

Figure 1. Process of computational diagnostic techniques for electrocardiogram signals.

2. Data Preprocessing

2.1. Noise in Electrocardiogram (ECG)

Normal ECG signals are time-varying signals with a small amplitude ranging from 10 μ V to 5 mV. Their typical value is 1 mV and their frequencies range from 0.05–100 Hz, mainly concentrated in the 0.05–35 Hz range. Most ECG analysis systems require relatively noiseless ECG signals to achieve accurate and reliable CVDs diagnosis [18]. However, in practice ECG signals are often disturbed by various noises and artifacts, including baseline drift, electrode contact noise, power-line interference (PLI), and electromyographic (EMG) noise, which lead to ECG distortion and affect feature extraction [19–21].

Patient movements, poor electrode contact, and changes in electrode skin impedance cause baseline wander and abrupt drift noises. During Holter monitoring, ECG monitoring in moving ambulance or exercise, baseline wander plays a significant role in noises [22]. The frequency range of the motion noise is usually 1–10 Hz. The amplitude of baseline wander caused by respiration varies by ~15% of peak-to-peak ECG amplitude at frequencies ranging from 0.15 Hz to 0.3 Hz [19]. It is noted that the severe baseline wanders or motion artifacts can distort the ST-segment and other low-frequency components of ECG signals. The ST segment distortion may lead to the wrong diagnosis of myocardial infarction, Brugada syndrome, and other ST-segment related abnormalities.

Power-line interference (PLI) (50/60 Hz) is mainly contributed by inductive and capacitive couplings of ubiquitous power lines in the ECG signal acquisition circuitry [20]. The lower frequency noise components of the PLI are mixed with the frequency content of the ECG signal. The severely structured noises can distort the morphological features such as amplitude, duration, and shape of low-amplitude local waves of the ECG signal. In particular, the P-wave distortions can lead to the wrong diagnosis of atrial arrhythmias such as atrial enlargement and fibrillation [23]. Proper shielding, avoiding loose connection in wires and appropriate placement of electrodes can minimize the PLI.

EMG noise is contributed by the electrical activity of muscles during periods of contraction. The frequency distribution of EMG noise is basically within 0 Hz to 500 Hz and mainly concentrates in the range from 50–150 Hz. EMG noise with a high frequency above 100 Hz can be removed by a low-pass filter with an appropriate cut-off frequency. Previous studies demonstrated that EMG artifacts significantly altered the shapes of local waves of the ECG signal because the frequency of the noise was considerably superimposed with that of the ECG signals in the range of 0.05–100 Hz.

abnormal T waves (e.g., inverted T waves) by chest leads on patients with pulmonary embolism (PE). Marcinkevics. R. et al. found that the T-wave amplitude of patients with arrhythmia right ventricular dysplasia (ARVD) was significantly different from that of normal patients [43]. U-wave is the last unsteady and smallest wave in the ECG, which shows a circular upward deflection. Sometimes, U-wave may not be observed because of its small size. The formation of U wave is controversial. Generally, U-wave is thought to represent repolarization of the Purkinje fibers. Usually, U wave has the same polarity as T wave. In clinical diagnosis, transient U-wave inversion can be caused by local myocardial ischemia or hypertension [44].

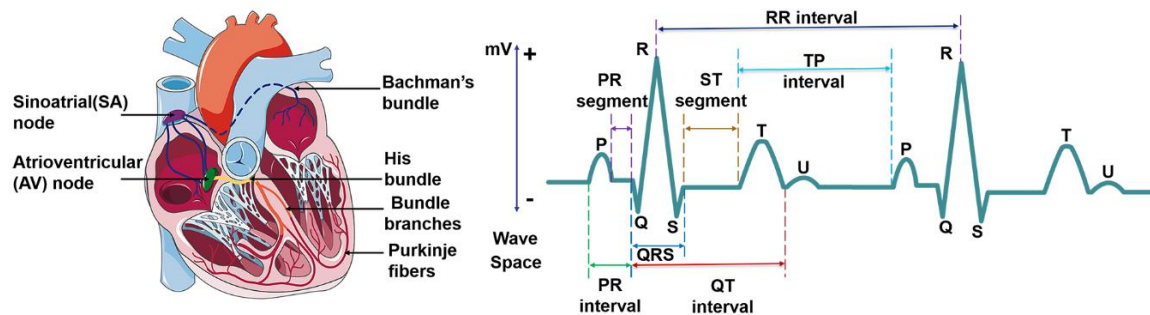


Figure 2. Cardiac electrical conduction system and the electrocardiogram signal.

