Automatic diagnosis of newly emerged heart failure from serial electrocardiography by repeated structuring & learning procedure

Sbrollini, A.; Barocci, M.; Mancinelli, M.; Paris, M.; Raffaelli, S.; Marcantoni, I.; ... ; Burattini, L.

Biomedical Signal Processing and Control 79 (2023) 104185

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Automatic diagnosis of newly emerged heart failure from serial electrocardiography by repeated structuring & learning procedure

Agnese Sbrollini [a], Maddalena Barocci [a], Martina Mancinelli [a], Michele Paris [a], Simone Raffaelli [a], Ilaria Marcantoni [a], Micaela Morettini [a], Cees A. Swenne [b], Laura Burattini [a,*]

[a] *Department of Information Engineering, Università Politecnica delle Marche, via Brecce Bianche 12, 60131 Ancona, Italy*
[b] *Cardiology Department, Leiden University Medical Center, PO Box 9600, 2300 RC Leiden, the Netherlands*

ARTICLE INFO

Keywords:
Deep Learning
Machine Learning
Artificial Neural Network
Repeated Structuring and Learning Procedure
Heart Failure
Serial Electrocardiography

ABSTRACT

Heart failure (HF) diagnosis, typically visually performed by serial electrocardiography, may be supported by machine-learning approaches. Repeated structuring & learning procedure (RS&LP) is a constructive algorithm able to automatically create artificial neural networks (ANN); it relies on three parameters, namely maximal number of hidden layers (MNL), initializations (MNI) and confirmations (MNC), arbitrarily set by the user. The aim of this study is to evaluate RS&LP robustness to varying values of parameters and to identify an optimized combination of parameter values for HF diagnosis. To this aim, the Leiden University Medical Center HF database was used. The database is constituted by 129 serial ECG pairs acquired in patients who experienced myocardial infarction; 48 patients developed HF at follow-up (cases), while 81 remained clinically stable (controls). Overall, 15 ANNs were created by considering 13 serial ECG features as inputs (extracted from each serial ECG pair), 2 classes as outputs (cases/controls), and varying values of MNL (1, 2, 3, 4 and 10), MNI (50, 250, 500, 1000 and 1500) and MNC (2, 5, 10, 20 and 50). The area under the curve (AUC) of the receiver operating characteristic did not significantly vary with varying parameter values ($P \geq 0.09$). The optimized combination of parameter values, identified as the one showing the highest AUC, was obtained for MNL = 3, MNI = 500 and MNC = 50 (AUC = 86 %; ANN structure: 3 hidden layers of 14, 14 and 13 neurons, respectively). Thus, RS&LP is robust, and the optimized ANN represents a potentially useful clinical tool for a reliable automatic HF diagnosis.

1. Introduction

Heart failure (HF) is a common and potentially fatal heart disease that currently affects about 2 % of the adult population, with peaks up to 10 % in subjects over 65 years old; the risk of death at one year from the first diagnosis is about 35 % [1,2]. According to guidelines of the European Society of Cardiology, HF is a clinical syndrome characterized by breathlessness, ankle swelling and fatigue. These symptoms are usually associated with clinical evidence, such as elevated jugular venous pressure, pulmonary crackles, and peripheral edema. These symptoms and clinical evidences are caused by a structural and/or functional cardiac abnormality, resulting in a reduced cardiac output and/or elevated intracardiac pressures at rest or during stress [3]. This broad definition reflects the complexity of the disease that has about seventeen primary aetiologies. However, more than two-thirds of HF cases can be attributed to four underlying conditions: ischemic heart disease, chronic obstructive pulmonary disease, hypertensive heart disease and rheumatic heart disease [4].

While the primary cause of HF may be extracardiac, presence of one or more underlying cardiac abnormalities is central for HF diagnosis. According to the current definition, HF is present when symptoms occur; however, some asymptomatic patients may present structural or functional cardiac abnormalities that are precursors of HF. Timely recognition and treatment of these precursors may help to contrast HF natural

Abbreviations: Acc, accuracy; ANN, artificial neural networks; AUC, area under the curve; CI, 95% confidence intervals; ECG, electrocardiogram; HF, Heart failure; HFDB, heart-failure database; LCT, learning computational time; MNC, maximal number of confirmations; MNI, maximal number of initializa-tions; MNL, maximal number of hidden layers; NTOT, total number of neurons; *OP*, operating point; P, level of statistical significance; ROC, receiver operating characteristic; RS&LP, repeated structuring & learning procedure; Se, sensitivity; Sp, specificity; TCT, testing computational time; VCG, vectorcardiogram.

* Corresponding author at: Via Brecce Bianche 12, 60131 Ancona, Italy.
E-mail address: l.burattini@univpm.it (L. Burattini).

https://doi.org/10.1016/j.bspc.2022.104185
Received 8 October 2021; Received in revised form 19 July 2022; Accepted 4 September 2022
Available online 15 September 2022
1746-8094/

development [5] and lead to positive outcomes [3]. Presence of several HF precursors implies changes in the electrical properties of the heart and, thus, variations of the electrocardiogram (ECG) with respect to normal. The ECG is the recording of the electrical activity of the heart; by its nature, it is a pseudo-periodic signal consisting in the repetition of a pattern showing a sequence of typical waves, which are: the P wave, reflecting atrial depolarization; the QRS complex, reflecting ventricular depolarization and hiding atrial repolarization; and the T wave, reflecting ventricular repolarization. Thus, morphological and temporal ECG features represent physiological phenomena occurring within the heart and may indicate the presence of cardiac abnormalities. HF occurrence is unlikely in patients with a completely normal ECG; however, most ECG abnormalities are not HF specific [3].

In this study we focused on the automatic early diagnosis of HF through serial ECG changes. Serial ECG changes consist in ECG differences observed when comparing two ECGs, one newly and one previously acquired from the same subject [6–8]. In daily clinical practice, serial electrocardiography is usually done by visual inspection. Nevertheless, its complexity has recently suggested machine-learning approaches [9–11]. Serial electrocardiographic analysis supported by the machine-learning algorithm called Repeated structuring & learning procedure (RS&LP) has provided important preliminary results in the detection of newly emerging or aggravating cardiac pathology [10,11], and represents an important example of how machine-learning approaches could support advances in clinics and healthcare. In order to provide examples of possible RS&LP clinical applications, the procedure was also used to detect newly emerged HF [11]. Being proposed just as an example, the application relied on an arbitrary and unoptimized setting of parameters; optimization of the parameter setting, however, becomes necessary when proposing RS&LP as a useful tool for HF diagnosis in the clinical practice.

Thus, the aim of the present study is to evaluate the robustness of the RS&LP to varying values of it's a-priory arbitrarily parameters and to identify a best combination of parameters for automatic HF diagnosis.

