Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2016) 24: 3335 – 3344
© TÜBİTAK
doi:10.3906/elk-1403-117

Predicting acute hypotensive episode by using hybrid features and a neuro-fuzzy network

Marzieh ABBASINIA, Fardad FAROKHI*, Shahram JAVADI
Electrical Engineering Department, Islamic Azad University Central Tehran Branch, Tehran, Iran

Received: 13.03.2014 • **Accepted/Published Online:** 05.07.2014 • **Final Version:** 20.06.2016

Abstract: This paper presents an approach for acute hypotensive episode (AHE) time series forecasting based on hybrid feature space and a neuro-fuzzy network. Prediction was accomplished through a combination of time domain and wavelet features by using six vital time series of each patient, obtained from MIMIC-II and available in the context of the Physionet-Computers in Cardiology 2009 Challenge. At first, statistical time domain features were used and then the wavelet coefficient was utilized for extracting time scale features. Further UTA feature selection was applied and 30 effective features were determined and achieved to predict AHE with 96.30 accuracy 1.5 h before AHE onset.

Key words: Acute hypotensive episode (AHE), prediction, neuro-fuzzy network (NF), wavelet transform, feature selection, mean arterial blood pressure (MAP)

1. Introduction

Time series forecasting is an important area of research and has many applications in various fields, such as engineering, economics, financial management, and medicine. The most common time series models include:

1) Linear model: autoregressive integrated moving average (ARIMA)

2) Nonlinear model: neural network models and fuzzy system models [1].

Recently, the application of artificial neural networks (ANNs) in time series forecasting has been growing and ANNs are used as important tools for prediction of nonlinear and nonstationary time series. Prediction of physiological time series helps physicians to choose more effective treatment and prevent occurrence of physiological disorders such as hypertension and hypotension, through proper intervention [2]. Hypotension is a common occurrence in the intensive care unit in which the blood pressure is abnormally low. It can deprive the brain and other vital organs of oxygen and nutrients, damage the organ, or eventually cause death, and so early detection of hypotension is essential [3].

In [4] the shape of atrial blood pressure (ABP) waveforms taken from invasive beat-by-beat blood pressure measurement devices was examined and parametric and nonparametric methods were suggested to characterize ABP before AHE.

In [5] a similarity-based searching and pattern matching algorithm was presented that were classified and predicted AHE. In [6] the variance of the ABP signal and the variance of the ABP wave's slope were used as relevant features to predict AHE.

AHE prediction was the subject of the Physionet and Computers in Cardiology for 2009 Challenge and

*Correspondence: f_farokhi@iauctb.ac.ir

a wide range of researchers were worked on it. In [7] generalized regression neural network models integrated in to a multimodel structure was proposed for prediction of MAP signal and diagnosis of AHE. Considering other vital signals in this methodology may reduce the number of multimodels or increased the horizon of prediction.

In [8] AHE was predicted by using Haar wavelet decomposition and reconstruction together with a feed-forward neural network structure and achieved effective prediction by using only ABP signals. In [2] a local linear neuro-fuzzy model was used in order to predict mean arterial blood pressure (MAP) time series. In [9], the predicting model was developed by using median and maximum of wavelet coefficient based on support vector machine (SVM). In [10] a hybrid methodology based on discrete wavelet transform and time delay embedding neural network structure was applied to MAP time series and estimated prediction. In [11] effective time domain features were selected by genetic algorithm and prediction performed with a simple linear support vector classifier (LSVC) that can find the optimal linear boundary between two classes.

All mentioned publications presented the best AHE prediction over a one hour forecast horizon and used a long period of ABP signals. Accordingly, in order to increase forecast horizon and reduce the observation window as compared to the previous studies, this paper attempts to consider other significant time series in addition to ABP signals and improving the generalizing ability of the intelligent network by considering a large scale database. All segments of the signal are depicted in Figure 1.

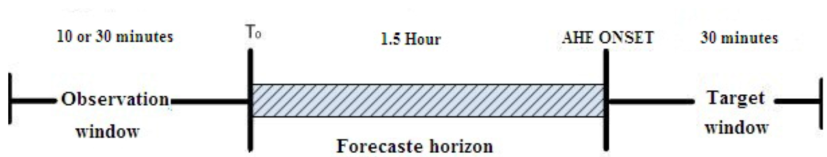

Figure 1. Data segmentation.

