



# A novel proposed CNN–SVM architecture for ECG scalograms classification

Oznur Ozaltin<sup>1</sup> · Ozgur Yeniay<sup>1</sup>

Accepted: 3 December 2022 / Published online: 15 December 2022

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

## Abstract

Nowadays, the number of sudden deaths due to heart disease is increasing with the coronavirus pandemic. Therefore, automatic classification of electrocardiogram (ECG) signals is crucial for diagnosis and treatment. Thanks to deep learning algorithms, classification can be performed without manual feature extraction. In this study, we propose a novel convolutional neural networks (CNN) architecture to detect ECG types. In addition, the proposed CNN can automatically extract features from images. Here, we classify a real ECG dataset using our proposed CNN which includes 34 layers. While this dataset is one-dimensional signals, these are transformed into images (scalograms) using continuous wavelet transform (CWT). In addition, the proposed CNN is compared to known architectures: AlexNet and SqueezeNet for classifying ECG images, and we find it more effective than others. This study, which not only performed CWT but also implemented short-time Fourier transform, examines the success in recognizing ECG types for the proposed CNN. Besides, different split methods: training and testing, and cross-validation are applied in this study. Eventually, CWT and cross-validation are the best pre-processing and split methods for the proposed CNN, respectively. Although the results are quite good, we benefit from support vector machines (SVM) to obtain the best algorithm and for detecting ECG types. Essentially, the main aim of the study increases classification results. In this way, the proposed CNN is utilized as deep feature extractor and combined with SVM. As a conclusion of this study, we achieve the highest accuracy of 99.21% from the proposed CNN–SVM when using CWT. Therefore, we can express that this framework can be used as an aid to clinicians for ECG-type identification.

**Keywords** Convolutional neural networks (CNN) · Continuous wavelet transform (CWT) · Feature extraction · Scalogram · Support vector machine (SVM)

## 1 Introduction

The qualitative processing and classification of biomedical signals is very important for diagnosis and therapy. Many methods are used to process biomedical signals. Some important methods are discrete Fourier transform (DFT), short-time Fourier transform (STFT), continuous wavelet transform (CWT), and discrete wavelet transform. The Fourier transformation provides a very good frequency range for stationary signals (Haberl et al. 1989). However, the time domain is almost non-existent.

This can lead to serious problems, especially if time-dependent characteristics are to be inferred. However, when signals are transformed with the wavelet transform, both frequency and time domains are distinguishable (Li et al. 1995). In other words, wavelet transform (WT) is a transformation technique that splits signals into different frequency components and processes each component with the time domain of the respective scale. In this study, we focus on electrocardiogram (ECG) signals. The signals resulting from the electrical activity of the heart, the main vital organ in the human body, are called an electrocardiogram (ECG). Sudden deaths from heart disease with coronavirus (COVID-19) are currently on the rise (<https://www.chss.org.uk/media-release/new-nhs-figures-show-dangerous-domino-effect-of-pandemic-on-progress-made-with-strokes-and-heart-disease/>). For this

✉ Oznur Ozaltin  
oznurozaltin@hacettepe.edu.tr

<sup>1</sup> Department of Statistics, Institute of Science, Hacettepe University, Beytepe, Ankara 06800, Turkey

reason, the processing and analysis of the signals received by the heart are very important for rapid diagnosis and treatment. In conventional methods, a suitable sampling method is used in the pre-processing phase of ECG signals and the signals are cleaned of noise. Then, the manual feature extraction phase begins, where it is very important to seek expert opinions. This phase is very critical as incorrect feature extraction can lead to misclassification of signals and serious errors in diagnosis and treatment. After all these phases are completed, classification is done using traditional classification algorithms. However, the studies show that the situation for deep learning algorithms has changed in recent years (Ozaltin et al. 2022; Özaltın and Yeniay 2021; Koc et al. 2022). Thanks to deep learning algorithms, successful classifications can be made automatically. In this way, the state of health of patients can be monitored with smartphones, watches, etc., even without an expert opinion.

The aim of the study was to recognize type of ECG efficiently via deep learning algorithm. Firstly, we collect the dataset from PhysioNet databases (Physionet 2020). The dataset consists of three different types: arrhythmia (ARR), congestive heart failure (CHF), and normal sinus rhythm (NSR). In this study, a novel convolutional neural networks (CNN) architecture, which is one of the deep learning algorithms, is proposed for automatic ECG signal classification. This newly proposed 34-layer CNN architecture is designed for two-dimensional images. In fact, the newly proposed CNN is considered not only ECG classification, but also other biomedical signals, images, etc. classification. In this context, the ECG signals are naturally transformed from one-dimensional signals into images by using a continuous wavelet transform (CWT) in the pre-processing phase. This wavelet transform has three different mother wavelet functions: Amor, Bump, and Morse, which are the most commonly used. The impact of these functions on classification performance is also examined. In this study, 360 Hz, 500 Hz and 1000 Hz sample lengths are examined whether the wave characteristics become more evident. Figure 1 shows the images (scalograms) obtained with different sampling lengths of ECG signals, 360 Hz, 500 Hz, and 1000 Hz, respectively. Therefore, a total of nine different datasets are obtained under these conditions. These datasets are classified separately with the same training options parameters using the proposed CNN, AlexNet, and SqueezeNet. After identifying the best wavelet function, sample length, and architecture, we additionally investigate another pre-processing method: STFT to measure ECG classification performance via different split methods: training and testing, and cross-validation. Finally, the proposed CNN is used as a deep feature

extractor from images and merged with support vector machines (SVM) to get trusted results.

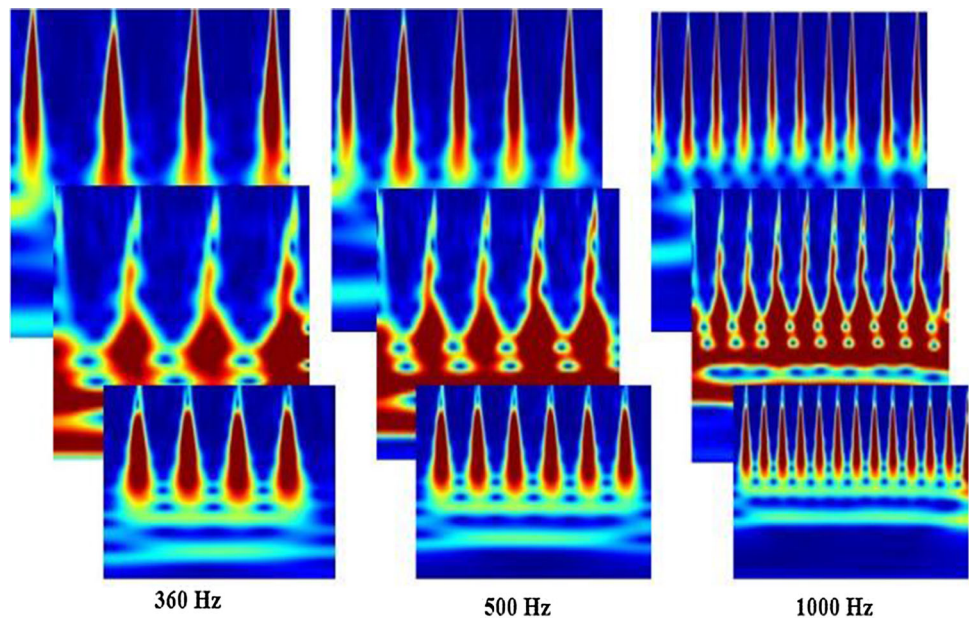
In this study, a hybrid algorithm is proposed to detect ECG types from acquired images based on a deep learning algorithm and a machine learning algorithm. The main contributions and novelties of this study are as follows:

- When using CWT, 500 Hz is observed as an efficient sample length while converting.
- Amor wavelet function has higher performance than others while applying CWT.
- A new CNN architecture called proposed CNN is presented and compared with AlexNet and SqueezeNet. Eventually, the proposed CNN has the highest performance.
- To measure the performance of the proposed CNN, STFT is also used as pre-processing method via different splitting methods: training and testing (80:20, 70:30), and k-fold cross-validation (5, 10). Finally, CWT is higher than it and cross-validation is the best splitting method.
- To improve classification performance, the proposed CNN is utilized as feature extractor and benefited from both fully connected layer and maximum pooling layer.
- Reduced features are classified using SVM.
- Consequently, the highest performance to recognize ECG types is acquired thanks to the proposed CNN–SVM hybrid algorithm.

## 1.1 Related studies

Nowadays, artificial intelligence is evolving day by day, and many studies are also being conducted to classify ECG signals and other biomedical signals using CNN architectures. Khorrami and Moavenian (2010) applied the CWT, discrete wavelet transform (DWT), and discrete cosine transform (DCT) to ECG signals. In addition, they compared SVM with multi-layer perceptron (MLP) algorithms in the classification phase. In particular, they found that combinations made with MLP (CWT-MLP, DWT-MLP, DCT-MLP) are superior to SVM. Al Rahhal et al. (2018) transformed signals from different datasets using CWT to identify arrhythmias in ECG signals. Also, they used the CNN algorithm and achieved an accuracy of 99% in the classification phase. Huang et al. (2019) converted ECG signals with STFT and obtained two-dimensional scalograms in their study. Moreover, they benefited from the CNN architecture for classifying these scalograms and achieved an accuracy of 99%. In addition, they also classified the one-dimensional ECG signals using CNN and found an accuracy of 90.93%. Krak et al. (2020) transformed ECG signals into the images using CWT and DWT in their study. Furthermore, they classified the images

**Fig. 1**  $227 \times 227 \times 3$  size of images with different sample lengths



using the CNN architecture and obtained an accuracy of 96% in the classification phase. Baloglu et al. (2019) designed a 10-layer end-to-end CNN architecture for the classification of multiclass one-dimensional ECG data and achieved an accuracy of a 99.78%. Mahmud et al. (2020) created a CNN architecture for multiclass one-dimensional ECG data and obtained an accuracy rate of 99.28%. Salem et al. (2018) utilized DenseNet architecture to classify transformed two-dimensional ECG data and achieved an accuracy of 97.23%. Zhao et al. (2020) proposed a CNN containing 24 layers for classifying transformed ECG data and achieved an accuracy of 87.1%. Xu and Liu (2020) created a CNN architecture in order to analyze ECG data recorded from a Holter device and achieved an accuracy of 99.4%. Rajkumar et al. (2019) suggested a CNN architecture for one-dimensional ECG data by using exponential linear unit (ELU) activation layers and achieved an accuracy of 93.6%. Hua et al. (2020) developed a CNN architecture for one-dimensional ECG signals and achieved an accuracy of 97.45%. Kiranyaz et al. (2015) proposed a CNN architecture for patient-specific real-time one-dimensional ECG classification and achieved an accuracy of 96.4%. Chen et al. (2020) suggested CNN + long short-term memory (LSTM) which can classify six kinds of ECG fragments. They have classified two ECG databases: MIT-BIH arrhythmia database and MIT-BIH arrhythmia database + Challenge2017, and achieved an accuracy of 99.32% and 97.15%, respectively, using CNN + LSTM. Sandeep et al. (2019) utilized the CNN architecture to classify ECG data and also achieved an accuracy of 90.63%. Furthermore, machine learning algorithms such as support vectors machine (SVM), K-nearest neighbors (KNN), decision tree (DT), extreme learning machine

(ELM), ensemble learning, and multi-layer perceptron (MLP) to classify ECG signals by many other researchers (Alickovic and Subasi 2015; Qaisar and Subasi 2020; Tuncer et al. 2022; Ceylan and Özbay 2007; Pławiak and Acharya 2020). Additionally, Table 1 shows recent studies on ECG signals classification.

