



HAL
open science

Detection of Heart Failure Using a Convolutional Neural Network via ECG Signals

Jad Botros, Farah Mourad-Chehade, David Laplanche

► **To cite this version:**

Jad Botros, Farah Mourad-Chehade, David Laplanche. Detection of Heart Failure Using a Convolutional Neural Network via ECG Signals. 2022 15th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Nov 2022, Beijing, China. pp.1-5, 10.1109/CISP-BMEI56279.2022.9980118 . hal-04454678

HAL Id: hal-04454678

<https://utt.hal.science/hal-04454678>

Submitted on 20 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Detection of Heart Failure Using a Convolutional Neural Network via ECG Signals

Jad Botros¹, Farah Mourad-Chehade¹ and David Laplanche^{1,2}

¹Computer Science and Digital Society Laboratory (LIST3N)

Université de Technologie de Troyes

²Pôle Santé publique, Hôpitaux Champagne Sud (HCS)

Troyes, France

jad.botros@utt.fr, farah.chehade@utt.fr, david.laplanche@hcs-sante.fr

Abstract—Heart failure (HF) is a chronic heart condition that increases mortality, morbidity, and healthcare costs. The electrocardiogram (ECG) is a noninvasive and straightforward diagnostic tool that can reveal detectable changes in HF. Because of their small amplitude and duration, these changes can be subtle and potentially misclassified during manual interpretation or when analyzed by clinicians. This paper reports a 7-layer deep convolutional neural network (CNN) model for HF automatic detection. The proposed CNN model requires only minimal pre-processing of ECG signals and does not require any engineered features. The model is trained and tested using an unbalanced and a balanced datasets extracted from the MIT-BIH and the BIDMC databases, achieving an accuracy of 99.73%, a sensitivity of 99.58%, and a specificity of 99.83% when the dataset is unbalanced and an accuracy of 99.26%, a sensitivity of 99.37%, and a specificity of 99.12% when the dataset is balanced.

Index Terms—Deep Learning, Binary Classification, Convolutional Neural Network, Heart Failure, Electrocardiogram.

I. INTRODUCTION

Heart failure (HF) is the inability of the heart to pump enough blood to meet metabolic demands or to supply venous return. It is a clinical syndrome caused by muscular tissue injuries in the heart. Ischemic heart disease, hypertension, and diabetes are the leading causes of these injuries. As the heart fails, patients experience a variety of symptoms such as dyspnea due to pulmonary congestion, edema and ascites due to impaired venous return. Common constitutional symptoms include nausea, loss of appetite, and fatigue. Several compensatory mechanisms occur as the failing heart attempts to maintain adequate function, such as increasing cardiac output, increasing ventricular volume and wall thickness, and maintaining tissue perfusion with augmented mean arterial pressure, among other things. Although these mechanisms may be beneficial in the early stages of HF, they all eventually lead to a vicious cycle of worsening HF [1].

According to statistics, HF affects 2 to 3% of the population in developed countries: approximately 15 million Europeans are affected by this syndrome [2]. Globally, more than 26 million people have been diagnosed with HF, creating a major public health issue as well as a significant economic burden [3]. The prevalence of this syndrome has been increasing in recent years and is expected to rise further as the population

ages and even patients with cardiac conditions, which may trigger the development of HF, live longer. Despite the advances in HF care, the disease still has a high morbidity and mortality rate: nearly half of all patients diagnosed with HF die within five years. HF, on the other hand, is associated with frequent hospitalization: in Europe, HF accounts for 5% of all acute hospital admissions [2]. HF care accounts for 1 to 2% of total health-care costs in developed countries. The cost of hospitalization accounts for the majority of total spending [4].