2. Related works

Several works have previously presented machine learning approaches for automatic HF detection [12]. Used clinical data are heterogeneous and include electrocardiographic data, echocardiographic data, electronic health records data and data from other sources (*e.g.*, post-mortem clinical analysis) [12]. Only few studies presented machine learning approaches for HF detection from ECG analysis [13–18]; proposed techniques include deep fully-connected neural networks [17], convolutional neural networks [13,19,20], long-short term memory [18], random forest classifiers [16] and support vector machine [14,15]. Most works considered directly the cardiac signals (ECG [13–16,19,20] or heart-rate series [18]) as input of the classifiers; only one study considered both demographic and electrocardiographic features [17]. Two studies [17,19] were performed on a huge amount of data (>50,000 patients); the others were performed on smaller datasets (less than 100 patients) selected from open access databases [13–16,18,20]. Some works aimed to discriminate HF patients from subjects showing normal sinus rhythm [13–18] and show their major limitation in not considering possible comorbidities, which represent clinical confounders and, thus, may jeopardize HF diagnosis. Only two works [19,20] considered the presence of other pathologies such as diabetes mellitus, hypercholesterolemia, renal disease, hypertension, coronary artery disease and myocardial infarction. In one study [19] the other pathologies were considered as comorbidities that could affect both HF patients as well as patients constituting the control group. In another study [20] pathologies other than HF were affecting the patients constituting the control group only.

3. Materials and methods

3.1. Repeated structuring & learning procedure

RS&LP is a recently presented constructive algorithm for automatic creation of a supervised and fully connected artificial neural network (ANN) [11]. In its general formulation, the procedure takes as input a set of data features (one input neuron for each feature), constructs the ANN according to the algorithm described below and classifies the data (one output neuron for each considered class, except for binary classifications for which only one output neuron is required). RS&LP creates the ANN by using a learning dataset composed of a training dataset and a validation dataset. Class weights, each defined as the inverse of the corresponding class prevalence, are considered [21] to compensate for potential disproportions among distributions of cases over the output classes. Each neuron is characterized by a sigmoid activation function; weights and bias, ranging between −1 and +1, are randomly initialized.

ANN construction (Fig. 1) occurs on the basis of an iterative procedure composed of three main phases, namely structuring phase, learning phase, and confirmation phase. The procedure starts from an original ANN composed of the input layer, one hidden layer constituted by one neuron, and the output layer. During the structuring phase, the original ANN is upgraded into several different candidate ANNs obtained by adding a neuron to an existing hidden layer or to a new hidden layer. Each candidate ANN must respect two structural rules: the number of layers cannot exceed the "maximal number of hidden layers" (MNL, the numerical value of which is initially set by the user); and the number of neurons in a layer cannot be larger than the number of neurons in the previous layer. The learning phase consists of training and validation subphases, both including several epochs during which training and validation errors are computed. Training is performed using the scaled-conjugate-gradients algorithm [22], a training algorithm presenting reliable performance in terms of computational effort, classification accuracy, even if applied to small datasets [23–25].Validation relies on the early stopping criterion to avoid overfitting [26]. When the learning phase starts, weights and biases of the neurons added during the structuring phase in each candidate ANN are initialized. Initialization is acceptable only if it implies a decrement of the training error after only one epoch. Thus, if initialization of a new neuron is not immediately acceptable, the neuron is re-initialized. The number of initializations of a new neuron cannot exceed the "maximal number of initializations" (MNI, the numerical value of which is initially set by the user). All candidate ANNs with an acceptable initialization are learnt. During the confirmation phase, the validation errors of all learnt candidate ANNs are compared with the validation error of the original ANN. If the validation error of one or more candidate ANNs is less or equal to the validation error of the original ANN, the candidate ANN with the smallest validation error becomes the new original ANN; if the validation error of all candidate ANNs are larger than the validation error of the original ANN, the original ANN remains as such. Then, the procedure starts anew by using the updated original ANN. RS&LP ends when there are no acceptable candidate ANNs, when a candidate ANN reached the "maximum number of confirmations" (MNC; the numerical value of which is initially set by the user), or when there are no misclassifications in the learning dataset. When one of the above-listed stopping criteria occurs, the original ANN is considered as the final ANN. The pseudocode ot the RS&LP is reported in Fig. 2; further details on the RS&LP can be found in [11].

To compensate for random initializations of neurons possibly leading to different final ANNs, RS&LP is run 100 times so that 100 final ANNs were obtained. Among them, the ANN with the largest area under the curve of the receiver operating characteristic on the learning dataset is considered to be the best ANN. For convenience, ANN structure is represented in terms of [N1, N2,…,NL], where Ni is the number of neurons in the i^{th} layer, with $i = 1,2…,NL$, with NL being the number of layers in the ANN and total number of neurons (NTOT).