The rest of the study is organized as follows: In Section 2, we present the materials and methods. Then, we explain the dataset, experimental setup, performance metrics, and experimental results in Section 3. Next, we discuss the results in Section 4. Finally, we conclude the study and state the future works.

## 2 Materials and methods

In this section, we first present pre-processing methods. Next, we introduce CNN, the proposed CNN, and pre-trained architectures: AlexNet (Krizhevsky et al. 2012) and SqueezeNet (Iandola et al. 2016). In the last, we present SVM and the proposed CNN–SVM architecture for classification of ECG dataset. Figure 2 shows the framework of this study.

### 2.1 Pre-processing methods

In this study, we propose a novel CNN it needs images; therefore, we transform one-dimensional signals into two-dimensional image datasets via continuous wavelet transform (CWT) and short-time Fourier transform (STFT).

**Table 1** Recent Studies on ECG signals classification

References	Dataset	Classes	Methods	Metrics
Xing et al. (2022)	MIT-BIH Arrhythmia Dataset	5 classes	SNN (Spiking Neural Network)	Accuracy 98.26% Sensitivity 94.75% F1-score 89.09%
Chen et al. (2022)	KMUH Dataset (Kaohsiung Medical University Hospital)	9 classes	DNN (Deep Neural Network)	Accuracy 96.02% Micro-F1 82.71%
Patezylski et al. (2022)	PTB (Physikalisch-Technische Bundesanstalt) Dataset	2 classes	CNN (1D)	Accuracy 90.4% Sensitivity 89.6% F1-score 89.9%
		5 classes	FSL (Few-Shot Learning + SVM)	Accuracy 79.0% Sensitivity 70.6% F1-score 70.6%
		20 classes	CNN (1D)	Accuracy 67.1% Sensitivity 32.4% F1-score 32.6%
Cheng et al. (2022)	MIT-BIH Arrhythmia Dataset	5 classes	U-net (1D)	Accuracy 95.5% Sensitivity 95.55% Specificity 97.64%
Jiao and Wu (2022)	MIT-BIH Arrhythmia dataset	5 classes	Capsule network (LSTM + CNN)	Accuracy 99.3% Sensitivity 83.1% F1-score 86.2%
Sepahvand and Mohammadi (2022)	Chapman ECG dataset (Chapman University and Shaoxing People's Hospital)	7 classes	CNN	Accuracy 98.15% Sensitivity 97.11% F1-score 97.55%
Eltrass et al. (2022)	MIT-BIH Arrhythmia dataset	3 classes	CNN (AlexNet)	Accuracy 98.74% Sensitivity 98.17% Specificity 99.00%
	MIT-BIH normal sinus rhythm dataset BIDMC dataset			
Kumar et al. (2022)	Collected ECG signals based on IoT (Internet of things)	2 classes	Coy-Grey Wolf optimization + CNN	Accuracy 95.05% Sensitivity 94.639% Specificity 94.639%
Meng et al. (2022)	MIT-BIH Arrhythmia dataset	3 classes	Transformer with LightConv attention + CNN embedding	Accuracy 99.32%

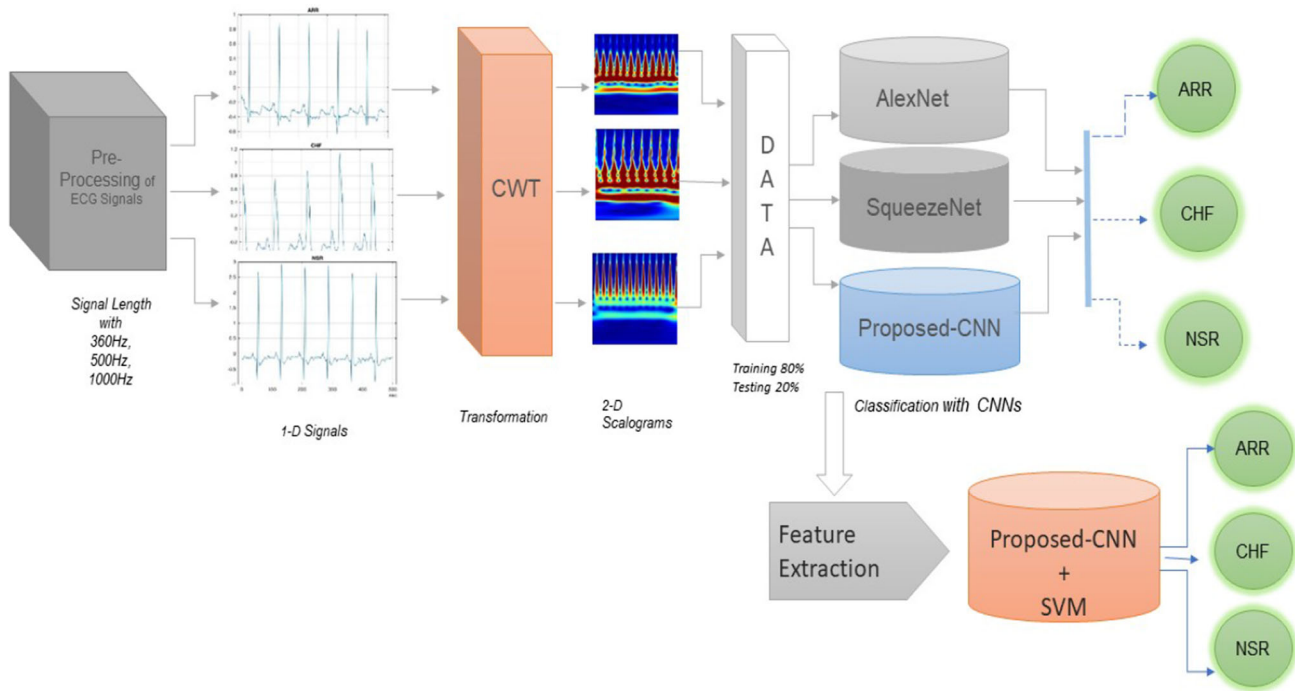


Fig. 2 Flowchart of this study

2.1.1 Max–min normalization

In this study, firstly, we normalize raw one-dimensional ECG signals using the minimum–maximum normalization method given formula in Eq. (1) as follows:

$$X = \frac{\text{signal} - \min(\text{signal})}{\max(\text{signal}) - \min(\text{signal})} \tag{1}$$

where  $X$  denotes the normalized ECG signal. Besides,  $\min(\cdot)$  is a minimum function, and  $\max(\cdot)$  is a maximum function.

2.1.2 Continuous wavelet transform

Continuous wavelet transform (CWT) is a transformation method. CWT allows simple analysis of its frequency components and can transform a one-dimensional signal into a two-dimensional scalogram by providing a mapping of the signal also on the time axis. The mathematical formulation of the CWT and WT family is offered in Eq. (2) and Eq. (3), respectively,

$$\text{CWT}(a, b) = \langle f, \psi_{a,b}^* \rangle = \int_{-\infty}^{+\infty} f(t) \psi_{a,b}^*(t) dt \tag{2}$$

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \tag{3}$$

where  $f(t)$  is a continuous signal function received in this study as an ECG signal function,  $\psi_{a,b}(t)$  is the mother

wavelet function,  $a$  indicates a scale parameter,  $b$  indicates the shift parameter or translation, and the symbol of  $*$  indicates the complex conjugate function (Lee and Choi 2019). Besides,  $\langle f, \psi_{a,b} \rangle$  is expressed as a function of the inner products of Eq. (2). It CWT( $a, b$ ) is regulated,

$$\text{CWT}(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(t) \psi\left(\frac{t-b}{a}\right) dt \tag{4}$$

will be in the form like in Eq. (4). The signal function  $f(t)$  can be converted from the inverse of CWT( $a, b$ ), as follows:

$$f(t) = \frac{1}{C} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \text{CWT}(a, b) \frac{\psi_{a,b}(t)}{|a|^{3/2}} da db \tag{5}$$

where  $C$  indicates the normalization constant depending on the choice of the mother wavelet function in Eq. (5) (Lee and Choi 2019).

Some mother wavelet functions as follows:

$$\psi_{\text{Morl}}(t) = e^{2\pi i t} e^{-\frac{t^2}{2\sigma^2}} = (\cos 2\pi t + i \sin 2\pi t) e^{-\frac{t^2}{2\sigma^2}} \tag{6}$$

$$\psi_{\text{Mexh}}(t) = \left(1 - \frac{t^2}{\sigma^2}\right) e^{-\frac{t^2}{2\sigma^2}} \tag{7}$$

$$\psi_{\text{Bump}}(ab) = e^{\left(1 - \frac{1}{1-ab-\mu/\sigma^2}\right)} \chi[\mu - \sigma, \mu + \sigma] \tag{8}$$

will be in the form in Eqs. (6–8). Here,  $\psi_{\text{Morlet}}(t)$ , Morlet,  $\psi_{\text{Mexh}}(t)$ , Mexican hat, and  $\psi_{\text{Bump}}(ab)$ , Bump, show the mother wavelet function (Lee and Choi 2019).

### 2.1.3 Short-time Fourier transform (STFT)

Short-time Fourier transform (STFT) is also a transformation method. The STFT is obtained from the discrete Fourier transform (DFT), to discover the sudden frequency and the sudden amplitude of localized waves with time-varying typical (Huang et al. 2019; Haykin and Veen 1999). The STFT uses a window function to extract time-domain information (Toma and Choi 2022). The window function possesses a certain interval, and the value of this window function outward of the interval is zero (Toma and Choi 2022). To calculate the frequency domain information, the window function shifts over all non-stationary signals and each time it is multiplied with the signal (Haykin and Veen 1999; Toma and Choi 2022). Further, the time–frequency spectrogram can be computed in a discretized non-stationary digital signal as given in Eq. (9) (Toma and Choi 2022),

$$\text{STFT}\{x[n]\} = X(m, \omega) = \sum_{n=-\infty}^{\infty} x[n]w[n-m]e^{-j\omega n} \quad (9)$$

where  $x[n]$  symbolizes signals and  $w[n]$  is the window function. In this study, we utilize the Kaiser function with a window size of 500 Hz. Thus, we convert ECG signals into ECG spectrums images with dimensions of  $227 \times 227 \times 3$ .