A combination of signs and symptoms, corroborated by tests, is required for the diagnosis of HF. The electrocardiogram (ECG) is a non-invasive measurement used to record the electrical activity of the heart. HF has an effect on such a signal. However, visual evaluation of the ECG signal obtained from the patient takes time, and manual interpretation is subject to inter-observer variability. The changes that hit the ECG signal are of a differential nature, making them difficult to be detected with the naked eye. This highlights the importance of using Computer Aided Diagnosis (CAD) systems to detect such abnormalities, because of their improved accuracy and reliability. A typical CAD system consists of four key procedures: signal pre-processing, extraction of unique features, selection of significant features, and classification.

This paper proposes a novel model based on deep learning to diagnose HF with a short-time record of ECG signals. The proposed model is a deep Convolutional Neural Network (CNN) having seven layers, with minimal pre-processing requirements. The main advantage of using a CNN over traditional machine learning methods is that it automatically detects important features without any human intervention. With its seven layers, the proposed architecture is smaller than currently available architectures in the literature, and thus it is less consuming in memory and computation time. This approach is illustrated using real ECG signals acquired from ambulatory holter and ECG recorders databases. Evaluation metrics demonstrate better performances, even with 2s ECG segments as input.

The rest of the paper is organized as follows. Section II presents related work. Section III describes the database and the proposed method. Section IV presents and discusses the obtained results. Finally, Section V provides the conclusion.

II. RELATED WORK

Many machine learning and deep learning techniques have been proposed in the literature to address the HF diagnosis issue. Asyali investigated the discrimination power of nine commonly used long term Heart Rate Variability (HRV) measures in his study [5]. He then used a Bayesian classifier to classify ECG signals as Normal or HF based on the standard deviation of NN intervals or SDNN. The obtained accuracy equals 93.24%, with sensitivity and specificity of 81.82% and 98.1% respectively. Melillo et al. distinguished between Normal and HF patients using a combination of HRV measures fed into a Classification and Regression Trees model (CART) [6]. The classification accuracy, sensitivity and specificity were 96.36%, 89.74% and 100% respectively. Liu et al. described a method for detecting HF using three other combinations of HRV measures and a Support Vector Machine (SVM) classifier [7]. The model was perfect in terms of accuracy, sensitivity and specificity. Masetic et al. used feature extraction and classification to achieve the same goal [8]. The autoregressive burg method was used to extract the features, which were then fed into a Random Forest classifier, which achieved 100% accuracy, sensitivity, and specificity.

Data pre-processing is an essential step in machine learning techniques because the quality of the data, and thus the information that can be extracted from it, directly affects the network's learning ability. Another important step in machine learning is feature selection. The selection process can be time-consuming and labor-intensive. Deep learning models have been used to optimize the performance of a CAD system in order to avoid the pitfalls of traditional machine learning. Chen et al. used RR interval segments to detect HF using a deep learning model based on sparse auto-encoders (SAE) [9]. The SAE were used to extract some features, which were then fed to a Deep Neural Network (DNN) for classification. The results showed a 72.44% accuracy, a 50.93% sensitivity, and a 80.93% specificity. Wang et al. presented a deep learning method based on Long Short Term Memory (LSTM) to classify ECG signals as Normal or HF [10]. A 500-point RR intervals segment was fed into the network. The model's accuracy was 84.91%, its sensitivity was 75.49%, and its specificity was 90.06%. Acharya et al. proposed an 11-layer Convolutional Neural Network (CNN) for categorizing ECG signals as Normal or HF [11]. The accuracy was 98.97%, the sensitivity was 98.87%, and the specificity was 99.01%. Wang et al. used LSTM and an inception module to detect HF. For an input of 500 points RR interval segment, the network achieved 86.42% accuracy, 74.91% sensitivity, and 91.21% specificity [12]. Padmavathi et al. discussed an 11-layer CNN approach for detecting

