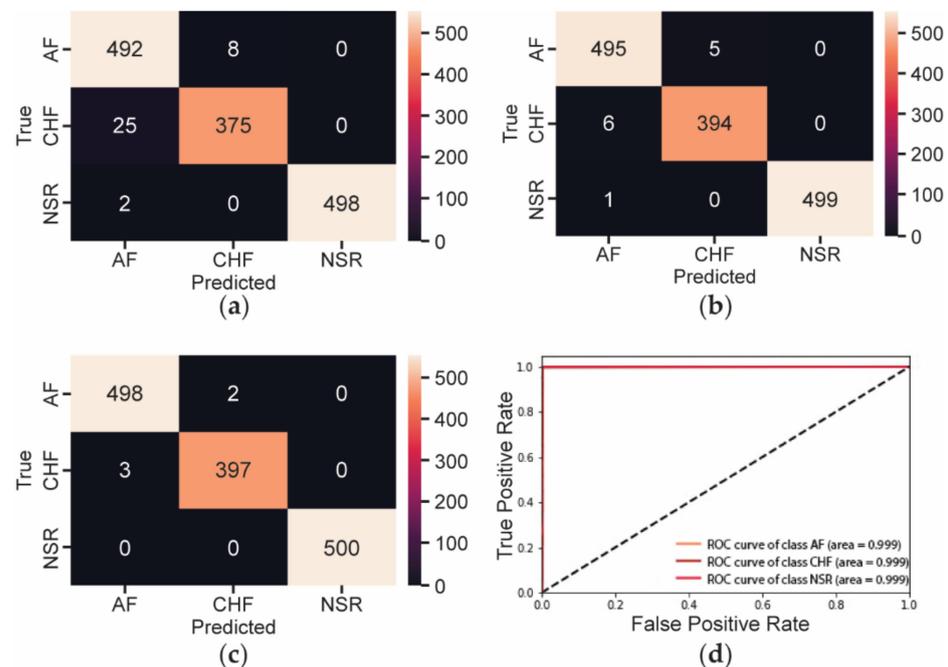


Figure 2. The proposed model performance. (a) The confusion matrix of test datasets for ECG signals with duration 3 s using architecture 4; (b) the confusion matrix of test datasets for ECG signals with duration 5 s using architecture 4; (c) the confusion matrix of test datasets for ECG signals with duration 8 s using architecture 4; and (d) the ROC curve of the CNN model for ECG signals with duration 8 s using architecture 4.

The receiver operating characteristic (ROC) curve showed the true positive rate (TPR) in y-axis against the false positive rate (FPR) in x-axis as shown in Figure 2d. As a summary of ROC curve, the area under curve (AUC) score showed the ability of the classifier to distinguish between classes. The greater the AUC score, the better the model’s ability in distinguishing between AF, CHF, and NSR conditions. As shown in Figure 2d, the AUC

score was 0.999. Therefore, we can conclude that the proposed model can generalize the test datasets well.

As shown in Table 3, the proposed CNN model successfully classified AF, CHF, and NSR from raw data of ECG signals, in contrast with previous studies, which used hand-crafted features extraction methods that potentially reduced the information contained in the ECG signal and required extensive simulation to obtain the optimal performance [14–17]. The classification accuracy performance from the previous studies in classifying AF and NSR are around 84.89–99.85% [19–22]. Furthermore, in classifying CHF and NSR, the aforementioned studies exhibited high performance classification accuracy, around 99.3–100%. However, the related previous studies that used CNN were binary classification studies that classified between AF and NSR or CHF and NSR. Therefore, the proposed method in this study successfully expanded the classification of heart disease based on ECG signals with high classification accuracy performance.

Table 3. Performance comparison with previous studies.