## 2.2 Convolutional neural network (CNN)

Convolutional neural network (CNN) emerges as a specialized deep learning approach for analyzing two-dimensional data. Not only it is preferred algorithm in the analysis of multidimensional data but also one-dimensional data. Other classifications and clustering algorithms are difficult to apply to real-time data due to their computational complexity (Narin 2020). For this reason, deep learning technology that can overcome this complexity evolves day by day. Moreover, CNN can perform feature extraction and classification automatically using raw data, so deep learning algorithms are very popular in the field of artificial intelligence. Further, it is found to give very good results of classification studies involving both big data and small data by researchers. Thanks to the CNN algorithm, ECG signals can be analyzed and observed on smartphones, watches, Holter monitoring devices, etc. (Huang et al. 2019).

The CNN processes an image in different layers and separates all its features. The most commonly used layers are:

1. Convolution layer,
  2. Nonlinear layer,
  3. Pooling layer,
  4. Flattening layer,
  5. Fully connected layer expressed as (Baloglu et al. 2019; Lee and Choi 2019; Acharya et al. 2017).
1. Convolutional Layer: The convolution process is the layer where the features of the image are determined. To determine more than one feature, the number of convolutional layers increases in the same proportion. This layer is the main building block of CNN.
  2. Nonlinear Layer: This layer is also known as the activation layer. It is used to realize the activation of the system with nonlinear functions. Rectified linear unit function (ReLU), which is widely used because it is faster than others, is preferred in recent years.
  3. Pooling Layer: Smaller matrices are obtained while preserving the properties of the existing input. In this way, the computational complexity is reduced.
  4. Flattening Layer: The matrix format data obtained from the previous step is prepared following the fully connected layer.
  5. Fully Connected Layer: It is the most important layer of convolutional neural network layers. The data are taken from the flattening layer and trained by the neural network and the learning process is performed.

## 2.3 Pre-trained architectures: AlexNet and SqueezeNet

AlexNet (Krizhevsky et al. 2012) has five convolution layers combined with max-pooling layers and three fully connected layers. It also includes a dropout layer and a softmax. Moreover, each layer is activated with the ReLU activation function. In 2012, it was used the ReLU activation function in place of the tanh function (Abdelmalek et al. 2019). Thus, it was seen that the architecture was accelerated. The total number of parameters is 62.3 million, and the input image size is  $227 \times 227$ .

SqueezeNet (Iandola et al. 2016) starts with an independent convolutional layer (conv1), follows by eight firing modules, and ends with the last convolutional layer (conv10). In total, it consists of ten convolutional layers, some max-pooling layers, and a SoftMax layer, in the recently presented version.

In this study, a novel CNN architecture is presented in the next section and it is compared with AlexNet and SqueezeNet on created different datasets.

## 2.4 Novel proposed CNN architecture

A CNN architecture usually consists of an input layer, some convolutional layers, some pooling layers, and a fully connected layer (Krak et al. 2020). In this study, we introduce a novel CNN architecture. It has seven convolutional layers, seven batch normalization layers, seven activation layers (ReLU), seven maximum pooling layers, and two fully connected layers with one dropout layer. Additionally, a SoftMax layer and a classification layer with an entropy approach are used as well. The convolution layers are effectively utilized for feature extraction from ECG image datasets. This is important since well feature extraction is also meaning very sensitive classification. Essentially, these layers are filtered to enhance the features of the primary signal while reducing the noise (Hua, et al. 2020; Li et al. 2018). The pooling layers reduce the dimension of the input images, and these are prepared for the next layer. Finally, extensive features in the fully connected layers are reduced with 0.5 probability by using the dropout layer and transferred to the SoftMax layer for the classification. Details of the parameters of the proposed CNN are given in Table 2.

The proposed CNN is a novel architecture that has different filter sizes, number of filters, strides, and padding. Fundamentally, we develop the architecture for biomedical image classification. However, it is tested on known classical datasets such as CIFAR-10, like other CNN architectures. Additionally, it is utilized on Physikalisch-Technische Bundesanstalt (PTB) Diagnostic ECG Database (Özaltın and Yeniay 2021; Goldberger, et al. 2000). This proposed CNN is performed for not only signals but also brain computed tomography, detailed in Ozaltın et al. (2022). Moreover, this proposed CNN is named as OzNet in studies of Ozaltın et al. (2022). And, this architecture obtains successful performances in these datasets (Fig. 3).

In this study, the proposed CNN is compared with AlexNet and SqueezeNet using same fine-tuning parameters. Stochastic gradient descent method (sgdm) is performed as the optimization algorithm, and the momentum parameter is determined as 0.95, and the learning rate is also started with 0.0001 as constant. Figure 4 shows the proposed CNN scheme.

## 2.5 Deep feature extraction

In this study, the proposed CNN can extract features from images effectively. Therefore, we use it both classifier and deep feature extractor. Although, when it is used for classification algorithm, the results are quite well, we decide to more improving results for obtaining the best one. Therefore, we designed a hybrid algorithm which is included the

proposed CNN and SVM. In this section of study, the proposed CNN is assigned as automatic feature extractor from ECG images and SVM is employed for classifier. In brief, we can explain the steps of how to work it as follows: (i) the proposed CNN is trained on ECG images, firstly. (ii) Reduced features are obtained from the proposed CNN of fully connected layer and 4096 features are collected for each image. (iii) To classify with these features, the dataset is split into 30% training set and 70% testing set. This is because we want to obtain trustworthy classification results owing to dropout layer would not have much influence (Elleuch et al. 2016; Srivastava et al. 2014). Then, the trained net is activated. (iv) SVM classifier is employed to detect type of ECG, effectively. The same stages are happened when reduced features are achieved from maximum pooling (Max-Pooling 7) layer. Figure 5 demonstrates the scheme of the proposed CNN–SVM.

## 2.6 Support vector machine (SVM)

Support vector machine (SVM) is a machine learning algorithm that an effective separation with a kernel-based method to the datasets for classification or regression (Koklu and Ozkan 2020). It is improved by Cortes and Vapnik (1995) for two classes. Then, the algorithm is advanced and generalized for multiclass and nonlinear datasets. In general, the dataset can be separated in high-dimensional feature space with a kernel function. Also, SVM can be overcome confused datasets and overfitting. The most common representation of the SVM function is  $f(x) = w^T \phi(x) + b$  where  $w \in R^n$   $b \in R$  and  $\phi(x)$  is a feature map.

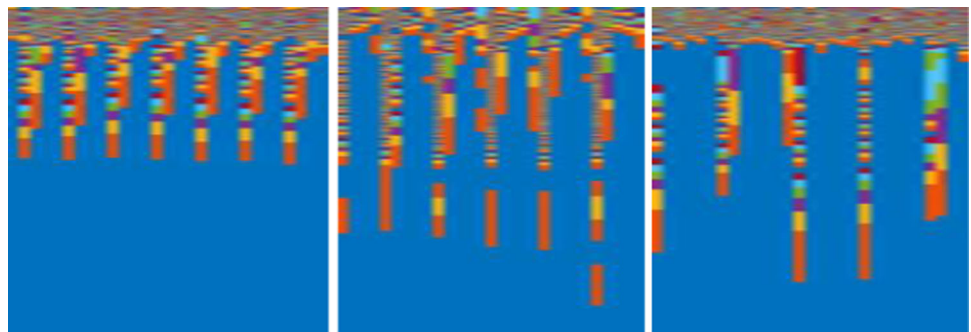
## 3 Results

### 3.1 ECG dataset

In this study, we benefit from three different ECG datasets from PhysioNet databases (Physionet 2020). Each raw ECG dataset is taken with a signal length of 1 h and sampled at 128 Hz. The first ECG dataset consists of the ECG recordings from 48 patients, which contain two leads. It is received from the MIT-BIH Arrhythmia Database and referred to as ARR (Goldberger, et al. 2000; Moody and Mark 2001). The next ECG dataset consists of the ECG recordings from 15 patients, which contain two leads. It comes from the BIDMC Congestive Heart Failure Database and is named CHF (Goldberger, et al. 2000; Baim et al. 1986). The final ECG dataset consists of the ECG recordings from 18 patients, containing two leads. It is obtained from MIT-BIH Normal Sinus Rhythm and

**Table 2** Proposed CNN architecture details

Layer name	Type	Layer parameters	Output shape
Input	Image Input	$227 \times 227 \times 3$ images with “zerocenter” normalization	$227 \times 227 \times 3$
Conv-1	Convolution 2D	Filter size = 64, number of filters = [5 5], stride = [1 1], padding = [1 1 1 1], BatchNormalization, ReLU	$225 \times 225 \times 64$
MaxPool-1	Max Pooling	Pool size = [3 3], stride = [2 2], padding = [0 0 0 0]	$112 \times 112 \times 64$
Conv-2	Convolution 2D	Filter size = 128, number of filters = [3 3], stride = [1 1], padding = [1 1 1 1], BatchNormalization, ReLU	$112 \times 112 \times 128$
MaxPool-2	Max Pooling	Pool size = [3 3], stride = [2 2], padding = [0 0 0 0]	$55 \times 55 \times 128$
Conv-3	Convolution 2D	Filter size = 128, number of filters = [13 13], stride = [1 1], padding = [0 0 0 0], BatchNormalization, ReLU	$55 \times 55 \times 128$
MaxPool-3	Max Pooling	Pool size = [3 3], stride = [2 2], padding = [0 0 0 0]	$27 \times 27 \times 128$
Conv-4	Convolution 2D	Filter size = 256, number of filters = [7 7], stride = [1 1], padding = [1 1 1 1], BatchNormalization, ReLU	$27 \times 27 \times 256$
MaxPool-4	Max Pooling	Pool size = [2 2], stride = [2 2], padding = [0 0 0 0]	$13 \times 13 \times 256$
Conv-5	Convolution 2D	Filter size = 128, number of filters = [3 3], stride = [1 1], padding = [1 1 1 1], BatchNormalization, ReLU	$13 \times 13 \times 128$
MaxPool-5	Max Pooling	Pool size = [3 3], stride = [2 2], padding = [0 0 0 0]	$6 \times 6 \times 128$
Conv-6	Convolution 2D	Filter size = 128, number of filters = [3 3], stride = [1 1], padding = [1 1 1 1], BatchNormalization, ReLU	$6 \times 6 \times 128$
MaxPool-6	Max Pooling	Pool size = [3 3], stride = [2 2], padding = [0 0 0 0]	$3 \times 3 \times 128$
Conv-7	Convolution 2D	Filter size = 128, number of filters = [3 3], stride = [1 1], padding = [1 1 1 1], BatchNormalization, ReLU	$3 \times 3 \times 128$
MaxPool-7	Max Pooling	Pool size = [2 2], stride = [2 2], padding = [0 0 0 0]	$1 \times 1 \times 128$
FC-8	Fully Connected	4096	$1 \times 1 \times 4096$
Drop-8	Dropout	50%	
FC-9	Fully connected	3 (number of classes)	$1 \times 1 \times 3$
Softmax	SoftMax		$1 \times 1 \times 3$
Output	Classification	Cross-entropy	