TABLE I
EVALUATION METRICS (IN %) OF THE DIFFERENT METHODS FOUND IN THE LITERATURE

| Method | Classifier | Accuracy | Sensitivity | Specificity |
|--------|-----------------------|----------|-------------|-------------|
| [5] | Bayesian Classifier | 93.24 | 81.82 | 98.1 |
| [6] | CART | 96.36 | 89.74 | 100 |
| [7] | SVM | 100 | 100 | 100 |
| [8] | RF | 100 | 100 | 100 |
| [9] | DNN | 72.44 | 50.93 | 80.93 |
| [10] | LSTM | 84.91 | 75.49 | 90.06 |
| [11] | CNN | 98.97 | 98.87 | 99.01 |
| [12] | Inception Module LSTM | 86.42 | 74.91 | 91.21 |
| [13] | CNN | 80.1 | 81 | 79.3 |
| [14] | CNN | 98.5 | 99.3 | 97.89 |

HF [13]. The model had an 80.10% accuracy rate, an 81% sensitivity, and a 79.30% specificity. Lih et al. created a 16-layer CNN-LSTM model for ECG signal classification. This one achieved 98.5% accuracy, 99.3% sensitivity and 97.89% specificity [14]. Table I summarizes the evaluation metrics of the different methods found in the literature.

III. MATERIALS AND METHOD

A. Database Description

The ECG signals used in this retrospective study are obtained from two publicly accessible databases: the MIT-BIH normal sinus rhythm (MIT-BIH NSR) database [15] and the BIDMC congestive heart failure (BIDMC CHF) database [16]. The former database contains eighteen long-term ECG signals from five healthy men aged 26 to 45 and thirteen healthy women aged 20 to 50. The ECG signals are sampled at 128 Hz and cover 20h of recordings. The latter database contains fifteen long-term ECG signals from eleven men aged 22 to 71 and four women aged 54 to 63 who are diagnosed with heart failure. The HF ECG signals are sampled at a frequency of 250 Hz, also with 20h of recordings. To ensure that all ECG signals are sampled at the same frequency, the MIT-BIH NSR recordings are upsampled to 250 Hz. The signals are then segmented into 2s fragments. This leads to 36 000 2-s ECG segments per signal. Following that, the segments are regularized using Z-score normalization to ensure that they all contribute equally to the analysis. Table II summarizes the characteristics of each database.

TABLE II
DATABASES CHARACTERISTICS

| Databases | Number of Signals | Properties |
|-------------|--------------------------|---|
| MIT-BIH NSR | 18 Long-Term ECG Signals | 5 men between 26 and 45 13 women between 20 and 50 fs=128Hz |
| BIDMC CHF | 15 Long-Term ECG Signals | 11 men between 22 and 71 4 women between 54 and 63 fs=250Hz |

The generated dataset lacks balance, with unequal representation of the two classes (60 % normal and 40 % HF). A balanced dataset is created to avoid the problem of data imbalance by taking equal numbers of ECG segments from the MIT-BIH NSR and the BIDMC CHF databases. The resulting dataset is composed of 240 000 2s-ECG fragments, with 120 000 for each class, normal or HF. More fragments could be taken, at the cost of higher computation times and memory consumption. In order to involve equally all subjects for each class, with 18 normal subjects and 15 HF patients: out of the 18 normal subjects, 15 are randomly chosen, and 8000 2-s ECG samples are randomly selected per patient. The obtained dataset will be used in the following for training and testing the model.

B. Method

This paper proposes a 7-layer CNN. The main advantage of using a CNN over its predecessors is that it can detect important features automatically without human intervention. Its low reliance on pre-processing reduces the need for human effort while expanding its capabilities. The model consists of two convolutional layers, two max-pooling layers, and three fully-connected layers. The convolutional layer is made up of a fixed number of filters and is used to extract the maps of the features. Convolutional layers perform the convolution of different filters with the input signal according to equation 1 in which Y is the result of the convolution of the signal x by the kernel f , n and k are positional indexes, and N is the size of the signal. The max-pooling layer, which takes the maximum value in a specific filter region, is used to reduce the network’s dimensionality. The fully connected layer, which is usually placed before the output layer, is used to aggregate data from the final feature map and generate the final classification. Fig. 1 shows the general structure of the proposed network.