Table 1. ECG features and the normal values for a healthy adult.

Features	Description	Amplitude	Duration	Disease Diagnosis	References
R-R interval	The interval between two successive R-waves of the QRS complex ventricular rate		0.6–1.2 s	Paroxysmal atrial fibrillation Congestive heart failure	[45,46]
P wave	Atrial depolarization	0.25 mV	0.08–0.11 s	Atrial fibrillation Atrial hypertrophy	[47]
P-R interval	The time between the onset of atrial depolarization and the onset of ventricular depolarization		0.12–0.2 s	Stroke	[41]
QRS complex	Ventricular depolarization	1.60 mV for R peak	0.06–0.1 s	Ventricular enlargement Heart failure Tachycardia Acute Coronary Syndrome	[48–50]
ST-segment	The interval between ventricular depolarization and repolarization		0.05–0.155 s	Myocardial ischemia or infarction	[51]
T wave	Ventricular repolarization	0.1–0.8 mV	0.05–0.25 s	Myocardial infarction Pulmonary embolism	[46,52,53]
U wave	The last phase of ventricular repolarization	May not be observed because of its small size	Unknown	Unknown	[44]
QT interval	The time is taken for ventricular depolarisation and repolarisation		0.35–0.44 s	Hypokalemia ventricular arrhythmias	[54]

Segments and intervals of the ECG signal reflect each stage and cardiac cycle of heart contraction, which should be completed within a specific period for a healthy person. The abnormal period indicates something wrong with the heart [55]. R-R interval is the time elapsed between two consecutive R waves of the QRS signal. It is usually employed to assess ventricular rate. During sinus rhythm, patients with short-term risk of paroxysmal atrial fibrillation (PAF) tend to show higher RR interval variability [45]. Q-T interval refers to the time from the beginning of QRS to the end of T wave, representing the total time required for ventricular depolarization and repolarization. It is worth noticing that P-R interval is