This paper presents an algorithm that is based on a hybrid time and time-scale feature vector and a powerful neuro-fuzzy classifier. This algorithm used a much larger database to validate and generalize the prediction. Section 2 of this study presents three different classifiers for AHE prediction: multilayer perceptron (MLP) network, k nearest neighbor (KNN), and neuro-fuzzy network. In the third section, the best structure is chosen and feature selection is done in order to maximize the forecasting horizon. There are four different types of features categorized into four groups named statistical time domain, time domain, time scale, and statistical time scale features.

2. Procedure methodology

There are various classifier structures for predicting time series. In this research, three different classifiers were tested using a hybrid feature vector over different forecasting horizon in order to determine maximum achievable effective forecast horizon for hypotension event prediction.

2.1. Feature extraction

Six vital time series that are more easily available for patients in the MIMIC-II Database were selected: heart rate (HR), diastolic blood pressure (DBP), systolic blood pressure (SBP), mean arterial blood pressure (MAP), SPO_2 saturation, and respiration rate (RESP).

Each signal was divided into three time intervals as described in section 3. Prediction was done based on the best observation window that is before AHE onset and which is named T_0. Features were extracted from each case base on the observation window of six time series.

The observation window of each signal was divided into five minutes windows for extracting features. Mean, median, variance, and skewness as important statistical features are extracted from the five minute selected windows. Mean and median represent the magnitude of each physiologic variable, variance describes the variability of ones, and skewness is the third moment of amplitude distribution [12].

Another type of feature extracted is the discrete wavelet transform features depicting the signal in both time and frequency (scale) domain. Wavelet transform is an effective tool for nonstationary signal processing.

Each signal can decompose to different ranges of frequency signals that are known as wavelet coefficients. The original signal passes through low-pass and high-pass filters and emerges approximation and details as shown in Figure 2.

Multiple-level decomposition was applied in this paper and approximations were decomposed in turn and it is depicted in Figure 3, so that one signal was broken down into many lower-resolution components. $A_L(t)$ and $D_L(t)$ are approximation and the detail coefficients at level L, respectively [13,14].

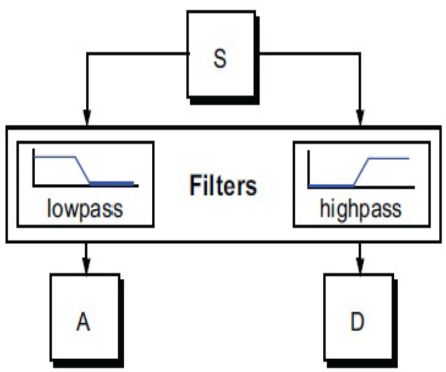

Figure 2. Wavelet transform.

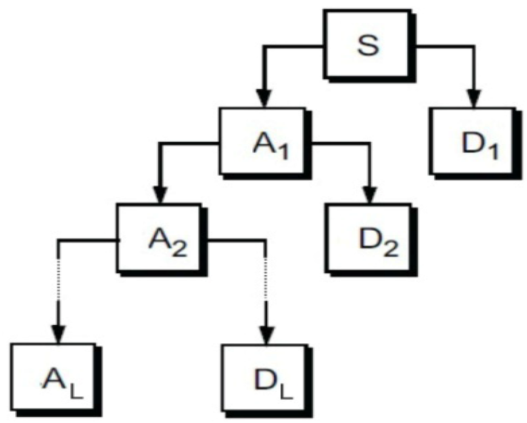

Figure 3. Multiple-level decomposition.

In previous studies, Haar wavelet [8] or Daubechies wavelet [9] was employed to extracting features and they employed only MAP signal. [15] stated that Haar wavelet is not smooth and suitable enough for nonstationary signals. In [16] it was shown that forecasting accuracy can be approved using Daubechies wavelet as compared to Haar.

Daubechies basis wavelet was chosen here and statistical analysis was performed on $A_L(t)$ and $D_L(t)$ of each time series.

2.2. Classification

The first classifier used was the well-known general model of multilayer perceptron (MLP), which is a feed forward artificial neural network and in this approach the back propagation technique was used for training. The structure of the MLP is shown in Figure 4. The MLP network is composed of hidden layers with hyperbolic tangent activation function and an output layer. The exact number of the hidden neurons was determined by trial and error and initial weights were selected randomly.