$$Y_n = \sum_{k=0}^{N-1} x_k f_{n-k}, \quad (1)$$

The first layer takes the 2-s ECG segment of 500 points as input. It is a convolutional layer with five 1×13 filters applied with a stride of 1 i.e. five feature maps are generated by convolving the different filters with the input 500 points ECG signal. The second layer is a max-pooling layer with a pool size of two and a stride of 4. This layer reduces the dimensions of the features maps by convolving a 1×2 filter with each of the previously generated feature maps. As a result, the number of parameters to learn is reduced, as is the amount of processing in the network. Consequently, the model is more robust to changes in feature position in the input. Following that, another convolutional layer of ten filters of size 1×9 each is applied with a stride of 1. This set of filters is applied to the dimensionally reduced feature maps extracting a higher-level features. The fourth layer is

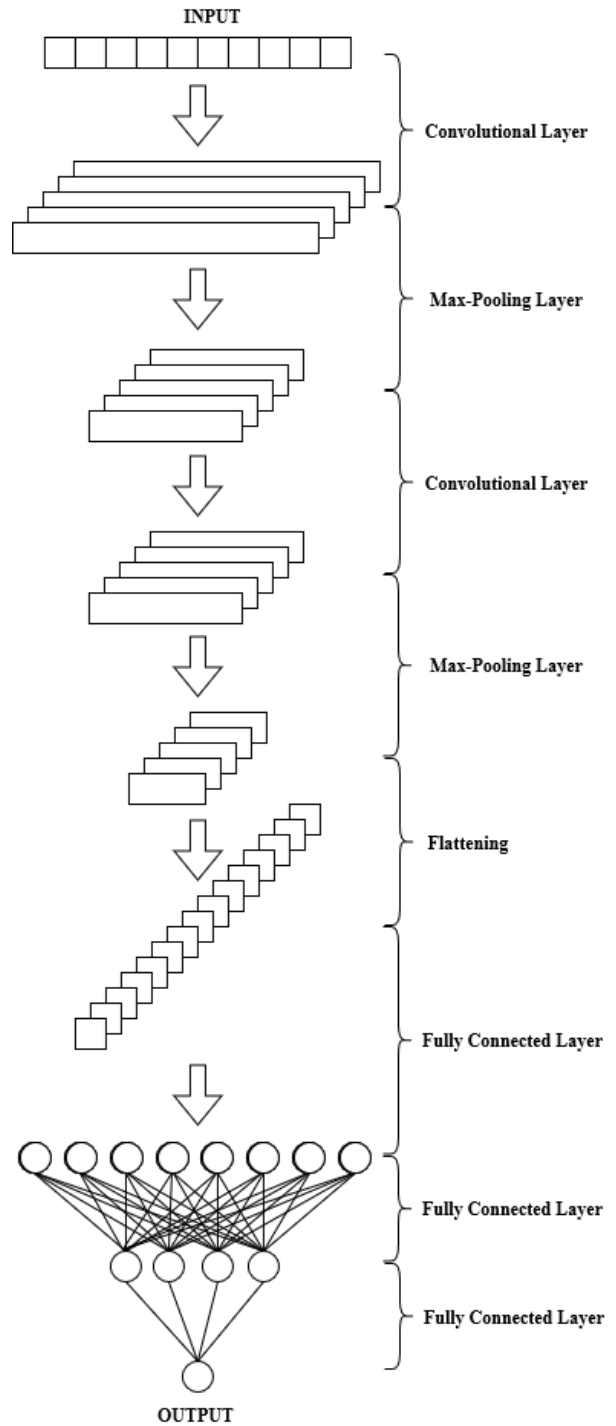


Fig. 1. General structure of the proposed model

a max-pooling layer with the same properties and tasks as the second layer. After that are three fully connected layers, each with 40, 20, and 2 units. The input to the first layer is nothing but a flattened version of the previous layer’s output. Except for the final layer, which uses the softmax activation function, all layers use the Leaky ReLU activation function.