**Fig. 3** Transformed images using STFT



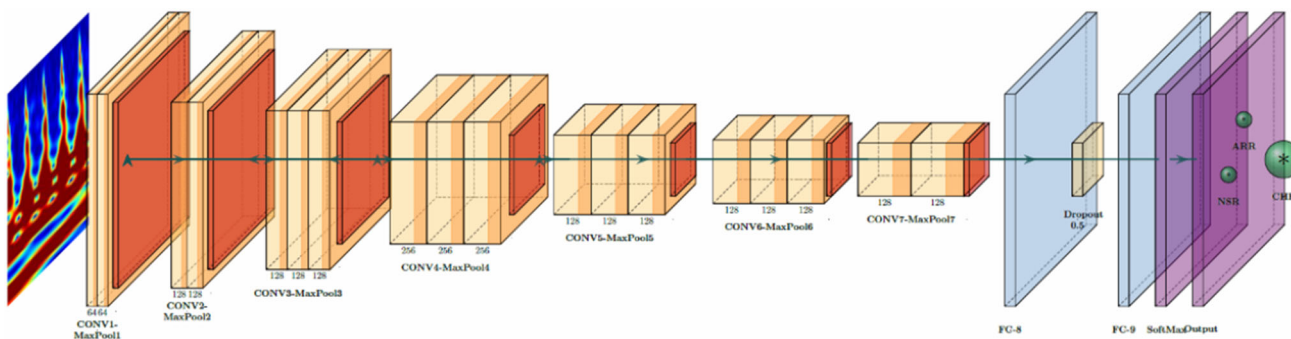


Fig. 4 Proposed CNN architecture

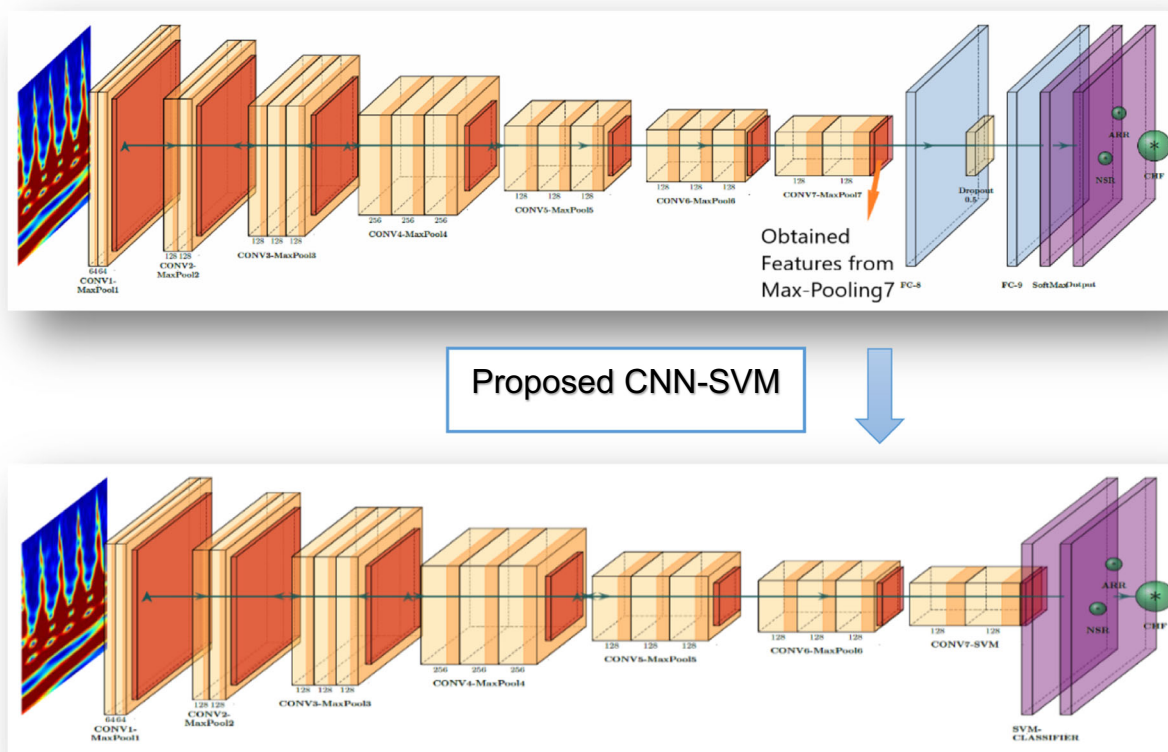


Fig. 5 Proposed CNN-SVM algorithm

referred to as NSR (Goldberger, et al. 2000). There are a total of 96 ARR, 30 CHF and 36 NSR in the ECG dataset. In fact, this dataset is not suitable for convolutional neural networks because of demand pattern. That is why we convert the signals into the images. First, we normalize the dataset using the max-min normalization method. Next, one-dimensional ECG signals are transformed into images utilizing CWT with different sampling lengths of signals, 360 Hz, 500 Hz, and 1000 Hz. This is because we want to compare which sample length is better to see differences. Besides, three different mother wavelet functions: Amor,

Bump, and Morse, are applied to each sample length to compare which mother wavelet function is better to detect differences. It also sizes each image to  $227 \times 227 \times 3$  and.jpg format. Therefore, we create nine different balanced datasets with identifying mother wavelet functions and signal lengths. Each dataset contains 900 images, and each class (ARR, CHF, and NSR) includes 300 images. After that, to compare the results, we also benefit from the STFT transform method to turn signals into images. Also, created this dataset consists of 900 images, and each class contains 300 images.

### 3.2 Experimental setup

In this study, we run AlexNet, SqueezeNet, and the proposed CNN to classify ECG datasets. In this study, we use splitting methods: training and testing sets, and cross-validation to compare affective classification performance. Primarily, the dataset is split conventionally as a training and testing set with 80:20 and 70:30 percentages. Next, k-fold cross-validation is performed, where k values are determined as 5 and 10. Further, we use the proposed CNN to automatically extract deep features. They are reached from the fully connected layer (FC-8) and maximum pooling layer (Max-Pooling 7), respectively. To classify these reduced features, we perform an SVM using Gaussian kernel function to detect ECG type from images. Therefore, we present a comprehensive study that effectively determines the ECG type.

### 3.3 Performance metrics

In this study, we review performance metrics of CNN architectures that are accuracy, sensitivity, specificity, precision, and *F1*-score in Eq. (8–12), as follows (Xu and Liu 2020; Abdelmalek et al. 2019):

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (10)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \times 100\% \quad (11)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \times 100\% \quad (12)$$

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\% \quad (13)$$

$$\text{F1 - Score} = \frac{2 \times \text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \times 100\% \quad (14)$$

where TP: true positive, FP: false positive, TN: true negative, and FN: false negative are expressed.

### 3.4 Experimental results

This study is conducted in a MATLAB 2021b environment with Intel Core i7-7500U CPU, NVIDIA GeForce GTX 950 M, 16 GB RAM and 64-bit Operating System. The aim of this study was to identify ECG types via CNN architectures and a designed hybrid algorithm. First, nine different ECG image datasets are created using CWT, and each is classified using AlexNet, SqueezeNet, and the proposed CNN with the same option parameters with 80:20 training and testing split percentages. In addition, the obtained results are tested with the Wilcoxon signed rank test. Tables 3 and 4 show both performance results and

paired comparisons for statistical significance. Besides, all comparisons are demonstrated in Fig. 5.

When Table 3 is examined in relation to the sample length of the ECG between AlexNet and the proposed CNN, AlexNet gets a maximum accuracy of 94.67% at a sample length of 500 Hz. Also, the proposed CNN achieves the maximum accuracy of 98.00% at a sample length of 500 Hz. Finally, SqueezeNet achieves a maximum accuracy of 94.67% with a sample length of 500 Hz, as given in Table 4. Therefore, we can indicate that 500 Hz is the best one for the sample length of ECG.

When Tables 3 and 4 are also examined in terms of the mother wavelet function, Amor and Morse provide almost similar results to classify images for AlexNet and our proposed CNN. However, these results do not apply to SqueezeNet. When SqueezeNet is examined for the mother wavelet function, Bump is found to be the best. So, if researchers want to use SqueezeNet, they can choose to use the bump wavelet function while performing CWT. When Tables 3 or 4 is investigated for the proposed CNN in terms of the mother wavelet function, Amor's choice for classifying the images is the best.

Although the results are quite good, we want to test these results for the reliability of this study using the nonparametric method, the Wilcoxon signed rank test. First, we make one hypothesis, which is a null hypothesis: there is no difference between AlexNet and the proposed CNN, and an alternative hypothesis: there is a difference between AlexNet and the proposed CNN. As a result, p value is obtained  $0.018 < 0.05$ , and hence, null hypothesis is rejected. In this study, a significant level is determined as 0.05. Therefore, we can statistically say that there is a difference between AlexNet and the proposed CNN.

Though the results are rather good, as given in Table 4, we want to test these results for the trustfully of this study using the Wilcoxon signed rank test. First, we make one hypothesis, which is a null hypothesis: there is no difference between SqueezeNet and the proposed CNN, and an alternative hypothesis: there is a difference between SqueezeNet and the proposed CNN. As a result, p value is obtained  $0.024 < 0.05$ , and hence, null hypothesis is rejected. Thus, we can statistically express that there is a difference between SqueezeNet and the proposed CNN.

As a result, the proposed CNN is the best choice to classify ECG datasets while using CWT and 80:20 training and testing percentages. Figures 6 and 7 display performance graphs for classification. In addition, Table 5 details the results with other performance metrics for each class.