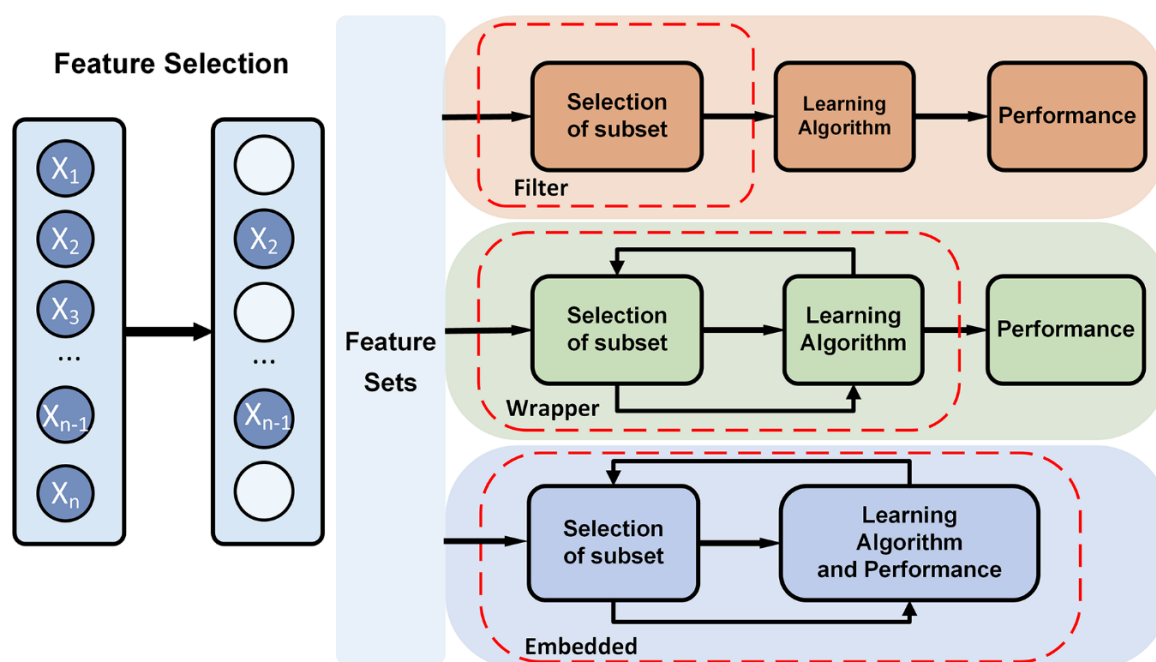


Figure 3. Feature selection methods, such as filter, wrapper, and embedded method.

Filter-based feature selection applies a selected metric to find irrelevant attributes and filters out the redundant data [95]. The selection process is independent of the training process. Filter-based methods rank the features as a pre-processing step before the learning algorithm, and select those features with high ranking scores. The score is computed by measuring the variance between the expected value of the information and the observed value. The evaluation metric of filter usually is used to analyze the internal features of the feature subset, which includes correlation, distance, information gain, and so on. In practice, filter-based feature selection can be initially screened by expert knowledge, then filtered by filtering methods. The characteristics that have been proven to be relevant to a particular disease or physiological response are often directly selected in feature selection. Besides, it is a common method to calculate a score for each feature column. Columns with poor feature selection scores are ignored. Filter-based feature selection provides a variety of performance criteria for assessing the information value, such as correlation coefficient, **mutual information**, **Kendall correlation**, **Spearman correlation**, **Chi Squared**, **Fisher score**, Laplacian score, Trace Ratio criterion, among which, **Fisher score is widely used metrics for supervised feature selection**. Fahim Sufi et al. ranked feature subsets according to a correlation-based heuristic evaluation function [96]. The algorithm selected features by calculating mean feature-class correlation and the average feature-feature intercorrelation. The two criteria ensure that irrelevant features and redundant features are removed from the attributes set, because they are not correlated with the class or other features. Some researchers evaluated each feature from the primary feature sets by Fisher score [97,98]. The fisher score selected each ECG feature independently according to their scores under the Fisher criterion, resulting in suboptimal subset of attributes. It is often used to select feature sets with lower dimension. The Filter method uses statistical indicators to score and filter each feature, focusing on the characteristics of the data itself. The advantage of the Filter method is that the calculation is fast and does not depend on a specific model. However, the final accuracy of the classification may be not high because the selected features are not customized for the specific model.

Wrapper-based feature selection utilizes a predefined classifier to evaluate the feature set. This method scores the features using the learning algorithm that will ultimately be employed in classification. The feature selection process is integrated with training process, and the prediction ability of the model is used as the selection criterion to evaluate the feature subset, such as classification

accuracy, complexity penalty factor. Forward and backward selection algorithm in multiple linear regression is a simple implementation of wrapper. **Sequential floating forward search (SFFS) algorithm utilizes sequential forward selection (SFS) and sequential backward selection (SBS)** in sequence to obtain the best ECG feature set. Llamedo and Martinez obtained an optimal ECG feature set containing eight features by SFFS [99]. The SFFS method is suitable for small-and medium-scale data [100]. KNN [101] and SVM [102] can be used as evaluation functions of the wrapper. Compared with the filter, the wrapper has better performance in generating high-quality subsets, but the data processing is computationally expensive since the learner needs to be trained many times during the feature selection process. Unlike filter selection, which does not consider subsequent classification algorithms, wrapped selection directly takes the performance of the final classification algorithms as the evaluation standard of the feature subset. In other words, wrapped feature selection is to select the most favorable feature subset for a given learning algorithm. However, the performance of the subset of features is affected by the particular learning algorithm. The stability and adaptability of the feature subset are poor because each additional feature must be constructed feature subset for evaluation. Wrapper-based feature selection has high time complexity and is not suitable for high dimensional data set.