References	No of Classes	Feature Extraction	Classifier Algorithm	Classification Performance
Rizal et al. (2015) [14]	3 classes (AF, CHF, NSR)	2–3 s ECG signals segment, Hjorth descriptor	K-mean clustering, K-NN, and MLP	Accuracy of 88.67%, 99.3%, and 99.3%
Hadiyoso et al. (2017) [15]	3 classes (AF, CHF, NSR)	2–3 s ECG signals segment, higher order complexity	K-NN	Accuracy of 94%
Yingthawornsuk et al. (2018) [16]	3 classes (AF, CHF, NSR)	2–3 s ECG signals segment, Hjorth descriptor	LS, ML, SVM	Accuracy of 84.89%, 88.22%, and 76.94%
Fuadah et al. (2022) [17]	3 classes (AF, CHF, NSR)	2–3 s ECG signals segment, Hjorth descriptor, Entropy based features	K-NN, SVM, RF, and ANN	Accuracy of 100%
Kamaleswaran et al. (2018) [18]	3 classes (normal, AF, and other rhythms)	9–61 s ECG signals segment, the 13 layers of CNN	No fully connected layer	F1-score of 0.83
Ping et al. (2020) [19]	2 classes (AF and normal)	5 s, 10 s, and 20 s ECG signals segment, 8 layers of CNN	1 LSTM layer and 1 fully connected layer	F1-score of 84.89% (5 s), 89.55% (10 s), and 85.64% (20 s)
Sidrah et al. (2020) [20]	2 classes (AF and normal)	10 s ECG signals segment, 10 layers of CNN	2 LSTM layers, 1 fully connected layer and 1 dropout layer	Accuracy of 86.5%
Georgios et al. (2020) [21]	2 classes (AF and normal)	1.3 s ECG signals segment (187 sample length data), 9 layers of CNN	1 LSTM layer, 1 fully connected layer, and 1 dropout layer	Sensitivity of 97.87% and specificity of 99.29%
Nurmaini et al. (2020) [22]	3 classes (AF, non-AF, normal)	9 s ECG signals segment, 10 layers of CNN	2 fully connected layers	Accuracy of 99.17%
Wang et al. (2020) [23]	2 classes (CHF and normal)	500, 1000, 2000 sample lengths, 4 convolutional layers, and 1 max-pooling layer	Global average pooling layer	Accuracy of 99.85%, 99.41%, and 99.17%, respectively
Ning et al. [24]	2 classes (CHF and normal)	1-min and 5-min segment of ECG signals	CNN and RNN	Accuracy of 98.5% (1-min) and 99.3% (5-min), respectively
Porumb et al. (2020) [25]	2 classes (CHF and normal)	5-min segment of ECG signals, 3 convolutional layers	2 fully connected layers	Accuracy of 100%
Padmavathi et al. (2020) [26]	2 classes (CHF, normal), 2 classes (CHF, MI), 2 classes (MI, normal), 3 classes (CHF, MI, normal), and 4 classes (AF, CHF, MI, normal)	EMD, 3 convolutional layers, and 3 max-pooling layers of CNN	1 fully connected layer	Accuracy of 80.1%, 85.9%, 65.4%, 63.5%, and 31.2%, respectively
Our Method	3 classes (AF, CHF, NSR)	3 s, 5 s, 8 s segment of ECG signals, 6 convolutional layers, 6 max-pooling layers, and a drop layer	1 fully connected layer	Accuracy of 96.786%, 99.143%, and 99.643% respectively

4. Discussion

In this study, we proposed a new configuration of the 1-D CNN to classify AF, CHF, and NSR conditions. While AF, CHF, and NSR classification using machine learning in the

previous studies relied on handcrafted features to make classifiers capable of classifying AF, CHF, and NSR conditions, we significantly advanced the method by using raw ECG signals as input for 1-D CNN. As a result, we improved the classification accuracy performance of the aforementioned studies that used the same datasets. We claim this advantage due to the ability of 1-D CNN to extract and learn the pattern of ECG signals rather than relying on specific features that might not completely represent the characteristics of ECG signals.

While determining the configuration of the 1-D CNN model, the number of filters and depth of the model must be considered. These parameters influence the feature maps as well as the model complexity. If the model is too simple, it will not be able to extract the unique features. On the other hand, if the model is too deep, it will increase the model complexity and slow the training process. Therefore, we evaluated four architectures with different depths of convolutional layers and different ECG segment duration as input. Architectures 1–4 consist of 3, 4, 5, and 6 convolutional layers, respectively. We selected the best architecture by evaluating the system's performance and applied statistical analysis using the ANOVA test with a statistical significance of $p < 0.05$. As shown in Table 2, architecture 4 provided the highest classification test accuracy and obtained a p -value of 9.62×10^{-7} ($p < 0.05$), which was the most statistically significant of the architectures in classifying AF, CHF, and NSR. According to the results, we can conclude that the number of convolutional layers significantly impacts classification performance related to the feature maps generated from feature extraction layers. The optimal number of convolutional layers has been selected through several simulations and evaluation of the system's classification performance.