When all of the performance metrics in Table 5 are examined, these proposed CNNs metrics are met at over 96%. Specifically, the NSR performances are considered to be %100 in terms of specificity and precision score. In

**Table 3** Wilcoxon signed rank test for proposed CNN and AlexNet

Datasets		AlexNet (accuracy%)	Proposed CNN (accuracy%)	$D_i$	$r( D_i )$	
ECG signal length	Wavelet function					
360 Hz	Amor	89.33	92.67	-3.34	7	
	Bump	91.33	92	-0.67	1	
	Morse	94	94	0	-	
500 Hz	Amor	94.67	<b>98</b>	-3.33	6	
	Bump	94	94	0	-	
	Morse	93.33	94.67	-1.34	2.5	
1000 Hz	Amor	93.33	95.33	-2	4.5	
	Bump	92	94	-2	4.5	
	Morse	93.33	94.67	-1.34	2.5	
Wilcoxon signed ranks test		$Z = -2.375, p - value = 0.018$			$\sum_{i=1}^{n=7} r( D_i ) = 28$	

**Table 4** Wilcoxon signed rank test for proposed CNN and SqueezeNet

Datasets		SqueezeNet (accuracy%)	Proposed CNN (accuracy%)	$D_i$	$r( D_i )$	
ECG signal length	Wavelet function					
360 Hz	Amor	90	92.67	-2.67	4	
	Bump	94	92	+ 2	2.5	
	Morse	89.33	94	-4.67	6.5	
500 Hz	Amor	87.33	98	-10.67	9	
	Bump	94.67	94	+ 0.67	1	
	Morse	90.67	94.67	-4	5	
1000 Hz	Amor	88.67	95.33	-6.66	8	
	Bump	92	94	-2	2.5	
	Morse	90	94.67	-4.67	6.5	
Wilcoxon signed ranks test		$Z = -2.255, p - value = 0.024$			$\sum_{i=1}^{n=7} r( -D_i ) = 41.5$ $\sum_{i=1}^{n=2} r( +D_i ) = 3.5$	

addition, its performances on other metrics are also over 98%.

When the metrics are examined, which the classifiers did well, it is noticeable that the proposed CNN’s F1-score is superior to the others in Table 5. Therefore, the proposed CNN is determined to be the best classifier in terms of performance metrics.

As a result of this part, the best signal length, mother wavelet function, and architecture are determined to be 500 Hz, Amor, and the proposed CNN, respectively. Thus, these foundations have shown that only one ECG dataset is classified. In addition, Fig. 8 shows the accuracy rate graph and the loss graph for the proposed CNN, while the signal length is 500 Hz and the wavelet function is Amor.

Having determined the proposed CNN as the best architecture for classifying ECG images, we examine the impact of other split methods on performance. First, the ECG image dataset created with 500 Hz sample length and Amor wavelet function using CWT is divided into 80:20 and 70:30 training and test sets, respectively, and then, we use a fivefold and tenfold cross-validation. The results are shown in Tables 6, 7, and 8.

When Table 6 is viewed, all mean performance metrics are observed above 96.52% and also the maximum standard deviation (Std) was 0.0173. Therefore, the proposed architecture is traditionally trained and tested to classify images.

According to Table 7, all mean performance metrics are above 95.3% and also the maximum standard deviation (Std) was 0.01705. Thus, it can be said that a training and testing split of 80:20 has the best performance for classifying ECG images while performing CWT.

According to Table 8, all average performance metrics are seen, and the maximum average accuracy of 97.22% is obtained through tenfold cross-validation. Concluding on the use of CWT, the performances expressed that the cross-validation is better than the split method for training and testing. Perfect performances for classifying ECG images are achieved using CWT and the proposed CNN. However, we would like to see how other pre-processing methods affect the performance of the proposed CNN using the same splitting methods. Therefore, we prefer to use STFT method which is performed widely. Its performances are shown in Tables 9, 10, and 11.

According to Table 9, all the average performance metrics are shown above 89.3% and also, the maximum standard deviation (Std) was 0.013438. Therefore, the proposed architecture is traditionally trained and tested at 80:20 to classify images using STFT.

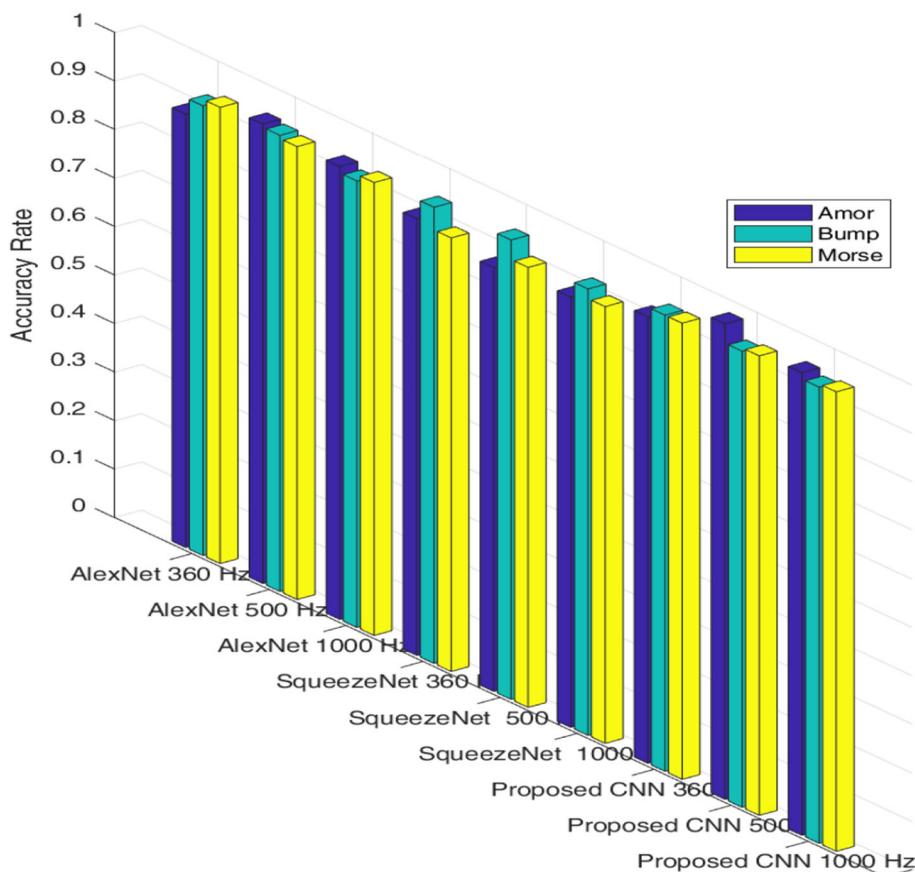
According to Table 10, all the average performance metrics are shown above 89.2% and also, the maximum standard deviation (Std) was 0.0169. Therefore, when the

proposed architecture is trained and tested at 70:30 to classify images using STFT, performance results are similar to 80:20 training and testing split.

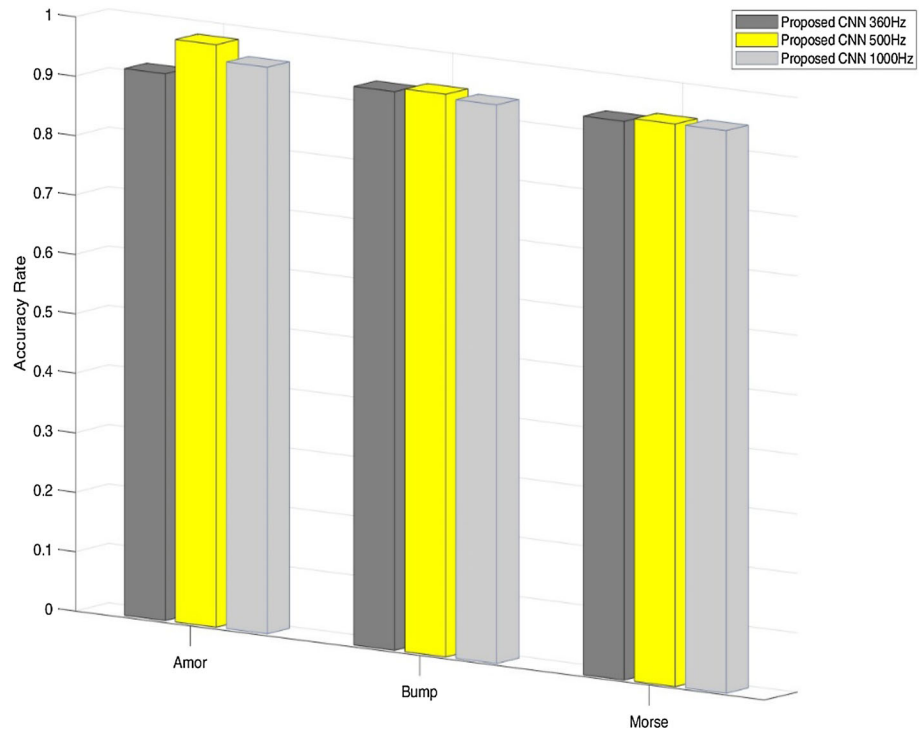
According to Table 11, all average performance metrics are observed, and the maximum average accuracy of 91.11% is achieved through fivefold cross-validation. Final on the use of STFT, the performances indicated that the cross-validation is better than the split method for training and testing. Compared with CWT, STFT is not preferred to create ECG images as its performances are lower than CWT using the proposed CNN. In general, however, the proposed CNN in this study achieves quite good classification performance for recognizing ECG types.

Indeed, in this study, our main contributor wants to find the best algorithm to detect ECG types. Thus, the proposed CNN is used as a deep feature extractor from images. Having trained proposed CNN for the ECG images using CWT through 80:20 splitting method because of the highest accuracy rate, reduced features are obtained from the fully connected (FC-8) layer and maximum pooling layer (Max-Pooling 7), respectively. These features are classified using SVM classifier. Therefore, we designed novel hybrid algorithm thanks to the proposed CNN and SVM. Table 12 exhibits performance results.