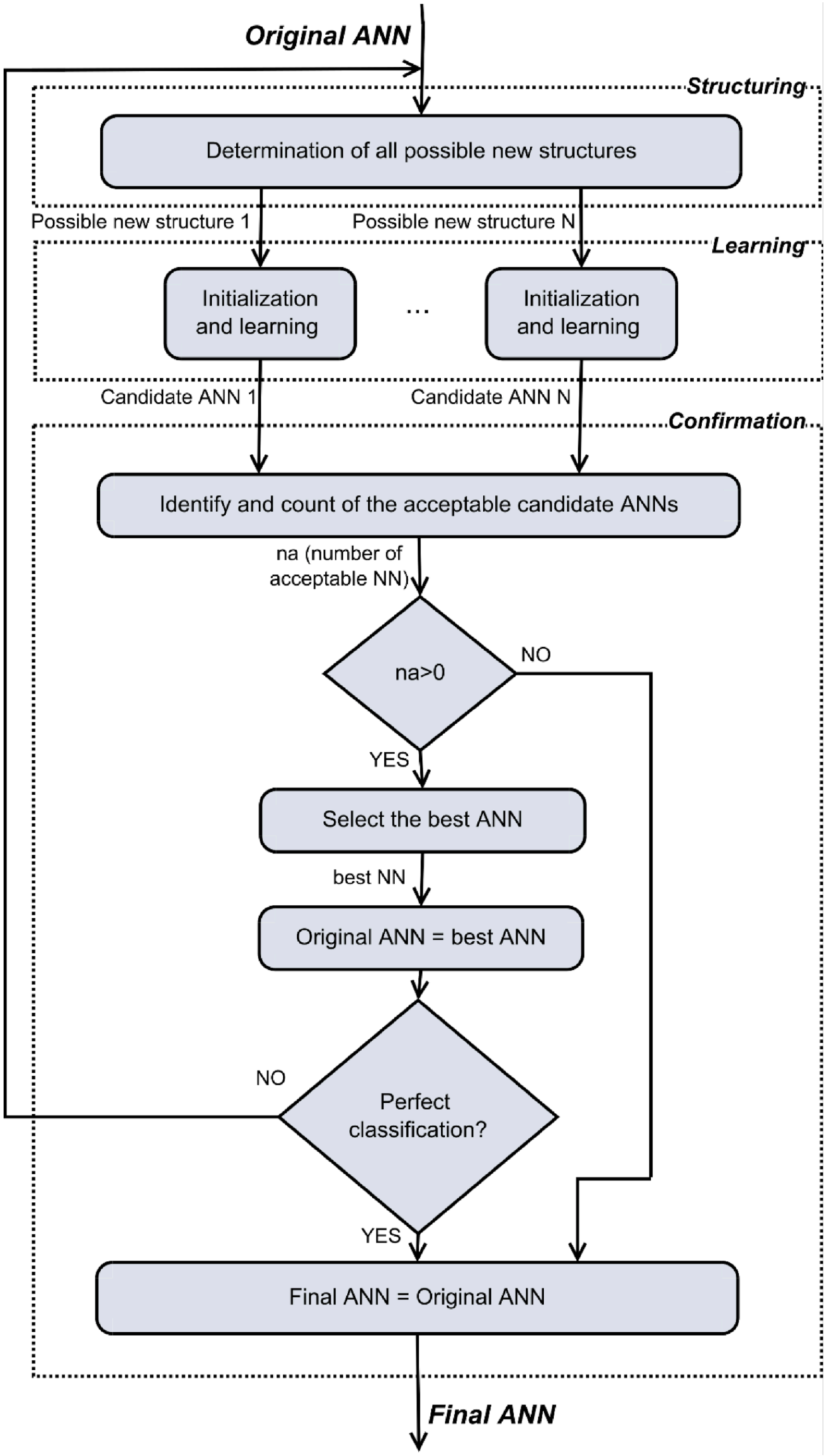

Fig. 1. Flowchart of the repeated structuring & learning procedure (RS&LP).

3.2. Heart failure detection by the repeated structuring & learning procedure

In this study, RS&LP was applied to serial electrocardiography for the detection of newly emerging HF. In this application the input set of data features consisted of 13 serial ECG features measured on the median beat of the vectorcardiogram (VCG) that is the orthonormal representation of the standard 12-lead ECG [27]. Given their associations with electrophysiological phenomena [11,27], the following 13 serial ECG features were considered: QRS-duration difference (ms), QT-

```
START
%LearnD is the learning dataset
Request: LearnD
%Structure of the initial NN (first iteration).
S0 ← [1];
%Training of the initial architecture with the learning data
EV0, ET0, NN0 ← Train S0, LearnD;
% NNOR is the original NN
NNOR ← NN0;
% SOR is the structure of original NN
SOR  ← S0;
%EVOR is the training error of the original NN
ETOR ← ET0;
%EVOR is the validation error of the original NN
EVOR ← EV0;
%Number of confirmation of the same structure
Conf ← 0;
%As long as a stopping criterium does not occur
  WHILE (Missclass NNOR, LearnD == 0) or (Conf < MNC) do
%Determination of the possible new structures. S-NN contain N possible new structures.
%Structures with number of layer higher than MNL are rejected.
    S-NN ← Find_New_Architectures SOR;
    FOR n ← 1 to N do
      S-NNn ← S-NN(n);
%Item is the number of initializations
      Item ← 0;
      WHILE (ETn>ETOR) or (Init<MNI)
%Initializialization of new neurons, mantaining the weights and the biases of the existing NN.
        IS-NNn ← Initialize S-NNn, NNOR ;
        Item  ← Item +1 ;
%Training of the initialized possible new structure with only one iteration.
        [EVj, ETj, NNj] ← Train_One_Step IS-NNn, LearnD;
      END
      IF Init == MNI
        Reject NNn;
      ELSE
%Training of the initialized possible ANN
        EVn, ETn, NNn ← Train IANNn, LearnD;
% Receiver Operating Characteristing (ROC) curve analysis and area under the curve (AUC) computation
        AUCn ← ROC_Analysis NNn, LearnD;
%Saving of the all possible candidate NNs
        PossNN(n) ← NNn;
        AUC(n) ← AUCn ;
      END
    END
    IF (PossNN==0)
      Conf == Conf +1 ;
    ELSE
%Selection of the possible NN with the maximum AUC
      NNOR ← PossNN (max AUC );
    END
  END
FinalNN ← NNOR
RETURN FinalNN
END
```

Fig. 2. Pseudocode of the repeated structuring & learning procedure (RS&LP).

interval difference (ms), difference in maximal QRS-vector magnitude (μV), difference in maximal T-vector magnitude (μV), QRS-integral vector magnitude difference (mV·ms), T-integral vector magnitude difference (mV·ms), QRS-complexity difference (%), T-wave complexity difference (%), magnitude of the ventricular-gradient difference vector (mV·ms), magnitude of the QRS-T spatial-angle difference (°), heart-rate difference (bpm), magnitude of J-vector difference vector (μV) and T-wave symmetry difference (%). Number of neurons in the ANN input layer was 13 (as the number of input features); number of neurons in the ANN output layer was 1 (binary ouptut indicating presence or absence of HF).

Suboptimized ANN for detection of newly emerged HF was constructed, trained and tested on the HF database (HFDB) [11,28] that was retrospectively derived from the clinical ECG database of the Leiden University Medical Center (Leiden, The Netherlands). All retrospective evaluations reported here were undertaken in compliance with the ethical principles of Helsinki Declaration and approved by the Leiden University Medical Center Medical Ethics Committee. The HFDB contains 129 10-second 12-lead ECG pairs acquired in patients who had experienced a myocardial infarction. All patients were clinically stable at the moment of their baseline ECG recording, which was a routine ECG performed at least six months after the acute event. The 81 patients who remained clinically stable and did not develop HF during the follow-up were selected as control patients; their follow-up ECG was a routine ECG performed approximately-one year after the acute myocardial infarction. The remaining 48 patients who developed HF during follow-up were selected as case patients; their follow-up ECG was a routine ECG performed at HF initial occurrence.

Table 1
Division of the HFDB, constituted by case patients and control patients, into learning and testing datasets, and further division of the learning dataset into training and validation datasets.

	Learning (50 %)			Testing (50 %)	Total (100 %)
	Training (80 %)	Validation (20 %)	Total (100 %)		
Case patients	18	6	24	24	48
Control patients	34	7	41	40	81
Total	52	13	65	64	129

All ECGs were processed by the custom-made LEADS software [29] that computes the VCG and measures, among others, all the ECG features needed for this study. Eventually, the 13 serial ECG features mentioned above were computed by subtracting baseline ECG feature values from the corresponding follow-up ECG feature values. No normalization was performed because not consistent with of what normally done in clinics.