The learning approach was implemented and corrected weights from the last layer to the first layer until satisfying the stop condition. The learning process was performed for prespecified iteration or until the error was less than a prespecified number. Finally, the output network was compared to the expected result [17].

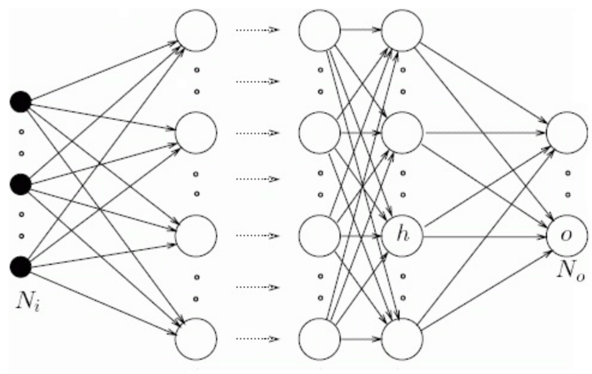

Figure 4. MLP structure.

The second classifier used was the K nearest neighbor (KNN). This classifier uses an instance-based method that classifies examples in the test set based on their similarity with examples in the training set. In this approach, KNN classified examples between two classes described in section 3 and compared them with the expected output.

The third classifier is a neuro-fuzzy network. There exist two approaches for designing a fuzzy system from input-output pairs:

1) Fuzzy IF-THEN rules are generated from input-output pairs, and the fuzzy system is then constructed according to choice of fuzzy inference engine, fuzzifier and defuzzifier.

2) The structure of the fuzzy system is specified first and some parameters are free to change and they are obtained according to the input-output pairs in training.

In this paper the second approach was applied. The used fuzzy structure is defined with product inference engine, singleton fuzzifier, center average defuzzifier, and Gaussian membership function. The output of the network is evaluated using Eq. (1).

$$f(x) = \frac{\sum_{l=1}^{M} \bar{y}^l \left[\prod_{i=1}^{n} \exp(-(\frac{x_i - \bar{x}_i^l}{\sigma_i^l})^2) \right]}{\sum_{l=1}^{M} \left[\prod_{i=1}^{n} \exp(-(\frac{x_i - \bar{x}_i^l}{\sigma_i^l})^2) \right]} \tag{1}$$

where M is the number of rules and $\bar{y}^l$, $\bar{x}_i^l$, σ_i^l are free parameters for IF-THEN parts of the rules. Figure 5 depicts the fuzzy system $f(x)$ of Eq. (1) such as a feed forward neuro-fuzzy network.

Designing the fuzzy system through gradient descent learning was done through the following steps:

Step1: the fuzzy system was considered as Eq. (1) and the initial parameters were randomly selected and the number of rules selected by trial and error (M).

Step2: for each input-output pair (x_0^p, y_0^p), p=1, 2,..., n, where p is the instance number of training set, for q=1,2,..., M each parameter is calculated layer by layer using Eq. (2) to Eq. (5).

$$z^l = \prod_{i=1}^{n} \exp(-(\frac{x_{0i}^p - \bar{x}_i^l(q)}{\sigma_i^l(q)})^2) \tag{2}$$

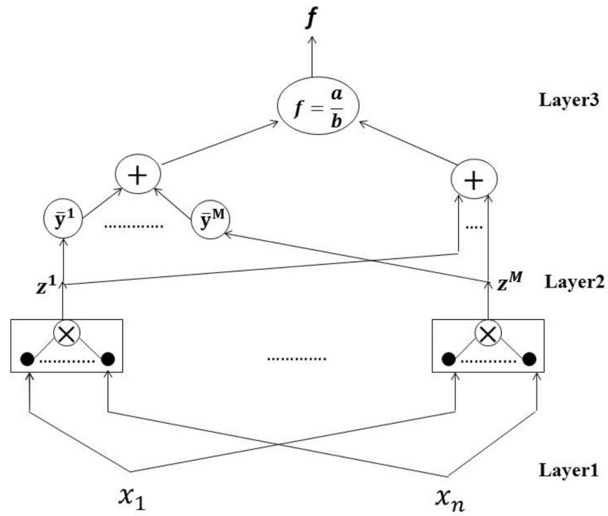

Figure 5. Network representation of the fuzzy system.