The model weights are initialized using Glorot uniform initialization, and they are updated using backpropagation with a batch size of 10. The model is built up over a period of sixty epochs. Let y_i be the label of segment i of the training databaset, that is $y_i = 1$ if the segment i is taken from a HF signal and $y_i = 0$ otherwise. Consider \hat{p}_i the predicted probability that the segment i is HF, obtained at the output of the network. For the binary classification problem, the binary cross entropy function is used to calculate the model’s loss, as shown in the following equation:

$$\mathcal{L}(q) = \frac{-1}{M} \sum_{i=1}^M y_i \cdot \log(\hat{p}_i) + (1 - y_i) \cdot \log(1 - \hat{p}_i),$$

where M is the total number of segments taken in the training phase and q is the index of the epoch. The cross entropy computes a score that represents the mean difference between the actual and the predicted values. The score is to be minimized, where the 0 value is a perfect cross entropy.

IV. RESULTS AND DISCUSSION

The model is validated using stratified 10-fold cross validation over the dataset of 208 000 segments. The dataset is divided into ten equal parts with an equal percentage of each class in each part, and the training and the validation are performed ten times, each time using 9 folds for the training and the remaining one for the validation. Following this stratified 10-fold cross-validation procedure, a blindfold validation of normal and HF ECG segments from 4 patients (32 000 segments) with complete holdout is conducted in order to further assess the proposed model. This process is repeated 10 times and Tables III and IV display and summarize the results from this fully held-out test set considering the unbalanced and balanced datasets respectively.

The test set of fully unseen patients shows that the proposed classifier achieves an overall average classification accuracy of 99.263%, correctly categorizing 99.374% of normal ECG fragments and 99.124% of ECG fragments exhibiting indications of HF. Three evaluation metrics are used in this paper to assess the performance of the proposed model that are accuracy, sensitivity and specificity, as listed below:

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP}$$

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

where TP , TN , FP and FN denote the true positive, the true negative, the false positive and the false negative rates.

Tables III and IV summarize the aforementioned measures for the executed ten folds. As can be seen, the presented model achieves an average accuracy of 99.73%, with an

TABLE III
TEN FOLDS EVALUATION METRICS (UNBALANCED DATASET)

| Fold | Accuracy (%) | Sensitivity (%) | Specificity (%) |
|--------------------|--------------|-----------------|-----------------|
| 1 | 99.70 | 99.62 | 99.75 |
| 2 | 99.94 | 99.96 | 99.91 |
| 3 | 99.98 | 99.97 | 100 |
| 4 | 99.95 | 99.93 | 99.96 |
| 5 | 99.83 | 99.53 | 99.80 |
| 6 | 99.91 | 99.97 | 99.82 |
| 7 | 98.90 | 97.50 | 99.89 |
| 8 | 99.40 | 99.56 | 99.27 |
| 9 | 99.99 | 99.98 | 99.99 |
| 10 | 99.95 | 100 | 99.90 |
| Average (%) | 99.73 | 99.58 | 99.83 |

TABLE IV
TEN FOLDS EVALUATION METRICS (BALANCED DATASET)

| Fold | Accuracy (%) | Sensitivity (%) | Specificity (%) |
|--------------------|--------------|-----------------|-----------------|
| 1 | 96.74 | 99.95 | 94.25 |
| 2 | 99.85 | 99.68 | 100 |
| 3 | 99.91 | 99.80 | 100 |
| 4 | 99.98 | 99.96 | 100 |
| 5 | 99.05 | 99.27 | 99.12 |
| 6 | 99.97 | 99.98 | 99.96 |
| 7 | 98.44 | 96.22 | 100 |
| 8 | 99.74 | 99.13 | 100 |
| 9 | 99.69 | 99.82 | 99.58 |
| 10 | 99.15 | 99.93 | 98.33 |
| Average (%) | 99.26 | 99.37 | 99.12 |