The HFDB was equally divided into a learning dataset and a testing dataset. The learning dataset was used for ANN creation by RS&LP; the testing dataset was used to assess classification performance. The learning dataset was further divided into a training dataset (80 % of the learning dataset) and a validation dataset (20 % of the learning dataset). The prevalence of cases and controls was maintained in all datasets. The distribution of case patients and control patients over the datasets is reported in Table 1.

3.3. Robustness analysis

In previous examples of RS&LP applications [10,11,30] values of MNL, MNI and MNC were arbitrarly set at 3, 500 and 10, respectively. Here, numerical setting of these parameters was varied to evaluate clinical performance of each ANN in the testing dataset. Value of each parameter was varied while keeping the values of the other two constant. Specifically, the following three tests were performed:

- Test 1 was performed to determine RS&LP robustness to varying value of MNL. Considered values of MNL were 1, 2, 3, 4 and 10, while values of MNI and MNC were kept constant at 500 and 10, respectively.
- Test 2 was performed to determine RS&LP robustness to varying value of MNI. Considered values of MNI were 50, 250, 500, 1000 and 1500, while values of MNL and MNC were kept constant at 3 and 10, respectively.
- Test 3 was performed to determine RS&LP robustness to varying value of MNC. Considered values of MNC were 2, 5, 10, 20 and 50, while values of MNL and MNI were kept constant at 3 and 500, respectively.

3.4. Statistical analysis

For each test, the best ANN was characterized by computing the area under the curve (AUC) of the receiver operating characteristic (ROC) and the associated 95 % confidence intervals (CI) in the testing dataset. ROCs obtained with all combinations of parameter values were compared using the DeLong's tests [31], setting a level of statistical significance (P) equal to 0.05. Finally, the operating point (OP), identified as the ROC point in which sensitivity (Se – represented in the vertical axes of ROC) equals specificiy (Sp – represented in the horizontal axes of ROC), was used to calculate number of true positives (TP, number of patients affected by HF classified as cases), true negatives (TN, number of patients not affected by HF classified as controls), false positives (FP, number of patients not affected by HF classified as cases) and false negatives (FN, number of patients affected by HF classified as controls). According with these definitions, values of accuracy (Acc), Se (equal to the vertical coordinate of OP on the ROC) and Sp (equal to the horizontal coordinate of OP on the ROC) were computed as follow:

$$Se = \frac{TP}{TP + FN} \tag{1}$$

$$Sp = \frac{TN}{TN + FP} \tag{2}$$

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

For each test, the numerical value of the analyzed parameter was selected considering those used to create the ANN having the highest AUC. The best paramenters configuration was finally identified by combining the selected values of MNL (from Test 1), MNI (from Test 2) and MNC (form Test 3).

3.5. Computational efficency

Computational efficiency of the RS&LP was evaluated in terms of computational time required to create the ANN over the learning dataset (LCT), and to classify data over the testing dataset (TCT). Processing was performed using MATLAB R2019b, running on an Intel(R) Core(TM) i7-2600 (RAM = 12 GB).

4. Results

Performance of ANNs obtained by performing the tests are reported in Table 2. Results of Test 1 indicate that, although the number of hidden layers increased with increasing MNL, it never became higher than 6, obtained for MNL equal to 10. NTOT also tended to increase with MNL, going from 26 (MNL = 1) to 63 (MNL = 10). The AUC values ranged from 77 % to 83 %, even though differences among the ROCs did not reach statistical significance ($P > 0.05$). ROCs relating to Test 1 are depicted in Fig. 3(a). The maximum value of AUC was 83 %, obtained for MNL = 3, thus representing the optimized MNL value in correspondence of which values of Acc, Se and Sp in $OP_{Se=Sp}$ were all 75 %.

Results of Test 2 show that NTOT varied from 28 (MNI = 1500) to 50 (MNI = 1000) without showing a clear trend. AUC values ranged from 68 % to 83 %, even though differences among ROCs did not reach statistical significance ($P > 0.05$). ROCs relating to Test 2 are depicted in Fig. 3(b). The maximum value of AUC was 83 %, obtained for MNI = 500, thus representing the optimized MNI value in correspondence of which values of Acc, Se and Sp in $OP_{Se=Sp}$ were all 75 %.

Results relative to Test 3 indicate that NTOT increased from 3 (MNC = 2) to 41 (MNC = 50) with increasing MNC. AUC values were quite stable, ranging from 78 % to 86 % ($P > 0.05$). ROCs relative to Test 3 are depicted in Fig. 3(c). The maximum value of AUC was 86 %, obtained only for MNC = 50, thus representing the optimized MNC value in correspondence of which values of Acc, Se and Sp in $OP_{Se=Sp}$ were all 75 %.

Table 2 also reports LCT values associated to all tests and combinations of parameter values; LCT value ranged from 2 h:44 min:28 s (Test 2 with MNL = 3, MNI = 50 and MNC = 10) to 212 h:44 min:11 s (Test 3 with MNL = 3, MNI = 500 and MNC = 50). TCT was 9 ms in all cases.

Considering the results of all tests, the best combination of parameter values is MNL = 3, MNI = 500 and MNC = 50. Thus, the ANN associated with this combination, having architecture equal to [14,13,13], AUC equal to 86 %, CI equal to 20 %, and Acc, Se and Sp all equal to 75 % (in $OP_{Se=Sp}$), LCT equal to 212 h:44 min:11 s, and CTC equal to 9 ms (Table 2; Fig. 4) represents the best ANN for the automatic diagnosis of newly emerged HF.

5. Discussion

This study evaluated the robustness of the RS&LP to varying values of parameters and identified the best combination of MNL, MNI, and MNC values for the automatic diagnosis of newly emerged HF from serial electrocardiography. The simultaneous availability of the previously proposed RS&LP and of the here-identified best combination of parameter values represents the main contribution of this work since makes RS&LP a tool immediately usable in clinics for HF diagnosis. Indeed, RS&LP may be used to diagnose different pathologies; however, to perform reliably, it needs to be associated with an optimal combination of parameter values specifically identified for that pathology.

As known, clinical interpretability of automatic decision support systems is essential in healthcare. To ensure interpretability of the results, our procedure for automatic HF diagnosis relies on intra-subject serial changes of ECG features and not on raw ECG data. Indeed, according to serial electrocardiography, absence of ECG changes indicates clinical stability of a patient, while occurrence of ECG changes may indicate emerging pathologies. By using the 13 serial ECG features, our machine-learning approach mimics and potentiates the decision procedure normally adopted by physicians, who visually compare two serial ECG tracings in search of clinically significant differences.

RS&LP robustness was evaluated by performing three tests during which one single parameter value was varied. Overall, thirteen different combinations of parameter values were considered. The obtained ANN

Table 2
Clinical performances of suboptimized artificial neural networks (ANNs) obtained by performing the three robustness tests.