$$b = \sum_{l=1}^{M} z^l \tag{3}$$

$$a = \sum_{l=1}^{M} \bar{y}^l(q)\, z^l \tag{4}$$

$$f = {}^{a}/_{b} \tag{5}$$

Step3: the free parameters were updated by using Eq. (6) to Eq. (8) for q=1, 2,..., M and i=1,2,..., n:

$$\bar{x}_i^l(q+1) = \bar{x}_i^l(q) - \alpha \frac{f-y}{b}\left(\bar{y}^l(q) - f\right) z^l \frac{2\left(x_{0i}^p - \bar{x}_i^l(q)\right)}{\sigma_i^{l2}(q)} \tag{6}$$

$$\bar{y}^l(q+1) = \bar{y}^l(q) - \alpha \frac{f-y}{b} z^l \tag{7}$$

$$\bar{\sigma}_i^l(q+1) = \bar{\sigma}_i^l(q) - \alpha \frac{f-y}{b}\left(\bar{y}^l(q) - f\right) z^l \frac{2\left(x_{0i}^p - \bar{x}_i^l(q)\right)}{\sigma_i^{l3}(q)} \tag{8}$$

where i is the number of features used, l is the number of rules l=1,2,..., M, and α is the learning rate and it is between 0 and 1.

Step4: if the error $|f - y_0^p|$ or mean square error (MSE) gets less than a prespecified value or until prespecified iteration or gets the best result before overtraining, which is used here, free parameters and M are saved and training stage finishes, else go step 5.

Step5: the process was repeated from step 2 by presenting inputs of the training set to the fuzzy system for the next iteration.

The learning algorithm applied here performed such as back-propagation error procedure; therefore, this algorithm is called the error back-propagation learning algorithm [18].

3. Result

3.1. Data

The presented algorithm was tested over the MIMIC-II database, which is composed of vital signals and is available at www.Physionet.org. Each selected record consisted of three time intervals: 1) 10 or 30 minute observation window 2) forecast window defined as the 1.5 hour period before T_0 or AHE onset 3) 30 minute target window. The target window of each record was classified into "Hypotensive" or "Control". Records in group H contain an episode of acute hypotension beginning during the forecast window and those in group C contain no episodes of acute hypotension during their stay in hospital or forecast window according to the MIMIC-II database. An AHE is defined as any period of 30 minutes or more during which at least 90% of the mean arterial blood pressure measurements were at or below 60 mmHg.

3.2. Training and test datasets

180 patients of both groups are randomly selected from MIMIC-II. Either the training set or test set included 90 cases. Each record includes six time series (HR, SBP, DBP, MAP, RESP, SPO_2) which has at least 60 minutes' data before T_0. All signals' sampling intervals are 60 s. The missing information of time series was obtained by linear interpolation. The range of variability of every signal was normalized to the [0,1] interval. Ten minutes' data before T_0 were used for time domain feature extraction and statistical time scale features are extracted from the last 30 minutes of the original signals for the 4th decomposition level and the Daubechies Wavelet transform (db4) was selected for appropriate content of the features in this level.

3.3. Acute hypotensive episode prediction result

The error back-propagation neuro-fuzzy structure proposed in section 2 has three layers: input layer, hidden layer, and output layer. The best number of hidden neurons was selected by trial and error.

Table 1 indicates the best predictive performance results for three networks. These algorithms were utilized for different forecast horizon as 1 hours, 1.5 hours and 2 hours. The best effective forecast horizon was selected 1.5 hours and 102 time domain features were used for all algorithms.

Table 1. Results of different neural networks for different forecast horizon.

Observation window: 10 min Forecast horizon: 60 min				
Algorithm	SE	SP	ACC	MSE
MLP $N_{102,12,8,10,2}$	81.48	88.89	85.19	0.002
KNN, K=3	81.48	70.37	75.93	3.4×10^{-4}
NF $N_{102,22,2}$	85.19	88.89	87.04	0.002
Observation window: 10 min Forecast horizon: 90 min				
Algorithm	SE	SP	ACC	MSE
MLP $N_{102,18,14,2,2}$	85.19	81.48	83.33	0.15
KNN , K=4	70.37	85.19	77.78	0.19
NF $N_{102,30,2}$	88.89	85.19	87.04	0.004
Observation window: 10 min Forecast horizon: 120 min				
Algorithm	SE	SP	ACC	MSE
MLP $N_{102,44,16,11,2}$	70.37	75.93	79.63	0.18
KNN , K=2	70.37	74.07	72.22	6.3×10^{-6}
NF $N_{102,30,2}$	85.19	81.48	83.33	8.1×10^{-4}