average sensitivity of 99.58% and an average specificity of 99.83% with an unbalanced dataset and an average accuracy of 99.26%, with an average sensitivity of 99.37% and an average specificity of 99.12%, with respectively low standard deviations with a balanced dataset. The number of epochs chosen is sufficient for the loss function to converge and reach a stable minimum. The model’s average loss is 0.024 when the dataset is unbalanced and 0.08 when the dataset is balanced. The code was written in Python and implemented in Google Colab.

In order to illustrate the performances of the proposed method, a comparison is established between the achieved results and those of other HF diagnosis methods using the same databases. Three existing methods are thus considered. The authors of [6] explored the discrimination power of various Heart Rate Variability (HRV) measures and chose three combinations of these measures to form the input to a CART classifier. The method in [11] used an 11-layer CNN to classify 2s ECG fragments, with minimal pre-processing, into N or HF. The authors of [13] upsampled the ECG signals to a common frequency, filtered them using empirical mode decomposition, and performed R-peak detection prior to beats segmentation. For classification, the beats were fed into an 11-layer CNN. Table V shows the test evaluation metrics of methods of the literature compared to the proposed model. The results of the method in [11] given in Table V are obtained with an unbalanced dataset in the higher row. When balanced

TABLE V
EVALUATION METRICS (IN %) OF DIFFERENT METHODS FOUND IN THE LITERATURE

| Method | Accuracy | Sensitivity | Specificity |
|-------------------------------------|----------|-------------|-------------|
| [6] | 96.36 | 89.74 | 100 |
| [11]-unbalanced | 98.97 | 98.87 | 90.01 |
| [11]-balanced | 94.40 | 94.68 | 94.12 |
| [13] | 80.10 | 81 | 79.30 |
| Proposed Model (Unbalanced dataset) | 99.73 | 99.58 | 99.83 |
| Proposed Model (Balanced dataset) | 99.26 | 99.37 | 99.12 |

data are considered, the accuracy drops to 94.40%, as do the sensitivity and the specificity, which drop from 98.87% to 94.68% and from 99.01% to 94.12%, respectively, as shown in the lower row. The results show that the proposed model performs better than existing methods. A higher specificity is achieved with method of [6], however it is at the cost of the sensitivity and the accuracy that are lower than the obtained one with the proposed model. The use of deep learning in the presented method, along with an efficient training and a suitable architecture, ensures high performances compared to traditional machine learning methods. Indeed, deep learning capabilities enable more reliable signal abstraction in high-dimensional space without the need for human intervention. When compared to the deep models architectures in [11] and [13], the proposed model is smaller, with 7 layers instead of 11, and faster when training time is taken into account.

V. CONCLUSION

An automated ECG classification system for the detection of heart failure is proposed in this paper. The proposed model consists of a 7-layer CNN and is fully automatic, with no pre-processing or feature selection stages required. Normal ECG recordings are obtained from the MIT BIH NSR database, and heart failure ECG recordings are obtained from the BIDMC CHF database. The two databases are freely accessible on the internet. To avoid data imbalance issues, the dataset is created with 120000 normal ECG segments and 120000 ECG segments from heart failure patients. The presented CNN classifier, which achieved high metrics, plays an important role in the detection and classification of ECG signals, as evidenced by its accuracy of 99.73%, sensitivity of 99.58% reflecting the true positive rate, and specificity of 99.83% which represents the true negative rate, when the dataset is unbalanced, and its accuracy of 99.26%, sensitivity of 99.37%, and specificity of 99.12%, when the dataset is balanced. Future work may include the fusion of various medical data, such as times series, punctual measurements and clinical notes, to ensure an efficient monitoring of the state of a subject over time and to offer a more precise classification with a stratification of the patients into different levels of heart failure. The problem of classification would then be defined as a multi-class hybrid classification, combining data of different natures.