As the most optimal architecture that successfully outperformed the other architecture, architecture 4 consists of six convolutional layers with 8, 16, 32, 64, 128, and 256 filters for layers one through six, respectively. A filter size of five was applied in each convolutional layer, followed by a max-pooling layer after each convolutional layer. We used a small filter size to assemble information on ECG signals in more detail. We started with a low number of filters to identify low features that combined to generate more complex shapes and increased the number of filters that aid in class separation. Furthermore, we set a dropout of 0.2 in the feature extraction layer to decrease the complexity of the network and reduce overfitting.

Moreover, in the classification layer, the output of the convolutional layer passes through one hidden layer toward the output layer with a fully connected layer. A fully connected layer will learn all features combined from the previous layer, leading to correct classification. Furthermore, several hyper parameters including optimization algorithms and learning rate were carefully selected after conducting extensive simulations to avoid over-fitting and increase robustness with good generalization capability.

Even though the clinical approaches for diagnosing AF and CHF are accurate, they still have some limitations, such as being time-consuming and differing ECG signal interpretations among clinicians. As a result, computer-aided diagnosis tools based on machine learning and deep learning are currently being considered for cardiac disease diagnosis. This study demonstrated several experiments to classify AF, CHF, and NSR conditions using a deep-learning concept on a single lead ECG signal with a short segment.

As a gold standard for diagnosing AF, CHF, and NSR conditions, ECG signal morphology contains essential cardiac condition information. Therefore, the ECG signal is used as input into the CNN model to classify AF, CHF, and NSR conditions. In ECG signal morphology, AF is characterized by a flat P wave associated with an increased rate due to atrial muscle loss and prolonged conduction time. Meanwhile, QRS prolongation (≥ 120 ms) strongly predicts diagnosing patients with CHF. Based on the uniqueness of the ECG signal morphology of AF and CHF, the proposed CNN architecture will learn the pattern and extract the essential information from each condition.

Furthermore, we identified and evaluated the generated model from leave-one cross-validation during training to classify the unseen data from different subjects to ensure the generalization ability of the proposed CNN model. According to the best classification

accuracy performance using the selected model out of 42 models, we obtained the highest test accuracy, 99.643%, with an AUC score of 0.999. Moreover, after evaluation using 42 models, this study yielded a classification test accuracy with a 95% confidence interval of 96.37–98.52%.

As a result, we believe that a deep-learning system might be used in computer-aided diagnosis to analyze ECG signals to classify AF, CHF, and NSR successfully. Additionally, this approach has the potential to be utilized as an assisting tool to help medical professionals perform more accurate diagnosis of AF and CHF conditions. However, to confirm the clinical feasibility of our proposed approach, we must conduct further investigation using a larger number of datasets.