According to the Table 1, each algorithm used the same feature vector, which consisted of time domain features as mean, median, variance, and skewness of the 2 last consequent five minutes of each signal together with the original data of the 10 last minutes of the signals and described the result as accuracy, sensitivity and specificity, described as follows:

Sensitivity (SE) is the percentage of records correctly identified as group H, specificity (SP) is the percentage of records correctly identified as group C, and accuracy (ACC) is the percentage of true prediction.

$$\text{Sensitivity} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \times 100 \tag{9}$$

$$\text{Specificity} = \frac{\text{TN}}{(\text{TN} + \text{FP})} \times 100 \tag{10}$$

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})} \times 100 \tag{11}$$

The terms of TP, TN, FN, and FP are true positive, true negative, false negative, and false positive, respectively.

Because of the NF network's good results, it was selected for feature selection in order to improve its overall performance.

As stated in Table 2, when the time domain features were applied to the network, the result as compared to previous papers was 87.04% accuracy, while the observation window was reduced to 10 minutes, and it was increased to 90.74% by adding time scale features to the vector. In order to select effective and consistent features, the feature selection technique was used. The UTA approach was proposed in [19] and it is based on a trained network that was used.

Table 2. Feature selection and prediction result

Algorithm	Feature	SP	SE	ACC	MSE
NF $N_{\mathbf{102,30,2}}$	102 time domain	88.89	85.19	87.04	0.004
1stage UTA NF $N_{\mathbf{16,30,2}}$	16 time domain	88.89	85.19	87.04	0.006
2stage UTA NF $N_{\mathbf{8,28,2}}$	8 time domain	85.19	88.89	87.04	0.004
NF $N_{\mathbf{146,24,2}}$	8 time domain+ 138 wavelet	96.30	85.19	90.74	5×10^{-4}
1 stage UTA NF $N_{\mathbf{45,38,2}}$	4 time domain+ 41 wavelet	96.30	96.30	96.30	8×10^{-4}
2 stage UTA NF $N_{\mathbf{30,34,2}}$	3 time domain+ 27 wavelet	92.59	1	96.30	7×10^{-5}

In the UTA method, the theory of feature effectiveness is based on the fact that if a feature is constant for all the instances then it could not participate in classification because the classifier could not differentiate between instances using that feature. In the UTA algorithm the features are selected one by one and a constant, which is the mean value of that feature in all instances, is replaced instead of that feature. The FP+FN minus the previous value (the FP+FN using original features) could be zero, positive, or negative, which means that the selected feature is not efficient, efficient, or destructive, respectively. UTA was implemented in the test dataset.

If $\text{FP+FN}_{new} > \text{FP+FN}_{old}$ then feature is efficient.

If $\text{FP+FN}_{new} < \text{FP+FN}_{old}$ then feature is not efficient.

If $\text{FP+FN}_{new} = \text{FP+FN}_{old}$ then feature is destructive.

In this paper UTA first was implemented on time domain features in 2 stages and eventually 8 relevant time domain feature were selected from 102 features. These features were extracted from 5 time series (SBP, DBP, MAP, RESP, and SPO_2). After that, time scale features were added to them and accuracy was increased by using the new feature vector; then UTA was applied in two times in order to minimizing the feature vector. Figure 6 depicts the last two stages feature selection.