Test	MNL	MNI	MNC	Architecture	AUC (%)	CI (%)	Acc (%)	Se (%)	Sp (%)	LCT (hh:mm:ss)
1	1	500	10	[26]	78	66–90	75	75	75	13:16:02
	2	500	10	[22,21]	80	68–92	72	71	73	16:28:12
	3	500	10	[12,12,8]	83*	72–94	75	75	75	18:04:30
	4	500	10	[15,15,10,10]	77	64–89	70	71	70	16:51:55
	10	500	10	[17,13,10,8,8,7]	77	64–89	67	67	68	22:43:46
2	3	50	10	[15,12,8]	68	54–82	63	63	63	2:44:28
	3	250	10	[26,10,10]	79	66–91	67	67	68	8:27:13
	3	500	10	[12,12,8]	83*	72–94	75	75	75	18:04:30
	3	1000	10	[17,17,16]	80	69–92	72	71	73	32:57:32
	3	1500	10	[14,7,7]	79	67–91	70	71	70	58:05:07
3	3	500	2	[1,1,1]	82	71–94	75	75	75	3:36:44
	3	500	5	[15,7,7]	85	74–95	75	75	75	6:30:07
	3	500	10	[12,12,8]	83	72–94	75	75	75	18:04:30
	3	500	20	[19,10,9]	78	66–91	70	71	70	39:17:36
	3	500	50	[14,14,13]	86*§	76–96	75	75	75	212:44:11

*suboptimized ANN with the highest AUC within a test; §optimized ANN.

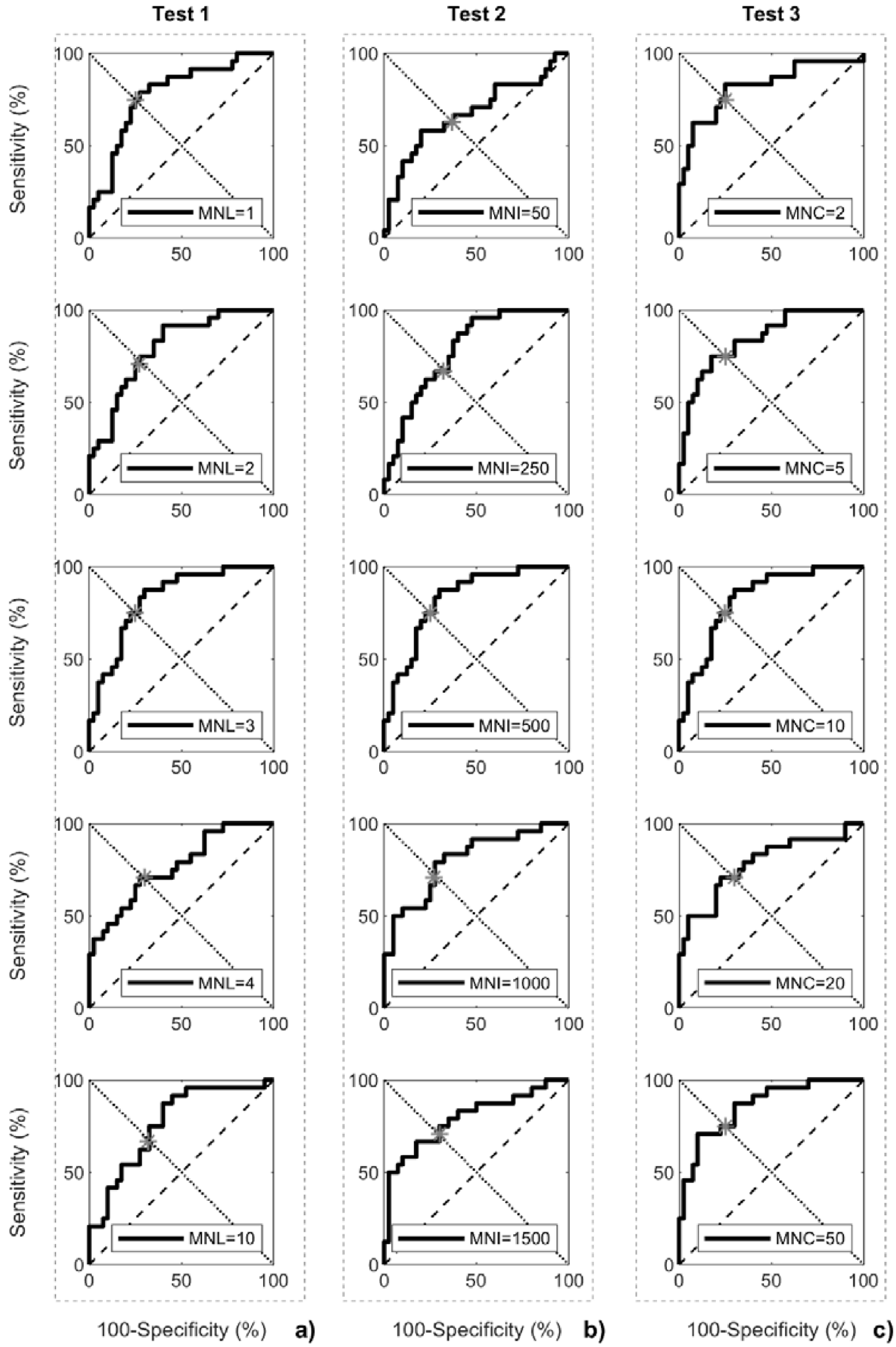

Fig. 3. Receiver operating characteristics (ROCs) obtained when performing Test 1 (panel a), Test 2 (panel b) and Test 3 (panel c), with a varying maximum number of hidden layers (MNL), a varying maximum number of initializations (MNI) and with a varying maximum number of confirmations (MNC), respectively. The operating points for which sensitivity equals specificity ($OP_{Se=Sp}$) are indicated with '×'.

was validated by using the train/test split validation procedure, guaranteeing the realization of a unique tested decision support system.

Results (Table 2) indicate that ANNs constructed by the RS&LP are more stable in terms of AUC and CI than in terms of structure. This is an expected and desirable finding; indeed, RS&LP was designed to automatically create an ANN by recursively alternating structuring and learning phases to optimize correctness of output classification, without considering a-priori architecture. The best parameter configuration is those having MNL, MNI and MNC equal to 3, 500 and 50, respectively. This combination of parameters may not be the optimal one, indeed it would be it in case of parameter independence. Nevertheless, it associates with a high value of AUC (86 %) and thus guarantees a good clinical performance. Additionally, this optimized combination of parameter values was associated with the longest LCT (212 h, 44 min and 11 s),