Author Contributions: This manuscript is the intellectual product of the entire team. Y.N.F. wrote the convolutional neural network source code and manuscript, performed data analysis, and interpreted the results. K.M.L. designed the study, reviewed, and revised the entire manuscript based on the results. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially supported by the Ministry of Food and Drug Safety (22213MFDS3922), the NRF (National Research Foundation of Korea) under the Basic Science Research Program (2022R1A2C2006326), and the MSIT (Ministry of Science and ICT), Korea, under the Grand Information Technology Research Center support program (IITP-2022-2020-0-01612) supervised by the IITP (Institute for Information & communications Technology Planning & Evaluation).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets presented in this study can be found in online repositories. The names of the repositories can be found at <https://physionet.org/content/afdb/1.0.0/>; <https://physionet.org/content/chfdb/1.0.0/>; <https://physionet.org/content/nsrdb/1.0.0/>.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mak, M.W.; Cheung, C.C. Towards End-to-End ECG Classification with Raw Signal Extraction and Deep Neural Networks. *IEEE J. Biomed. Health Informatics* **2019**, *23*, 1574–1584.
2. Lippi, G.; Sanchis-Gomar, F.; Cervellin, G. Global epidemiology of atrial fibrillation: An increasing epidemic and public health challenge. *Int. J. Stroke* **2021**, *16*, 217–221. [[CrossRef](#)] [[PubMed](#)]
3. Lippi, G.; Sanchis-Gomar, F. Global epidemiology and future trends of heart failure. *AME Med. J.* **2020**, *5*, 6. [[CrossRef](#)]
4. Taye, G.T.; Hwang, H.-J.; Lim, K.M. Application of convolutional neural network for predicting the occurrence of ventricular tachyarrhythmia using heart rate variability features. *Sci. Rep.* **2020**, *10*, 6769. [[CrossRef](#)] [[PubMed](#)]
5. Taye, G.T.; Shim, E.B.; Hwang, H.-J.; Lim, K.M. Machine Learning Approach to Predict Ventricular Fibrillation Based on QRS Complex Shape. *Front. Physiol.* **2019**, *10*, 1193. [[CrossRef](#)]
6. Witvliet, M.P.; Karregat, E.P.M.; Himmelreich, J.C.L.; de Jong, J.S.S.G.; Lucassen, W.A.M.; Harskamp, R.E. Usefulness, pitfalls and interpretation of handheld single-lead electrocardiograms. *J. Electrocardiol.* **2021**, *66*, 33–37. [[CrossRef](#)]
7. Chen, L.Y.; Soliman, E.Z. P Wave Indices—Advancing Our Understanding of Atrial Fibrillation-Related Cardiovascular Outcomes. *Front. Cardiovasc. Med.* **2019**, *6*, 53. [[CrossRef](#)]
8. Staerk, L.; Sherer, J.A.; Ko, D.; Benjamin, E.J.; Helm, R.H. Atrial Fibrillation: Epidemiology, Pathophysiology, Clinical Outcomes. *Circ. Res.* **2017**, *120*, 1501–1517. [[CrossRef](#)]
9. Rasmussen, M.U.; Kumarathurai, P.; Fabricius-Bjerre, A.; Larsen, B.S.; Domínguez, H.; Davidsen, U.; Gerds, T.A.; Kanters, J.K.; Sajadieh, A. P-wave indices as predictors of atrial fibrillation. *Ann. Noninvasive Electrocardiol.* **2020**, *25*, e12751. [[CrossRef](#)]
10. Jahmunah, V.; Oh, S.L.; Wei, J.K.E.; Ciaccio, E.J.; Chua, K.; San, T.R.; Acharya, R.U. Computer-aided diagnosis of congestive heart failure using ECG signals—A review. *Phys. Med.* **2019**, *62*, 95–104. [[CrossRef](#)]
11. Heidenreich, P.A.; Bozkurt, B.; Aguilar, D.; Allen, L.A.; Byun, J.J.; Colvin, M.M.; Deswal, A.; Drazner, M.H.; Dunlay, S.M.; Evers, L.R.; et al. 2022 AHA/ACC/HFSA Guideline for the Management of Heart Failure: Executive Summary. *J. Am. Coll. Cardiol.* **2022**, *79*, 1757–1780. [[CrossRef](#)] [[PubMed](#)]
12. Askenazi, J.; Parisi, A.F.; Cohn, P.F.; Freedman, W.B.; Braunwald, E. Value of the QRS complex in assessing left ventricular ejection fraction. *Am. J. Cardiol.* **1978**, *41*, 494–499. [[CrossRef](#)]
13. Kashani, A.; Barold, S.S. Significance of QRS complex duration in patients with heart failure. *J. Am. Coll. Cardiol.* **2005**, *46*, 2183–2192. [[CrossRef](#)] [[PubMed](#)]