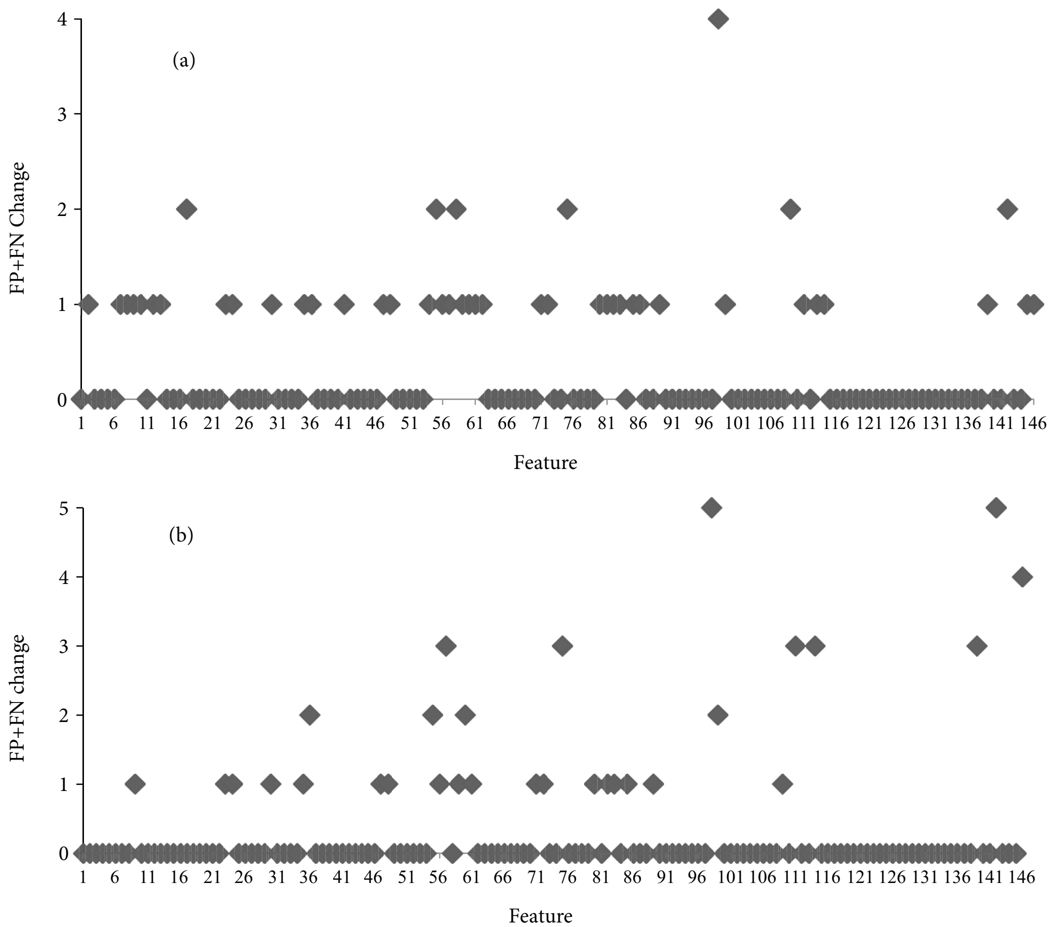

Figure 6. UTA algorithm on time scale and time domain features (a) 1Stage UTA in NF network (b) 2Stage UTA in NF network.

Optimal features were obtained from 5 time series: HR, SBP, DBP, MAP, and RESP, and the feature selection process reduced features from 146 to 30 effective and consistent features. None of the SPO_2 extracted features were determined as effective in AHE prediction.

Table 2 shows the results of the UTA process. The best neuro-fuzzy network has 30 features, which consisted of 3 time domain features that were extracted from SBP, DBP, and MAP signals and 27 time scale features extracted from HR, SBP, DBP, MAP, and RESP signals. Table 3 depicts the details of the best neuro-fuzzy network.

4. Comparison of the proposed method with other approaches

Table 4 shows a comparison of the best prediction between previous works, which are mentioned in the first section, and the proposed method. Most of them used training and test datasets in the Physionet-Computers in Cardiology 2009 Challenge, except [7–12] in which they selected large dataset from MIMIC-II. In this paper each sample is related to one patient and the first experience of AHE is predicted, but in [7–12] some records were divided into some segments as a sample in which both groups (H and C) may contain one patient.

Table 3. Results of the best algorithm.

Algorithm	Feature	Control class	Hypotensive class
		TP_1=25	TP_2=27
NF N30, 34, 2 with	3 time domain+	TN_1=27	TN_2=25
UTA feature selection	27 wavelet	FN_1=2	FN_2=0
		FP_1=0	FP_2=2

Table 4. Comparison between the different methods.