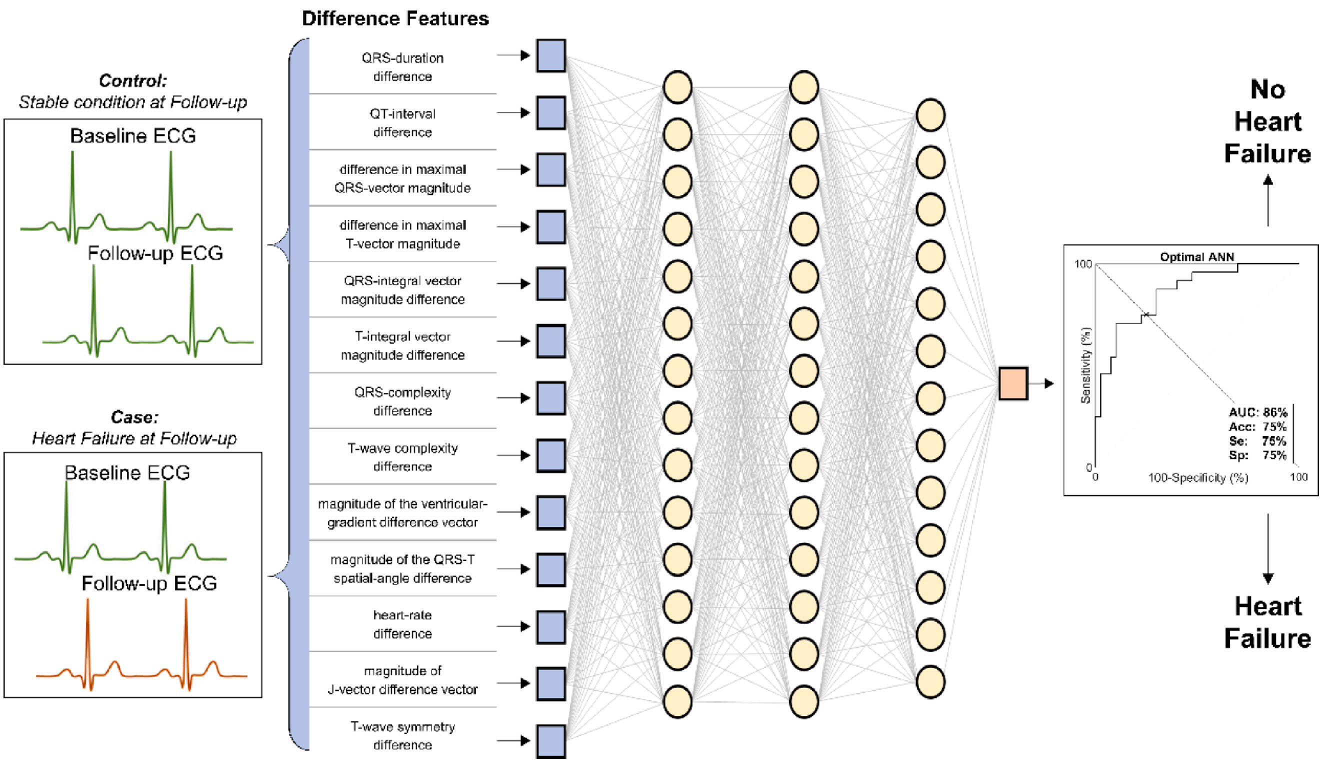

Fig. 4. Artificial neural network obtained with the optimized configuration of repeated structuring & learning procedure parameters for automatic diagnosis of newly emerged heart failure with [14,14,13] architecture and associated area under the curve (AUC) of 86%.

mostly due to the high values of the parameters, particularly of MNI. However, once created with the optimized combination of parameter values, the ANN performed testing classification fast (TCT = 9 ms) suggesting its possible use in clinical applications, even in real-time scenarios.

The best ANN was identified based on AUC and not on Acc, Se, and Sp. Indeed, computation of the latter requires choice of an operating point on the ROC, choice that should be left to the clinician and could vary depending on the clinical condition, anamnesis, and HF risk class of the patient. As an example, we reported Acc, Se, and Sp value relative to the commonly used operating point for which Se equals Sp; this value was 75 % for the optimized ANN.

In general, definition of the ANN architecture is critical, independently by the method applied to construct it. Indeed, use of too few neurons and/or hidden layers may lead to underfitting, whereas use of too many neurons and/or hidden layers may lead to overfitting [32]. Well defined rules for architecting ANNs have not been drawn up yet and definition of the appropriate number of neurons and layers still requires several trials and computations. RS&LP does not require an a priori definition of the architecture of the ANN to be created; rather, it adaptively defines it by continuously optimizing classification correctness. ANN growth is encouraged by trying different candidate ANNs (thus avoiding underfitting) and discouraged by imposing that candidate ANNs must improve performance (thus avoiding overfitting). The RS&LP underlying hypothesis is that there may exist several ANN architectures that may lead to the best possible classification; the reached optimized architecture depends on random initializations.

Several constrictive algorithms have been previously presented in the literature [33,34] but, to our knowledge, none for clinical or healthcare applications. The constructive nature of the RS&LP, that uniquely iteratively optimizes ANN architecture and its weights and biases, makes it particularly suitable to be applied to both relatively small databases, like the one used here or in our previously works [10,11], as well as to big data. To further avoid generalization problems due to the small size of the available dataset, in the present study the constructed ANN was also evaluated on the validation dataset where we applied the early-stopping criteria and selected suboptimized ANN as the one with the highest AUC among the 100 ANNs created with different random neuron initializations. Many clinical databases are limited in size and their statistical modelling is often rather conventional and miss ANN flexibility to handle non-linear interactions between features. Our present and previous applications of RS&LP to the same HF database [11] indicate that RS&LP performance (AUC = 86 %) is superior to that of logistic regression (AUC = 61 %) and of standard ANN method with a-priori fixed architecture (AUC = 83 %). When applied to databases of small size, the ANN architecture obtained with the RS&LP typically includes a low number of layers, but in case of big data applications, the RS&LP has the potentiality to create more complex ANN architectures able to manage different types of clinical data.

A qualitative comparison of the RS&LP performance against that of other machine learning approaches [13–20] aiming to detect HF by using cardiac signals is reported in Table 3. The studies differed in terms of used algorithm, presence of clinical confounders, samples size of analyzed populations and signals in input of the classifier. Many studies show very high performance; however, their clinical applicability could be limited due to lack of interpretability and explainability (cardiac signals [13–16,18–20] instead of features are used as input for the classifiers). The need of "Explainable Artificial Intelligence"[35] is a priority in clinical/ healthcare applications where machine learning approaches should not only perform automatic diagnosis of a pathology, but also explain why that classification was provided. Use of features (as done in the present study and in [17]) instead of signals as input of the classifier is often preferred in these cases. Indeed, if properly selected, the features have a physiological meaning that allows clinicians to interpret the results provided by the automatic analysis. Additionally, in real scenarios patients may be affected by comorbidities so that the capability to discriminate HF patients from subjects with normal sinus rhythm only [13–18] appears reductive. Anyway, considering the high versatility of ANN, future studies will aim to implement the RS&LP for the structuring and learning of convolutional and recurrent neural

Table 3
Qualitative comparison of machine learning approaches for heart failure detection from cardiac signals.