REFERENCES

- [1] Kemp, C. & Conte, J. The pathophysiology of heart failure. *Cardiovascular Pathology*. **21**, 365-371 (2012)
- [2] Braunschweig, F., Cowie, M. & Auricchio, A. What are the costs of heart failure?. *Europace*. **13**, ii13-ii17 (2011)
- [3] Ponikowski, P., Anker, S., AlHabib, K., Cowie, M., Force, T., Hu, S., Jaarsma, T., Krum, H., Rastogi, V., Rohde, L. & Others Heart failure: preventing disease and death worldwide. *ESC Heart Failure*. **1**, 4-25 (2014)
- [4] Berry, C., Murdoch, D. & McMurray, J. Economics of chronic heart failure. *European Journal Of Heart Failure*. **3**, 283-291 (2001)
- [5] Asyali, M. Discrimination power of long-term heart rate variability measures. *Proceedings Of The 25th Annual International Conference Of The IEEE Engineering In Medicine And Biology Society (IEEE Cat. No. 03CH37439)*. **1** pp. 200-203 (2003)
- [6] Melillo, P., Fusco, R., Sansone, M., Bracale, M. & Pecchia, L. Discrimination power of long-term heart rate variability measures for chronic heart failure detection. *Medical Biological Engineering Computing*. **49**, 67-74 (2011)
- [7] Liu, G., Wang, L., Wang, Q., Zhou, G., Wang, Y. & Jiang, Q. A new approach to detect congestive heart failure using short-term heart rate variability measures. *PLoS One*. **9**, e93399 (2014)
- [8] Masetic, Z. & Subasi, A. Congestive heart failure detection using random forest classifier. *Computer Methods And Programs In Biomedicine*. **130** pp. 54-64 (2016)
- [9] Chen, W., Liu, G., Su, S., Jiang, Q. & Nguyen, H. A CHF detection method based on deep learning with RR intervals. *2017 39th Annual International Conference Of The IEEE Engineering In Medicine And Biology Society (EMBC)*. pp. 3369-3372 (2017)
- [10] Wang, L., Zhou, W., Liu, N., Xing, Y. & Zhou, X. CHF Detection with LSTM Neural Network. *2018 40th Annual International Conference Of The IEEE Engineering In Medicine And Biology Society (EMBC)*. pp. 514-517 (2018)
- [11] Acharya, U., Fujita, H., Oh, S., Hagiwara, Y., Tan, J., Adam, M. & San Tan, R. Deep convolutional neural network for the automated diagnosis of congestive heart failure using ECG signals. *Applied Intelligence*. **49**, 16-27 (2019)
- [12] Wang, L. & Zhou, X. Detection of congestive heart failure based on LSTM-based deep network via short-term RR intervals. *Sensors*. **19**, 1502 (2019)
- [13] Padmavathi, C. & Veenadevi, S V. Heart Disease Recognition from ECG Signal Using Deep Learning. *International Journal of Advanced Science and Technology*. **29**, pp. 2303-2316 (2020)
- [14] Lih, O., Jahmunah, V., San, T., Ciaccio, E., Yamakawa, T., Tanabe, M., Kobayashi, M., Faust, O. & Acharya, U. Comprehensive electrocardiographic diagnosis based on deep learning. *Artificial Intelligence In Medicine*. **103** pp. 101789 (2020)
- [15] Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P., Mark, R., Mietus, J., Moody, G., Peng, C. & Stanley, H. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation*. **101**, E215-20 (2000,6)
- [16] Baim, D., Colucci, W., Monrad, E., Smith, H., Wright, R., Lanoue, A., Gauthier, D., Ransil, B., Grossman, W. & Braunwald, E. Survival of patients with severe congestive heart failure treated with oral milrinone. *J. Am. Coll. Cardiol.* **7**, 661-670 (1986,3)