14. Rizal, A.; Hadiyoso, S. ECG signal classification using Hjorth Descriptor. In Proceedings of the 2015 International Conference on Automation, Cognitive Science, Optics, Micro Electro-Mechanical System and Information Technology (ICACOMIT), Bandung, Indonesia, 29–30 October 2015; pp. 87–90. [[CrossRef](#)]
15. Hadiyoso, S.; Rizal, A. Electrocardiogram Signal Classification Using Higher-Order Complexity of Hjorth Descriptor. *Adv. Sci. Lett.* **2017**, *23*, 3972–3974. [[CrossRef](#)]
16. Yingthawornsuk, T.; Temsang, P. *Cardiac Arrhythmia Classification Using Hjorth Descriptors*; Springer: Cham, Switzerland, 2019; Volume 807. [[CrossRef](#)]
17. Fuadah, Y.N.; Lim, K.M. Optimal Classification of Atrial Fibrillation and Congestive Heart Failure Using Machine Learning. *Front. Physiol.* **2022**, *12*, 761013. [[CrossRef](#)]
18. Kamaleswaran, R.; Mahajan, R.; Akbilgic, O. A robust deep convolutional neural network for the classification of abnormal cardiac rhythm using varying length single lead electrocardiogram. *Inst. Phys. Eng. Med.* **2018**, *39*, 035006.
19. Ping, Y.; Chen, C. Automatic Detection of Atrial Fibrillation Based on CNN-LSTM and Shortcut Connections. *Healthcare* **2020**, *8*, 139. [[CrossRef](#)]
20. Sidrah, L.; Dashtipour, K. Detection of Atrial Fibrillation Using a Machine Learning Approach. *Information* **2020**, *11*, 549. [[CrossRef](#)]
21. Georgios, P.; Haris, K. Automated Atrial Fibrillation Detection using a Hybrid CNN-LSTM Network on Imbalanced ECG Datasets. *Biomed. Signal Process. Control* **2020**, *63*, 102194. [[CrossRef](#)]
22. Nurmaini, S.; Tondas, A.E. Robust detection of atrial fibrillation from short-term electrocardiogram using convolutional neural networks. *Futur. Gener. Comput. Syst.* **2020**, *113*, 304–317. [[CrossRef](#)]
23. Wang, L.; Zhou, W. Deep Ensemble Detection of Congestive Heart Failure Using Short-Term RR Intervals. *IEEE Access* **2019**, *7*, 69559–69574. [[CrossRef](#)]
24. Ning, W.; Li, S. Automatic Detection of Congestive Heart Failure Based on a Hybrid Deep Learning Algorithm in the Internet of Medical Things. *IEEE Internet Things J.* **2020**, *8*, 12550–12558. [[CrossRef](#)]
25. Porumb, M.; Iadanza, E. A convolutional neural network approach to detect congestive heartfailure. *Biomed. Signal Process. Control* **2019**, *55*, 101597. [[CrossRef](#)]
26. Padmavathi, C.; Veenadevi, S.V. Heart disease recognition from ECG signal using deep learning. *Int. J. Adv. Sci. Technol.* **2020**, *29*, 2303–2316.
27. Yuen, B.; Hoang, M.T.; Dong, X.; Lu, T. Universal activation function for machine learning. *Sci. Rep.* **2021**, *11*, 18757. [[CrossRef](#)] [[PubMed](#)]
28. Das, P.; Das, P.; Kundu, A. Delayed Feedback Controller based Finite Time Synchronization of Discontinuous Neural Networks with Mixed Time-Varying Delays. *Neural Process. Lett.* **2019**, *49*, 693–709. [[CrossRef](#)]
29. Moody, G.B.; Mark, R.G. A new method for detecting atrial fibrillation using R-R intervals. *Comput. Cardiol.* **1983**, *10*, 227–230.
30. Baim, D.S.; Colucci, W.S.; Monrad, E.S.; Smith, H.S.; Wright, R.F.; Lanoue, A.; Gauthier, D.F.; Ransil, B.J.; Grossman, W.; Braunwald, E. BIDMC Congestive Heart Failure Database. PhysioNet. 2000. Available online: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6480269/> (accessed on 15 February 2021).
31. Moody, G. MIT-BIH Normal Sinus Rhythm Database. PhysioNet. 1999. Available online: <https://physionet.org/content/nsrdb/1.0.0/> (accessed on 15 February 2021).
32. Novakovic, J.; Veljovi, A.; Ilic, S.; Papic, Z.; Tomovic, M. Evaluation of Classification Models in Machine Learning. *Theory Appl. Math. Comput. Sci.* **2017**, *7*, 39–46.