Ref.	Algorithm	Clinical Confounders	Data Sample Size	Input	Interpretability	Results
[13]	K-nearest neighbours and convolutional neural networks	No	73	ECG	No	Acc = 98.97 %
[14]	Support vector machine	No	76	ECG	No	Acc = 99.66 %
[15]	Support vector machine	No	33	64 ECG samples	No	Acc = 97.27 %
[16]	Random forest classifier	No	63	ECG	No	Acc = 99.86 %
[17]	Deep neural networks	No	55,163	Demographic and electrocardiographic features	Features	AUC = 89 %
[18]	Long short-term memory	No	156 divided into two databases: 1. 73 2. 83	RR time series	No	Acc = 98.9 % Acc = 87.6 %
[19]	Convolutional neural networks	Comorbidities in some patients: diabetes mellitus, hypercholesterolemia, renal disease, hypertension, coronary artery disease, myocardial infarction.	163,892	ECG	No	AUC = 89 %
[20]	Convolutional neural networks	Comorbidities in some patients: hypertension.	40	ECG	No	Acc > 99.8 % in several experiments
This work	Neural Networks	Comorbidities in all patients: myocardial infarction.	129	13 serial ECG difference features	Features	AUC = 86 %

Acc: accuracy; AUC: area under the receiver operating curve.

networks and test its performance in more complex clinical scenarios, always guaranteeing clinical interpretability.

It is finally important to observe that, as previously said, here RS&LP was optimized to work on serial electrocardiography to mimic and potentiate the diagnostic procedure adopted by clinicians while ensuring interpretability. However, in some practical cases baseline ECG may not be available. In those cases, RS&LP could still be thought as a tool to discriminate HF occurrence, but with an architecture and a combination of parameter values that should be specific for that application. Determination of these architecture and combination of parameter values is beyond the scope of this paper but will be matter of future studies.

6. Conclusion

Automatic diagnosis of newly emerged heart failure can occur through our optimized supervised fully connected artificial neural network created using the repeated structuring & learning procedure that can thus be proposed as a useful diagnostic tool for the clinical practice.

Funding

This research received no external funding.

CRediT authorship contribution statement

Agnese Sbrollini: Conceptualization, Methodology, Writing – original draft, Visualization. **Maddalena Barocci:** Software, Validation. **Martina Mancinelli:** Software, Validation. **Michele Paris:** Software, Validation. **Simone Raffaelli:** Software, Validation. **Ilaria Marcantoni:** Formal analysis, Visualization. **Micaela Morettini:** Investigation, Resources. **Cees A. Swenne:** Methodology, Data curation. **Laura Burattini:** Writing – review & editing, Supervision, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

References

[1] M. Metra, J.R. Teerlink, Heart failure, Lancet. 390 (10106) (2017) 1981–1995, https://doi.org/10.1016/S0140-6736(17)31071-1.
[2] J.JV. McMurray, M.A. Pfeffer, Heart failure, Lancet. 365 (9474) (2005) 1877–1889, https://doi.org/10.1016/S0140-6736(05)66621-4.
[3] P. Ponikowski, A.A. Voors, S.D. Anker, H. Bueno, J.G.F. Cleland, A.J.S. Coats, V. Falk, J.R. González-Juanatey, V.-P. Harjola, E.A. Jankowska, M. Jessup, C. Linde, P. Nihoyannopoulos, J.T. Parissis, B. Pieske, J.P. Riley, G.M.C. Rosano, L. M. Ruilope, F. Ruschitzka, F.H. Rutten, P. van der Meer, ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure, Eur. Heart J. 37 (2016) (2016) 2129–2200, https://doi.org/10.1093/eurheartj/ehw128.
[4] B. Ziaeian, G.C. Fonarow, Epidemiology and aetiology of heart failure, Nat. Rev. Cardiol. 13 (6) (2016) 368–378, https://doi.org/10.1038/nrcardio.2016.25.
[5] K. Ng, S.R. Steinhubl, C. deFilippi, S. Dey, W.F. Stewart, Early Detection of Heart Failure Using Electronic Health Records, Circ. Cardiovasc. Qual. Outcomes. 9 (6) (2016) 649–658, https://doi.org/10.1161/CIRCOUTCOMES.116.002797.
[6] W.R. Harlan, A. Graybiel, R.E. Mitchell, A. Oberman, R.K. Osborne, Serial electrocardiograms: Their reliability and prognastic validity during a 24-yr period, J. Chronic Dis. 20 (11-12) (1967) 853–867, https://doi.org/10.1016/0021-9681(67)90023-9.
[7] Z. Gao, Y. Li, Y. Sun, J. Yang, H. Xiong, H. Zhang, X. Liu, W. Wu, D. Liang, S. Li, Motion Tracking of the Carotid Artery Wall From Ultrasound Image Sequences: a Nonlinear State-Space Approach, IEEE Trans. Med. Imaging. 37 (1) (2018) 273–283, https://doi.org/10.1109/TMI.2017.2746879.
[8] W. Li, J. Li, Local Deep Field for Electrocardiogram Beat Classification, IEEE Sens. J. 18 (4) (2018) 1656–1664, https://doi.org/10.1109/JSEN.2017.2772031.

[9] C.C. Haar, R.J.G. Peters, J. Bosch, A. Sbrollini, S. Gripenstedt, R. Adams, E. Bleijenberg, C.J.H.J. Kirchhof, R. Alizadeh Dehnavi, L. Burattini, R.J. Winter, P. W. Macfarlane, P.G. Postema, S. Man, R.W.C. Scherptong, M.J. Schalij, A.C. Maan, C.A. Swenne, An initial exploration of subtraction electrocardiography to detect myocardial ischemia in the prehospital setting, Ann. Noninvasive Electrocardiol. 25 (3) (2020), https://doi.org/10.1111/anec.v25.310.1111/anec.12722.

[10] A. Sbrollini, M. de Jongh, C. Cato ter Haar, R. W Treskes, S. Man, L. Burattini, C. A. Swenne, Serial ECG Analysis: Absolute Rather Than Signed Changes in the Spatial QRS-T Angle Should Be Used to Detect Emerging Cardiac Pathology, in: Comput. Cardiol. (2010)., 2018. doi:10.22489/CinC.2018.099.

[11] A. Sbrollini, M.C. De Jongh, C.C. Ter Haar, R.W. Treskes, S. Man, L. Burattini, C. A. Swenne, Serial electrocardiography to detect newly emerging or aggravating cardiac pathology: a deep-learning approach, Biomed. Eng. Online. 18 (2019) 15, https://doi.org/10.1186/s12938-019-0630-9.

[12] C.R. Olsen, R.J. Mentz, K.J. Anstrom, D. Page, P.A. Patel, Clinical applications of machine learning in the diagnosis, classification, and prediction of heart failure, Am. Heart J. 229 (2020) 1–17, https://doi.org/10.1016/j.ahj.2020.07.009.