**Fig. 6** Performance comparison of different sample lengths and mother wavelet function using CNN architectures



**Fig. 7** Performance comparison of different sampling lengths and mother wavelet function using the proposed CNN



**Table 5** Performance metrics of proposed CNN, AlexNet, and SqueezeNet architectures

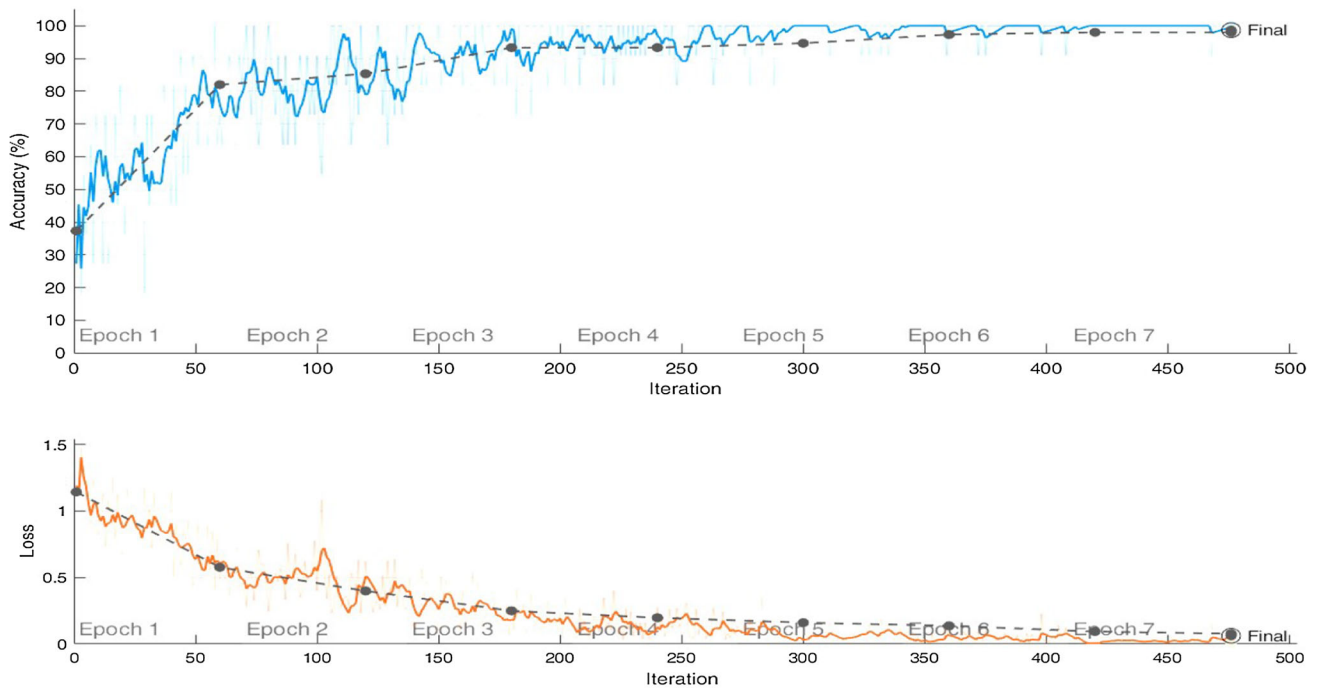
CNN architecture	Class name	Sensitivity (%)	Specificity (%)	Precision (%)	F1-score (%)	Test Accuracy Rate (%)
Proposed-CNN <sup>a</sup>	ARR	97.96	98.02	96	96.97	<b>98</b>
	CHF	98	99.02	98	98	<b>98</b>
	NSR	98.04	<b>100</b>	<b>100</b>	99.01	<b>98</b>
Proposed-CNN <sup>b</sup>	ARR	93.88	96.04	92	92.93	95.33
	CHF	96.15	<b>100</b>	<b>100</b>	98.04	95.33
	NSR	95.92	98.01	94	94.95	95.33
AlexNet <sup>a</sup>	ARR	92.31	97.96	96	94.12	94.67
	CHF	97.87	96.12	92	94.85	94.67
	NSR	94.12	98.6	96	95.05	94.67
SqueezeNet <sup>a</sup>	ARR	94.34	100	100	97.09	94.67
	CHF	90.57	97.94	96	93.20	94.67
	NSR	100	96.05	88	93.62	94.67

<sup>a</sup>ECG sample length 500 Hz, <sup>b</sup>ECG sample length 1000 Hz

\*Bold values indicate the maximum metrics

According to Table 12, all performance metrics are increased for two different processes. However, the highest accuracy of 99.21% is achieved when retrieving features from Max-Pooling 7 layer. In this study, while using CWT, the proposed CNN–SVM is seen as the best algorithm for recognizing ECG types. Additionally, Fig. 9 displays a confusion matrix of the proposed CNN–SVM with the highest.

This study is conducted not only with CNN, but also with an SVM classifier, which is very successful in image classification. The combination of these two methods, which are very successful individually, has proven itself very well. Table 13 shows a comparison of all methods in terms of performance metrics while using CWT.



**Fig. 8** Accuracy rate and loss graph of training progress using the proposed CNN

**Table 6** Proposed CNN performance metrics over five training sessions with an 80:20 training and testing split using CWT

Training number	Sensitivity (%)	Specificity (%)	Precision (%)	F1-score (%)	Test accuracy rate (%)
1	94.86	97.60	94.67	94.66	94.67
2	95.53	97.72	95.33	95.33	95.33
3	96.05	98.12	96.00	96.00	96
4	98.06	99.03	98.01	97.99	98
5	98.68	99.45	98.67	98.67	98.67
Mean + Std	96.64 ± 1.65	98.38 ± 0.82	96.54 ± 1.72	96.53 ± 1.73	96.53 ± 1.73

**Table 7** Proposed CNN performance metrics over five training sessions with a 70:30 training and testing split using CWT

Training number	Sensitivity (%)	Specificity (%)	Precision (%)	F1-score (%)	Test accuracy rate (%)
1	92.96	96.48	93.49	92.93	92.96
2	94.44	97.22	94.57	94.43	94.44
3	96.67	98.33	96.70	96.66	96.67
4	95.92	97.96	96.24	95.96	95.93
5	97.04	98.51	97.06	97.04	97.04
Mean + Std	95.41 ± 1.69	97.70 ± 0.84	95.61 ± 1.52	95.4 ± 1.705	95.41 ± 1.69

## 4 Discussion

In this study, we aim to investigate whether ECG types are distinguishable from ECG-created images using deep learning structures and which type of ECG images (CWT

or STFT) is efficient in recognizing ECG types using deep learning. Actually, our study possesses some advantages and disadvantages as follows:

Advantages of this study are as follows: (i) Different sample lengths (360 Hz, 500 Hz, and 1000 Hz) are

**Table 8** Proposed CNN performance metrics with fivefold and tenfold cross-validation using CWT

<i>k</i>	Sensitivity (%)	Specificity (%)	Precision (%)	<i>F1</i> -score (%)	Accuracy rate (%)
5	96.44	98.22	96.44	96.44	96.44
10	<b>97.22</b>	<b>98.61</b>	<b>97.23</b>	<b>97.22</b>	<b>97.22</b>

\*Bold values indicate the maximum average metrics

**Table 9** Proposed CNN performance metrics over five training sessions with an 80:20 training and testing split using STFT

Training number	Sensitivity (%)	Specificity (%)	Precision (%)	<i>F1</i> -score (%)	Test accuracy rate (%)
1	87.78	93.89	87.74	87.73	87.78
2	90.55	95.27	90.69	90.57	90.56
3	88.89	94.44	89.15	88.89	88.89
4	89.44	94.72	90.28	89.27	89.44
5	90.56	95.28	91.04	90.29	90.56
Mean + Std	89.44 ± 1.177	94.72 ± 0.588	89.78 ± 1.3438	89.35 ± 1.142	89.45 ± 1.18

**Table 10** Proposed CNN performance metrics over five training sessions with a 70:30 training and testing split using STFT

Training number	Sensitivity (%)	Specificity (%)	Precision (%)	<i>F1</i> -score (%)	Test accuracy rate (%)
1	87.04	93.52	88.02	86.72	87.04
2	91.48	95.74	91.77	91.43	91.48
3	89.25	94.62	90.05	89.14	89.26
4	89.25	94.63	89.34	89.23	89.26
5	89.63	94.81	89.95	89.71	89.63
Mean + Std	89.33 ± 1.58	94.66 ± 0.789	89.83 ± 1.35	89.25 ± 1.69	89.33 ± 1.578

**Table 11** Proposed CNN performance metrics with fivefold and tenfold cross-validation using STFT

<i>k</i>	Sensitivity (%)	Specificity (%)	Precision (%)	<i>F1</i> -score (%)	Accuracy rate (%)
5	91.11	95.56	91.16	91.06	91.11
10	87.66	93.83	87.87	87.57	87.66

researched while using CWT, and 500 Hz is seen as an efficient sample length when one-dimensional signals are converted into images. (ii) Different mother wavelet functions (Amor, Morse, and Bump) are examined which one is more efficient on CNN architectures classification performance while performing CWT. (iii) This study presents a novel CNN architecture, called proposed CNN, and it is compared with AlexNet and SqueezeNet. (iv) Amor wavelet function is viewed successfully when using AlexNet and the proposed CNN, and the Bump wavelet function is high performance for SqueezeNet. (v) The proposed CNN has the highest performance in generating ECG datasets and is tested for significant differences via the Wilcoxon signed rank test. (vi) CWT is compared with the STFT method using the proposed CNN. (vii) Performances are measured on different splitting methods:

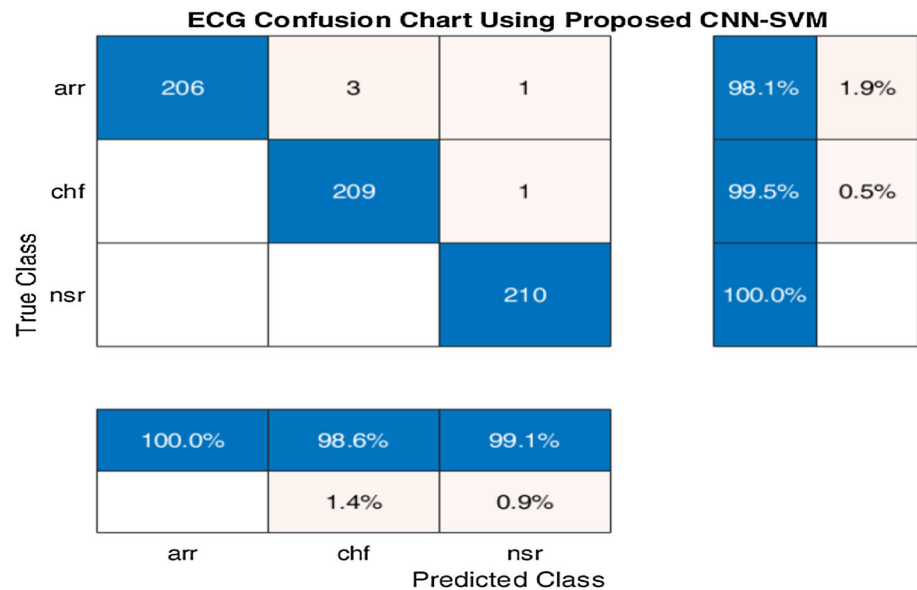
training and testing (80:20, 70:30), and *k*-fold cross-validation (5, 10). (viii) The proposed CNN is performed as a deep feature extractor and provides from fully connected and maximum pooling layer. (ix) As a result, a new hybrid algorithm with the proposed CNN and SVM is designed. In this stage, SVM is used as a classifier to increase the performance of the distinguishability of ECG types. Disadvantages of this study are researched limited ECG types (ARR, CHF, and NSR) and the number of individuals.