Embedded feature selection is built into the construction of the machine learning algorithm. It provides a trade-off solution between filter method and wrapper method, which can solve the high redundancy of the filter algorithm and the computational complexity of the wrapper algorithm. The embedded feature selection is automatically performed during the learner training process [96]. Compared to the other two methods, the searching and selection process of features' subset is built into classifier construction. Regularization and tree-based methods are widely used in embedded methods. The regularization models in form of $\ell_{2,1}$ -norm regularized regression models, such as Lasso, sparse linear discriminant analysis, and regularized support vector machine, are widely used in embedded methods [103]. Regularization is to impose additional constraints or penalties on the loss function when training a neural network, which can reduce the complexity and instability of the model in the learning process, thus avoiding overfitting and improving generalization ability. Decision tree is a classic embedded feature selection method, such as ID3, C4.5, CART algorithm. Features with good ability of classification are selected in the nodes of the tree, and then the selected feature subsets are used to perform the learning tasks. Feature subsets are selected during the process of decision tree generation. The random forest has the advantages of high accuracy, good robustness and easy to use, which makes it one of the most popular machine learning algorithms. The random forest provides two methods of feature selection, including mean decrease impurity and mean decrease accuracy. Tree-based prediction models can be used to calculate the importance of features, and thus to remove irrelevant features. Embedded feature selection can be applied to high dimensional data sets, but the design of the embedded method is tightly coupled with a specific learning algorithm, which in turn limits its application to other learning algorithms.

3.2.2. Feature Extraction

The disadvantage of feature selection is that unselected features are simply moved out, which will reduce the accuracy and efficiency of learning algorithms. Feature extraction considers all features and maps the useful information into a low-dimensional feature space, which is more commonly used in the selection of feature sets with insufficient prior knowledge and high-dimensionality. By choosing an appropriate dimension reduction method, the invalid information of the original feature set can be removed and the effective information of the original feature set is retained to the greatest extent. Typical dimensionality reduction methods include principal component analysis (PCA), linear discriminant analysis (LDA), independent component analysis (ICA), and generalized discriminant analysis (GDA) [104].

The PCA method is a linear dimensionality reduction method that maps the original features into a low-dimensional space while retaining the variance. PCA is the most widely used form of dimensionality reduction, which preserves the maximum amount of variance of the original data.

network takes raw data as input, which is much more robust than feature engineering. Acharya et al. proposed a 11-layer deep CNN for diagnosis of myocardial infarction based on ECG data with and without noise, and achieved accuracy of 93.53% and 95.22%, respectively [130]. The experiment showed that the presence or absence of noise hardly affected the accuracy of the end-to-end neural network, with a difference of less than 2%, which reflects the amazing robustness of the end-to-end neural network. Besides, the end-to-end model also solves some problems that traditional analysis methods are difficult to solve. Traditional analysis methods usually analyze data of a specified length, such as single-beat ECG data, which leads to the loss of time domain information and is not conducive to adjustment and modification. Compared with traditional analysis methods, the End-to-End model can input long-term data at a time, and the algorithm barely changes, which is beneficial to expansion and modification. Acharya et al. developed a CNN-based neural network for detecting arrhythmia. The network supported input of 2 s or 5 s of ECG data, achieving accuracy of 92.50% and 94.90%, respectively [131]. This showed that long-term ECG data can improve the effectiveness of the classifier to a certain extent. The end-to-end model can also be used to solve individual automatic optimization problems. In the method based on feature engineering, the automatic optimization of individuals is one of the most difficult tasks, because the distribution of training data and the strong offset between patients leads to poor performance of the model in practice [86,132,133].