[13] U.R. Acharya, H. Fujita, S.L. Oh, Y. Hagiwara, J.H. Tan, M. Adam, R.S. Tan, Deep convolutional neural network for the automated diagnosis of congestive heart failure using ECG signals, Appl. Intell. 49 (1) (2019) 16–27, https://doi.org/10.1007/s10489-018-1179-1.

[14] A.A. Bhurane, M. Sharma, R. San-Tan, U.R. Acharya, An efficient detection of congestive heart failure using frequency localized filter banks for the diagnosis with ECG signals, Cogn. Syst. Res. 55 (2019) 82–94, https://doi.org/10.1016/j.cogsys.2018.12.017.

[15] K.Y.K. Liao, C.C. Chiu, S.J. Yeh, A novel approach for classification of congestive heart failure using relatively short-term ECG waveforms and SVM Classifier, in: Lect. Notes Eng. Comput. Sci., 2015: pp. 47–50.

[16] Z. Masetic, A. Subasi, Congestive heart failure detection using random forest classifier, Comput. Methods Programs Biomed. 130 (2016) 54–64, https://doi.org/10.1016/j.cmpb.2016.03.020.

[17] J. Kwon, K.-H. Kim, K.-H. Jeon, H.M. Kim, M.J. Kim, S.-M. Lim, P.S. Song, J. Park, R.K. Choi, B.-H. Oh, Development and Validation of Deep-Learning Algorithm for Electrocardiography-Based Heart Failure Identification, Korean Circ. J. 49 (2019) 629, https://doi.org/10.4070/kcj.2018.0446.

[18] L. Wang, X. Zhou, Detection of Congestive Heart Failure Based on LSTM-Based Deep Network via Short-Term RR Intervals, Sensors. 19 (2019) 1502, https://doi.org/10.3390/s19071502.

[19] Z.I. Attia, S. Kapa, F. Lopez-Jimenez, P.M. McKie, D.J. Ladewig, G. Satam, P. A. Pellikka, M. Enriquez-Sarano, P.A. Noseworthy, T.M. Munger, S.J. Asirvatham, C.G. Scott, R.E. Carter, P.A. Friedman, Screening for cardiac contractile dysfunction using an artificial intelligence–enabled electrocardiogram, Nat. Med. 25 (1) (2019) 70–74, https://doi.org/10.1038/s41591-018-0240-2.

[20] W. Yang, Y. Si, D.i. Wang, G. Zhang, X. Liu, L. Li, Automated intra-patient and inter-patient coronary artery disease and congestive heart failure detection using EFAP-Net, Knowledge-Based Syst. 201-202 (2020) 106083, https://doi.org/10.1016/j.knosys.2020.106083.

[21] G. King, L. Zeng, Logistic Regression in Rare Events Data, Polit. Anal. 9 (2) (2001) 137–163, https://doi.org/10.1093/oxfordjournals.pan.a004868.

[22] M.F. Møller, A scaled conjugate gradient algorithm for fast supervised learning, Neural Networks. 6 (4) (1993) 525–533, https://doi.org/10.1016/S0893-6080(05)80056-5.

[23] F. Dario Baptista, S. Rodrigues, F. Morgado-Dias, Performance comparison of ANN training algorithms for classification, in: 2013 IEEE 8th Int. Symp. Intell. Signal Process., IEEE, 2013: pp. 115–120. doi:10.1109/WISP.2013.6657493.

[24] B. Sharma, P.K. Venugopalan, Comparison of Neural Network Training Functions for Hematoma Classification in Brain CT Images, IOSR J. Comput. Eng. 16 (1) (2014) 31–35, https://doi.org/10.9790/0661-16123135.

[25] V.K. Garg, R.K. Bansal, Comparison of neural network back propagation algorithms for early detection of sleep disorders, in: 2015 Int. Conf. Adv. Comput. Eng. Appl., IEEE, 2015: pp. 71–75. doi:10.1109/ICACEA.2015.7164648.

[26] L. Prechelt, Early Stopping — But When?, in: G. Montavon, G.B. Orr, K.-R. Müller (Eds.), Neural Networks Tricks Trade. Lect. Notes Comput. Sci., second, Springer, Berlin, 2012: pp. 53–67. doi:10.1007/978-3-642-35289-8_5.

[27] S. Man, A.C. Maan, M.J. Schalij, C.A. Swenne, Vectorcardiographic diagnostic & prognostic information derived from the 12-lead electrocardiogram: Historical review and clinical perspective, J. Electrocardiol. 48 (4) (2015) 463–475, https://doi.org/10.1016/j.jelectrocard.2015.05.002.

[28] M.C. De Jongh, A. Sbrollini, A.C. Maan, E.T. Van der Velde, M.J. Schalij, C.A. Swenne, Progression towards Heart Failure after Myocardial Infarction Is Accompanied by a Change in the Spatial QRS-T Angle, in: Comput. Cardiol. (2010)., 2017. doi:10.22489/CinC.2017.292-342.

[29] H.H.M. Draisma, C.A. Swenne, H. van de Vooren, A.C. Maan, B. Hooft van Huysduynen, E.E. van der Wall, M.J. Schalij, LEADS: an interactive research oriented ECG/VCG analysis system, in: Comput. Cardiol. 2005, IEEE, 2005: pp. 515–518. doi:10.1109/CIC.2005.1588151.

[30] D. Marinucci, A. Sbrollini, I. Marcantoni, M. Morettini, C.A. Swenne, L. Burattini, Artificial Neural Network for Atrial Fibrillation Identification in Portable Devices, Sensors. 20 (2020) 3570, https://doi.org/10.3390/s20123570.

[31] E.R. DeLong, D.M. DeLong, D.L. Clarke-Pearson, Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach, Biometrics. 44 (3) (1988) 837, https://doi.org/10.2307/2531595.

[32] A.C. Ian Goodfellow, Yoshua Bengio, Front Matter, Elsevier (2014), https://doi.org/10.1016/B978-0-12-391420-0.09987-X.

[33] K.M.R. Alam, N. Siddique, H. Adeli, A dynamic ensemble learning algorithm for neural networks, Neural Comput. Appl. 32 (12) (2020) 8675–8690, https://doi.org/10.1007/s00521-019-04359-7.

[34] O. Irsoy, E. Alpaydin, Continuously Constructive Deep Neural Networks, IEEE Trans. Neural Networks Learn. Syst. 31 (4) (2020) 1124–1133, https://doi.org/10.1109/TNNLS.2019.2918225.

[35] D. Gunning, M. Stefik, J. Choi, T. Miller, S. Stumpf, G.-Z. Yang, XAI—Explainable artificial intelligence, Sci. Robot. 4 (2019) eaay7120, https://doi.org/10.1126/scirobotics.aay7120.