Many approaches are used for the classification of arrhythmia (ARR), congestive heart failure (CHF), and normal sinus rhythm (NSR) datasets. Basically, successful classification is very important for diagnosis and treatment. Therefore, in this study, we propose a novel 34-layer deep learning algorithm, called proposed CNN. Besides this ECG dataset, other datasets have also been classified using

**Table 12** Performance metrics of proposed CNN–SVM algorithm

Layer name	Sensitivity (%)	Specificity (%)	Precision (%)	<i>F1</i> -score (%)	Test accuracy rate (%)
Max Pooling-7	<b>99.206</b>	<b>99.66</b>	<b>99.213</b>	<b>99.206</b>	<b>99.21</b>
FC-8	98.72	99.5	98.75	98.732	98.73

\*Bold values indicate the maximum metrics of this study

**Fig. 9** Confusion matrix of the proposed CNN–SVM**Table 13** Comparison of all methods in terms of performance metrics when using CWT

Classification algorithm	Sensitivity (%)	Specificity (%)	Precision (%)	<i>F1</i> -score (%)	Test accuracy rate (%)
Proposed CNN	96.64	98.38	96.54	96.53	96.53
SVM	85.56	93.68	85.56	85.51	85.56
Proposed CNN–SVM	99.206	99.66	99.213	99.206	99.21

our proposed CNN, such as the PTB ECG dataset, CT images of brain hemorrhages, and the CIFAR-10 dataset. As is known, the pre-trained CNN architectures are tested on the traditional dataset. In addition, the proposed CNN architecture is also tested on the CIFAR-10 dataset in this study and examined whether it could make a successful classification. The CIFAR-10 dataset consists of 10 classes and 60,000 images. Similarly, this huge dataset is split 80% for training and 20% for testing, as shown in the study. In this way, 50,000 images are trained and 10,000 images are also tested. Also, the same option parameters are applied to both sets of data. Table 14 shows the proposed CNN success on different datasets. In addition, Fig. 10 displays the confusion matrix for the CIFAR-10 dataset.

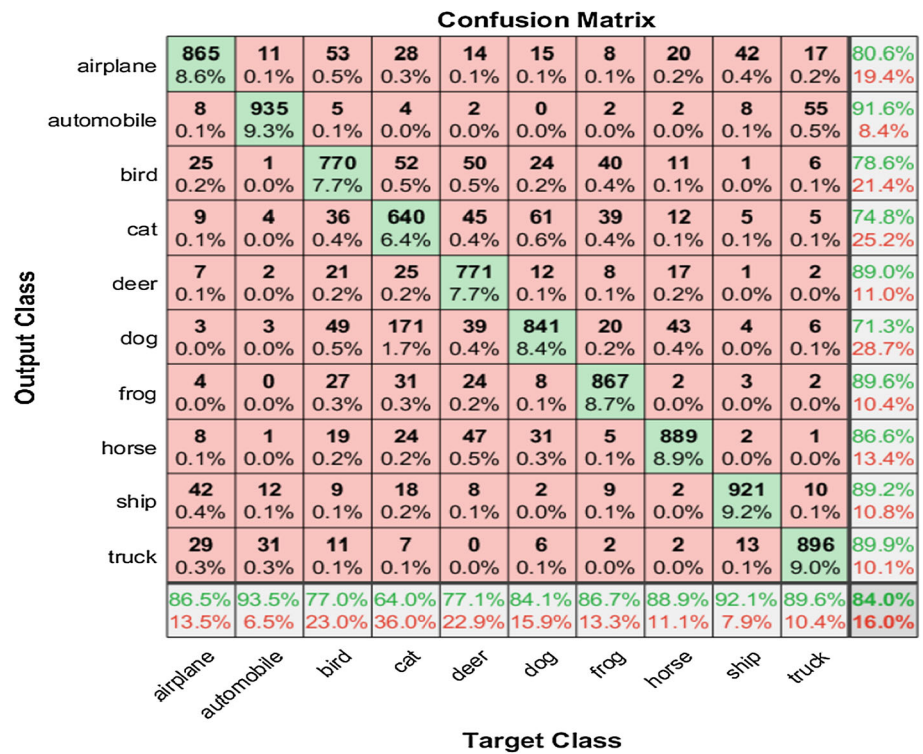
As can be seen, the performance of the proposed CNN is very good. However, as mentioned earlier, this CNN must be excellent for classifying biomedical signals or images. Therefore, the proposed CNN is merged with SVM for perfect classification. In general, if a CNN architecture has a fully connected layer, that layer is used for obtaining features and combined with SVM. Of course, this method offers good advantages because of the extracted features. However, the deep learning algorithm (also CNN) is a complex nonlinear model and is referred to as a black box (Guidotti et al. 2018). Accordingly, it has to be investigated which last layers have good properties within this probabilistic process. Among all these considerations, the characteristics in the Max-Pooling 7 (just before the FC-8 layer) are also examined in the present study. According to



**Table 14** The proposed CNN performance on different datasets

Datasets	Number of Class	Sensitivity (%)	Specificity (%)	Precision (%)	F1-score (%)	Accuracy rate (%)
PTB ECG dataset (Özaltın and Yeniay 2021)	2	96.42	94.96	95	95.56	95.6
CIFAR-10	10	83.95	98.22	84.10	83.87	84
ECG dataset in this Study	3	96.64	98.38	96.54	96.53	96.53
CT images of brain hemorrhage (Ozaltin et al. 2022)	4	91.95	94.93	93.17	92.53	92.85

**Fig. 10** Confusion matrix of the proposed CNN for the CIFAR-10 dataset



the knowledge gained in this study, it is necessary to examine the features in the last layers for a more sensitive analysis, which are listed in Table 12. Apart from this, when the literature is searched on the same property ECG dataset, the proposed CNN-SVM hits the top in terms of accuracy rate, detailed in Table 14.

### 5 Conclusion

Many of sudden deaths from heart disease continue to increase these days with the coronavirus (COVID-19). Based on this, the automatic classification of the signals received from the heart is of great importance for diagnosis and treatment. In this study, we classify ECG types using our proposed CNN, which has overcome overfitting with

the dropout layer. This CNN is also performed on other datasets, shown in Table 14. In addition, the proposed CNN is compared to AlexNet and SqueezeNet on nine different ECG image datasets processed via CWT using three different wavelet functions and three different sample lengths. All results show that the best sample length is 500 Hz and the best mother wavelet function is “Amor.” Also, the comparison of classification success in terms of the overall accuracy rate of the proposed CNN, AlexNet, and SqueezeNet is 98%, 94.67%, and 94.67%, respectively. Therefore, the proposed CNN architecture performs the best classification on the ECG image dataset generated with the Amor wavelet function and the 500 Hz sample length by using CWT. However, we want to search how another pre-processing method affects classification success and so, we generate new ECG images using STFT with 500 Hz

**Table 15** The comparison of classification performances for different studies on ECG signals

Study	Pre-processing method	Algorithm	Accuracy (%)
Çinar and Tuncer (2021)	STFT	CNN (AlexNet-SVM)	96.77
Eltras et al. (2021)	CQ-NSGT*	CNN (AlexNet)	98.82
Gaddam et al. (2021)	CWT	CNN (AlexNet)	95.67
Golgowski and Osowski (2020)	CWT	CNN	82.06
	DWT	Extra random forests	97.78
Krak et al. (2020)	CWT	CNN	96
Krishnakumar et al. (2021)	CWT	CNN (GoogleNet)	96.88
Kumari et al. (2020)	DWT	SVM	95.92
Nahak and Saha (2020)	RR	SVM	86.77
	Wavelet with AR	SVM	92.22
	Fusion of features	SVM	93.33
Olanrewaju et al. (2021)	CWT	CNN (AlexNet)	98.7
Rahuja and Valluru (2021)	CWT	CNN (AlexNet)	97.3
Proposed CNN	CWT	CNN	96.53
Proposed CNN-SVM	CWT	CNN-SVM	99.21

\*Constant-Q non-stationary Gabor transform

sample length. In this way, we use not only a splitting method as training and testing (80:20, 70:30), but also cross-validation implemented on two created datasets. According to the ECG image dataset generating via CWT, when the dataset split training and testing as 80:20, all mean performance metrics are over 96.5%, and also maximum standard deviation (Std) is 0.0173 on testing the ECG dataset. When the dataset split training and testing as 70:30, all average performance metrics are over 95.3%, and the highest Std is 0.01705. Further, as fivefold and tenfold cross-validation methods are implemented on the dataset, average accuracies are 96.44% and 97.22%, respectively. Also, the maximum average accuracy of 97.22% is obtained through tenfold cross-validation. Resulting of the use of CWT, the performances expressed that cross-validation is better than training and testing. According to the ECG image dataset creating via STFT, when the dataset split training and testing as 80:20, all average performance metrics are above 89.3% and also the maximum Std is 0.013438. While the dataset split training and testing as 70:30, all mean performance metrics are above 89.2% and also the maximum Std is 0.0169. Besides, when fivefold and tenfold cross-validation methods are applied on the dataset, average accuracies are 91.11% and 87.66%, respectively. All these results show that CWT is better than STFT to detect types of ECG.

The main purpose of the study is to find an excellent classification algorithm for recognizing the ECG types. Therefore, the proposed CNN is merged with SVM. In this stage of the study, the proposed CNN is employed as a

deep feature extractor from ECG images generated with CWT. In general, if any CNN architecture has a fully connected layer, it is used for obtaining features. It is highlighted that it can provide an advantage to examine features from the last layers of CNN, such as the max-pooling layer, in this study. To improve the proposed CNN performance, Max-Pooling 7 and FC-8 layers are used attaining reduced features, and the results are detailed in Table 12. As a result, the highest success with an accuracy of 99.21% is achieved by Max-Pooling 7 layer. When comparing to other studies on similar ECG datasets, the proposed CNN-SVM is considered the best performing for classification, detailed in Table 15.

This study applies deep learning algorithms for ECG-type detection as an assisting decision support system. As such, clinicians will not spend much more time identifying ECG types, and the proposed pipeline will help physicians and professionals better identify ECG types in a hospital setting. In future work, we will continue to search for the detection of various diseases on signals or images by deep learning algorithms.

**Acknowledgements** This study is based on Oznur Ozaltin's Ph.D. thesis and supervised by Ozgur Yeniay.

**Author contributions** OO idealized this study and analyzed the data. OY supervised the research and approved the final draft.