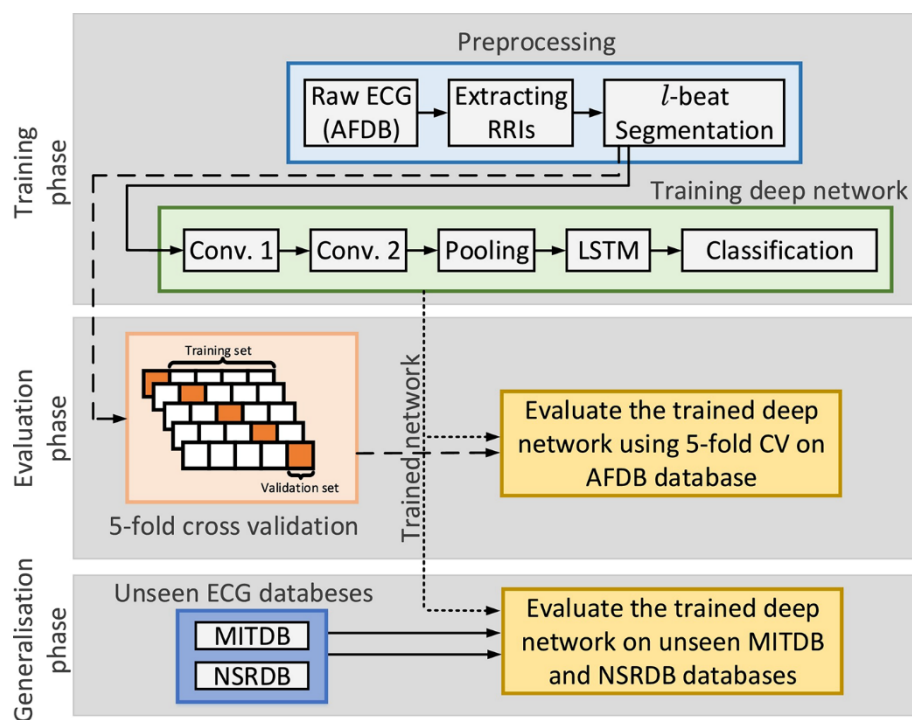


Figure 5. Flowchart of the Convolutional- and Recurrent-Neural Networks (reproduced with permission from the authors of [129]). The model consists of a training phase for estimation of the optimal parameters of model, an evaluation phase for validating performance measures and a generalization phase to report performance on previously unseen data sets.

Table 2. ECG analysis with end-to-end approaches.

Tasks	Database	Model	Signal	Performance (%)	References
AF detection	MIT-BIH	CNN and RNN	250 samples	Acc = 97.10 Sen = 98.98 Spe = 96.95	[129]
Myocardial infarction detection	PTB	CNN	651 samples	Acc = 93.53 Sen = 93.71 Spe = 92.83	[130]
CVD detection	INCART	1D-CapsNet	514 samples	Acc = 99.44 Sen = 99.7 Spe = 98.1	[134]
MI classification	PTB	CNN	651 samples	Acc = 99.78 Sen:above 99 Spe:above 99	[135]
Arrhythmia detection	MIT-BIH	CNN	500 samples	Acc = 92.50 Sen = 98.09 Spe = 93.13	[131]
Classification of ECG signal	MIT-BIH	DNN	300 samples	Acc = 98.6 Sen = 92.4 Spe = 99.29	[132]
Classification of ECG signal	MIT-BIH	1D-CNN	128 samples	Acc = 99 Sen = 93.9 Spe = 98.9	[124]
Classification of ECG signal	A synthetic dataset by using an ECG simulator	Short-Time Fourier Transform and CNN	2426 samples	Acc = 99.2	[136]
AF detection	IEEE-TBME	CNN	512 samples	Acc = 96.1 Sen = 97.0 Spe = 100.0	[137]
Classification of ECG signal	TNMG	DNN	2800 samples	Spe:above 99 AUC:above 97	[138]
AF detection	iRhythm Technologies	DNN	256 samples	Sen:above 90 Spe:above 90 AUC = 0.999	[14]
Classification of ECG signal	MIT-BIH	DNN	360 samples	Acc = 98.97 Sen = 97.68 Spe = 99.89	[139]
AF detection	MIT-BIH	CNN	360 samples	Acc = 99.45 Sen = 99.29 Acc = 95.49	[140]
MI detection	PTB	CNN	800 samples	Sen = 94.85 Spe = 97.37 Acc = 91.33	[141]
AF detection	MIT-BIH	CNN	3600 samples	Sen = 83.91 Spe:above 99	[142]

Accuracy(Acc), Sensitivity(Sen), Specificity(Spe), Area Under Curve(AUC), MIT-BIH arrhythmia Database (MIT-BIH), Physikalisch-Technische Bundesanstalt diagnostic ECG database (PTB), St.-Petersburg Institute of Cardiological Technics database (INCART), IEEE-TBME PPG Respiratory Rate Benchmark data set (IEEE-TBME). Telehealth Network of Minas Gerais (TNMG).

Carrera et al. provided an end-to-end model that can automatically adjust each user's parameters by using fine-tuning techniques in deep learning to make the model perform superior on individual patients [143]. Once a dedicated model is trained for a specific patient, it can be potentially used alone to quickly and accurately classify a long ECG data stream, which is greatly beneficial for wearable ECG monitoring. Attia et al. trained CNN networks based on ECG data for identifying patients with asymptomatic left ventricular dysfunction (ALVD). The network model not only screened out ALVD, but also found patients with a 4-fold risk of developing ALVD in the future [144], which provides users with a low-cost, non-invasive ALVD screening tool.

End-to-end neural networks also have some disadvantages. First, because there is no prior knowledge to support, end-to-end neural networks rely on high-quality data sets compared to classic machine learning methods. A larger and more comprehensive data set will help the neural network to fit well. Besides, neural networks with deep architectures, which involve a large number of parameters, easily overfit the training data. Hence, certain training skills are required in the training process. Third, the model becomes a black box, which reduces the interpretability of the neural network. The model

design can only be modified by model structure and training methods to obtain better results. However, because of its powerful functions, end-to-end neural networks will still be one of the main research directions of ECG analysis.

5. Electrocardiogram (ECG) Databases

A wide variety of standard ECG databases are available online for assessment of computational algorithms in ECG signal analysis [2,3], such as Massachusetts Institute of Technology-Beth Israel Hospital (MIT-BIH) arrhythmia database, Physikalisch-Technische Bundesanstalt diagnostic ECG database (PTB), American Heart Association (AHA) database, St.-Petersburg Institute of Cardiological Technics Database (INCART), University of Toronto Database (UofTDB), Fantasia Database, and so on. MIT-BIH databases provided by MIT consists of several subsets, such as MIT-BIH ST Change database, MIT-BIH atrial fibrillation database, MIT-BIH arrhythmia, MIT-BIH long term database and MIT-BIH supraventricular arrhythmia. The MIT-BIH arrhythmia is well-known and most widely used by researchers. The database was taken from 47 subjects at a sampling rate of 360 samples per second with 11-bit resolution. Each segment is half an hour long. Table 3 summarizes the main public databases available for ECG analysis in detail. Different methods (e.g., sampling frequency, sample number, and leads) for collecting ECG data are utilized by these databases. However, the limited number of public ECG databases are still the main limitation in ECG analysis. Setting up standard and new publicly available databases that contain a large number of patients with labels acquired using the latest medical guidelines is urgent for the development of computational algorithms for ECG based diagnosis.

Table 3. Brief description of ECG databases [2,3].

Database	Subjects	Records	Duration (min)	Frequency (Hz)	Leads	Resolution (bit)
MIT-BIH Arrhythmia	47	48	30	360	12	11
MIT-BIH AF	25	25	10 h	250	2	12
MIT-BIH ST Change	28	28	13–67	360	1–2	N/A
MIT-BIH Long Term	7	7	14–22h	128	2lead:12 1lead:10	2lead:6 1lead:3
MIT-BIH SUPRA	N/A	78	30	128	10	2
PTB	290	549	N/A	1k	12 + 3 Frank-lead	16
AHA	N/A	10	30	250	2	12
INCART	32	75	30	257	N/A	12
UofTDB	1020	1020+	2 to 5	200	1	12
Fantasia	40	40	120	250	N/A	N/A

6. Applications

ECG records the electrical signals of myocardial cells in each cardiac cycle, which contains a large amount of information related to the heart. Hence, it is widely used in the field of medical healthcare, such as diagnosis of heart disease, prediction of cardiovascular disease, stress and sleep monitoring, and wearable device based online health monitoring.

6.1. Disease Diagnosis

Coronary heart disease is the leading cause of death. The world health organization classifies coronary heart disease into five categories, including asymptomatic myocardial ischemia, angina, myocardial infarction, ischemic heart failure, and sudden death. The main cause of coronary heart disease is coronary atherosclerosis, which is an inflammatory response to chronic multifactorial injury of the vessel walls, resulting in the formation of atherosclerotic plaques. These deposits cause the surface