**Funding** The authors took on no certain funding for this study.

**Data availability** Data can be available from <https://www.physionet.org/>.

## Declarations

**Conflict of interest** The authors announced that they had no conflicts of interest to report related to this study.

## References

- Abdelmalek B, Ahmed K, Amine TM (2019) Lightweight CNNs-Based Object Detection for Embedded Systems implementation. In: 1st Conference on innovative trends in computer science (CITCS'2019)
- Acharya UR, Fujita H, Oh SL, Hagiwara Y, Tan JH, Adam M (2017) Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals. *Inf Sci* 415:190–198
- Al Rahhal MM, Bazi Y, Al Zuairi M, Othman E, BenJdira B (2018) Convolutional neural networks for electrocardiogram classification. *J Med Biol Eng* 38(6):1014–1025
- Alickovic E, Subasi A (2015) Effect of multiscale PCA de-noising in ECG beat classification for diagnosis of cardiovascular diseases. *Circuits Syst Signal Process* 34(2):513–533
- Baim DS et al (1986) Survival of patients with severe congestive heart failure treated with oral milrinone. *J Am Coll Cardiol* 7(3):661–670
- Baloglu UB, Talo M, Yildirim O, San Tan R, Acharya UR (2019) Classification of myocardial infarction with multi-lead ECG signals and deep CNN. *Pattern Recognit Lett* 122:23–30
- Ceylan R, Özbay Y (2007) Comparison of FCM, PCA and WT techniques for classification ECG arrhythmias using artificial neural network. *Expert Syst Appl* 33(2):286–295
- Chen C, Hua Z, Zhang R, Liu G, Wen W (2020) Automated arrhythmia classification based on a combination network of CNN and LSTM. *Biomed Signal Process Control* 57:101819
- Chen C-Y et al (2022) Automated ECG classification based on 1D deep learning network. *Methods* 202:127–135
- Cheng X et al (2022) Efficient hardware design of a deep U-net model for pixel-level ECG classification in healthcare device. *Microelectron J* 126:105492
- Çınar A, Tuncer SA (2021) Classification of normal sinus rhythm, abnormal arrhythmia and congestive heart failure ECG signals using LSTM and hybrid CNN-SVM deep neural networks. *Comput Methods Biomech Biomed Engin* 24(2):203–214
- Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20(3):273–297
- Elleuch M, Maalej R, Kherallah M (2016) A new design based-SVM of the CNN classifier architecture with dropout for offline Arabic handwritten recognition. *Procedia Comput Sci* 80:1712–1723
- Eltrass AS, Tayel MB, Ammar AI (2021) A new automated CNN deep learning approach for identification of ECG congestive heart failure and arrhythmia using constant-Q non-stationary Gabor transform. *Biomed Signal Process Control* 65:102326
- Eltrass AS, Tayel MB, Ammar AI (2022) Automated ECG multi-class classification system based on combining deep learning features with HRV and ECG measures. *Neural Comput Appl* 34(11):8755–8775
- Gaddam PG, Sreehari R (2021) Automatic classification of cardiac arrhythmias based on ECG signals using transferred deep learning convolution neural network. *J Phys Conf Ser* 2089(1):012058
- Goldberger AL et al (2000) PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation* 101(23):e215–e220
- Golowski M, Osowski S (2020) Anomaly detection in ECG using wavelet transformation. In: IEEE 21st International Conference on Computational Problems of Electrical Engineering (CPEE), 2020. IEEE, pp 1–4
- Guidotti R, Monreale A, Ruggieri S, Turini F, Giannotti F, Pedreschi D (2018) A survey of methods for explaining black box models. *ACM Comput Surv (CSUR)* 51(5):1–42
- Haberl R, Jilge G, Pulter R, Steinbeck G (1989) Spectral mapping of the electrocardiogram with Fourier transform for identification of patients with sustained ventricular tachycardia and coronary artery disease. *Eur Heart J* 10(4):316–322
- Haykin S, Veen BV (1999) *Signals and Systems*. Wiley, Hoboken
- Hua X et al (2020) A novel method for ECG signal classification via one-dimensional convolutional neural network. *Multimed Syst* 1:1–13
- 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
- Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K (2016) SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size. arXiv preprint [arXiv:1602.07360](https://arxiv.org/abs/1602.07360)
- Jiao Y, Qi H, Wu J (2022) Capsule network assisted electrocardiogram classification model for smart healthcare. *Biocybern Biomed Eng* 42(2):543–555
- Khorrami H, Moavenian M (2010) A comparative study of DWT, CWT and DCT transformations in ECG arrhythmias classification. *Expert Syst Appl* 37(8):5751–5757
- Kiranyaz S, Ince T, Gabbouj M (2015) Real-time patient-specific ECG classification by 1-D convolutional neural networks. *IEEE Trans Biomed Eng* 63(3):664–675
- Koc M, Sut SK, Serhatlioglu I, Baygin M, Tuncer T (2022) Automatic prostate cancer detection model based on ensemble VGGNet feature generation and NCA feature selection using magnetic resonance images. *Multimed Tools Appl* 81(5):7125–7144
- Koklu M, Ozkan IA (2020) Multiclass classification of dry beans using computer vision and machine learning techniques. *Comput Electron Agric* 174:105507
- Krak I, Stelia O, Pashko A, Efremov M, Khorozov O (2020) Electrocardiogram classification using wavelet transformations. In: 2020 IEEE 15th International conference on advanced trends in radioelectronics, telecommunications and computer engineering (TCSET). IEEE, pp 930–933
- Krishnakumar S, Yasodha M, Priyadharshini JV, Janney JB, Divakaran S, Christy VL (2021) Detection of arrhythmia and congestive heart failure through classification of ECG signals using deep learning neural network. In: 2021 International conference on advancements in electrical, electronics, communication, computing and automation (ICAECA), 2021. IEEE, pp 1–7
- Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst* 25:1097–1105
- Kumar A, Kumar S, Dutt V, Dubey AK, García-Díaz V (2022) IoT-based ECG monitoring for arrhythmia classification using Coyote Grey Wolf optimization-based deep learning CNN classifier. *Biomed Signal Process Control* 76:103638
- Kumari CU, et al (2020) Heart rhythm abnormality detection and classification using machine learning technique. In: 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI), (48184). IEEE, pp 580–584

- Lee HK, Choi Y-S (2019) Application of continuous wavelet transform and convolutional neural network in decoding motor imagery brain-computer interface. *Entropy* 21(12):1199
- Li C, Zheng C, Tai C (1995) Detection of ECG characteristic points using wavelet transforms. *IEEE Trans Biomed Eng* 42(1):21–28
- Li J, Si Y, Xu T, Jiang S (2018) Deep convolutional neural network based ECG classification system using information fusion and one-hot encoding techniques. *Math Problems Eng* 2018:1
- Mahmud T, Fattah SA, Saqib M (2020) Deeparnet: An efficient deep cnn architecture for automatic arrhythmia detection and classification from denoised ecg beats. *IEEE Access* 8:104788–104800
- Meng L, Tan W, Ma J, Wang R, Yin X, Zhang Y (2022) Enhancing dynamic ECG heartbeat classification with lightweight transformer model. *Artif Intell Med* 124:102236
- Moody GB, Mark RG (2001) The impact of the MIT-BIH arrhythmia database. *IEEE Eng Med Biol Mag* 20(3):45–50
- Nahak S, Saha G (2020) A fusion based classification of normal, arrhythmia and congestive heart failure in ECG. In: 2020 National conference on communications (NCC). IEEE, pp 1–6
- Narin A (2020) Detection of focal and non-focal epileptic seizure using continuous wavelet transform-based scalogram images and pre-trained deep neural networks. *IRBM*
- Olanrewaju RF, Ibrahim SN, Asnawi AL, Altaf H (2021) Classification of ECG signals for detection of arrhythmia and congestive heart failure based on continuous wavelet transform and deep neural networks. *Indones J Electr Eng Comput Sci* 22(3):1520–1528
- Ozaltin O, Coskun O, Yeniay O, Subasi A (2022) Classification of brain hemorrhage computed tomography images using OzNet hybrid algorithm. *Int J Imag Syst Technol* 2022:1. <https://doi.org/10.1002/ima.22806>
- Özaltın Ö, Yeniay Ö (2021) Ecg classification performing feature extraction automatically using a hybrid CNN–SVM algorithm. In: 2021 3rd International congress on human-computer interaction, optimization and robotic applications (HORA). IEEE, pp 1–5
- Physionet (2020). <https://www.physionet.org/>. Accessed 01 June 06 14:00
- Pałczyński K, Śmigiel S, Ledziński D, Bujnowski S (2022) Study of the few-shot learning for ECG classification based on the PTB-XL dataset. *Sensors* 22(3):904
- Plawiak P, Acharya UR (2020) Novel deep genetic ensemble of classifiers for arrhythmia detection using ECG signals. *Neural Comput Appl* 32(15):11137–11161
- Qaisar SM, Subasi A (2020) Cloud-based ECG monitoring using event-driven ECG acquisition and machine learning techniques. *Phys Eng Sci Med* 43(2):623–634
- Rahuja N, Valluru SK (2021) A deep neural network approach to automatic multi-class classification of electrocardiogram signals. In: 2021 International Conference on Intelligent Technologies (CONIT). IEEE, pp 1–4
- Rajkumar A, Ganesan M, Lavanya R (2019) Arrhythmia classification on ECG using Deep Learning. In: 2019 5th international conference on advanced computing & communication systems (ICACCS). IEEE, pp 365–369
- Salem M, Taheri S, Yuan JS (2018) ECG arrhythmia classification using transfer learning from 2-dimensional deep CNN features. In: 2018 IEEE biomedical circuits and systems conference (BioCAS). IEEE, pp 1–4
- Sandeep K, Kora P, Swaraja K, Meenakshi K, Pampana L (2019) ECG classification using machine learning. *Int J Recent Technol Eng (IJRTE)* 8(4):2492–2494
- Sepahvand M, Abdali-Mohammadi F (2022) A novel method for reducing arrhythmia classification from 12-lead ECG signals to single-lead ECG with minimal loss of accuracy through teacher-student knowledge distillation. *Inf Sci* 593:64–77
- Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res* 15(1):1929–1958
- Toma TI, Choi S (2022) A comparative analysis of 2D deep CNN models for arrhythmia detection using STFT-based long duration ECG spectrogram. In: 2022 13th International conference on ubiquitous and future networks (ICUFN). IEEE, pp 483–488
- Tuncer T, Dogan S, Plawiak P, Subasi A (2022) A novel discrete wavelet-concatenated mesh tree and ternary chess pattern based ECG signal recognition method. *Biomed Signal Process Control* 72:103331
- Xing Y et al (2022) Accurate ECG classification based on spiking neural network and attentional mechanism for real-time implementation on personal portable devices. *Electronics* 11(12):1889
- Xu X, Liu H (2020) ECG heartbeat classification using convolutional neural networks. *IEEE Access* 8:8614–8619
- Zhao Y, Cheng J, Zhang P, Peng X (2020) ECG classification using deep CNN improved by wavelet transform. *Comput Mater Continua* 64(3):1615–1628

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.