Method	Signal	SE	SP	ACC	Observation window	Forecast horizon
[7] Generalized regression neural network (GRNN)	MAP	-	78.4	82.8	-	60 min
[12] Principal Component Analysis and MLP neural network	MAP-DBP-SBP-HR	81.2	86.4	86.3	30 min	60 min
Proposed approach (time domain features and neuro-fuzzy network without feature selection)	MAP-DBP-SBP-HR-RESP-SPO2	88.89	85.19	87.04	10 min	90 min
[9] Wavelet transform and support vector machine classifier	MAP	81.25	-	92.96	60 min	60 min
[11] Feature selection through genetic algorithm and simple linear support vector classifier	MAP	92.40	94.42	93.65	90 min	60 min
[8] wavelet transform and feed-forward neural network	MAP	94.74	93.55	94	120 min	60 min
[10] wavelet transform and time delay embedding neural network	MAP	93.3	96	95	-	60 min
Proposed approach (Hybrid features with UTA feature selection and neuro-fuzzy network)	MAP-DBP-SBP-HR-RESP-SPO2	1	92.59	96.30	30 min	90 min

Table 3 depicts the ability of the proposed method before and after using hybrid features and feature selection in comparison with other methods. Some previous works stated ACC as the final result; comparison is done based on the best ACC while in the proposed method either ACC or MSE is calculated.

5. Conclusion

The concept of ANN is a powerful tool for time series forecasting. This paper investigated three different neural and neuro-fuzzy networks. The proposed method used time domain features together with time scale features in a hybrid vector to predict hypotension 1.5 hours before using neuro-fuzzy classifier. Further it utilized UTA feature selection to find effective features in order to minimize the feature vector size and improve the classification performance. The proposed model achieved 100% sensitivity, 92.59% specificity, and 96.30% accuracy. Using the hybrid effective features together with the neuro-fuzzy network led to a longer forecast horizon and reduced the observation window; also it has an acceptable accuracy compared to all the presented approaches as listed in Table 4. The presented approach, unlike previously presented researches, is generalized and proves its ability on the whole MIMIC-II and it may have practical application.

References

[1] Chen Y, Yang Bo, Dong J, Abraham A. Time-series forecasting using flexible neural tree model. Information Science 2005; 174: 219-235.

[2] Janghorbani A, Arasteh A, Moradi MH. Application of local linear neuro-fuzzy model in prediction of mean arterial blood pressure time series. 2010 17th Iranian conference of Biomedical Engineering (ICBME); November 2010; pp. 1-4.

[3] Moody GB, Lehman LH. Predicting acute hypotensive episodes. Computers in Cardiology 2009; pp. 541-544.

[4] Bassale J. Hypotension Prediction Arterial Blood Pressure Variability. Portland State University; 2001.

[5] Lehman LH, Saeed M, Moody GB, Mark RG. Similarity-based searching in multi-parameter time series databases. Computers in Cardiology 2008; pp. 653-656.

[6] Crespo C, Mcnames J, Aboy M, Bassale J, Ellenby M, Lai S, Goldstein B. Precursors in The Arterial Blood Pressure Signal To Episodes Of Acute Hypotension In Sepsis. Proceedings of the 16th International EURASIP Conference Biosignal 2002; 16: 206-208.

[7] Rocha T, Paredes S, Carvalho P, Henriques J. Prediction of Acute Hypotensive Episodes by means of neural network multi-model. Computer in Biology and Medicine 2011; 41: 881-890.

[8] Rocha T, Paredes S, Carvalho P, Henriques J, Harris M. Wavelet based Time series forecast with Application to Acute Hypotensive Episodes Prediction. Annual International conference of the IEEE Engineering in medicine and Biology society (EMBC); 2010; pp. 2403-2406.

[9] Wang ZH, Lai L, Xiong D, Wu X. Study on predicting Method for Acute Hypotensive Episodes based on wavelet Transform and Support Vector Machine. 3rd International Conference on Biomedical Engineering and Informatics (BMEI) 2010; 3: 1041-1045.

[10] Hoseinnia M, Sadr D. A Hybrid Approach for predicting Acute Hypotensive Episodes. 5th Symposium on Advances in Science & Technology; May 2011.

[11] Afsar FA. Prediction of Acute Hypotensive Episodes in patients Taking Pressor Medication using Modeling of Arterial Blood Pressure waveform. 2010 4th International Conference on Bioinformatics and Biomedical Engineering (ICBBE); June 2010; pp. 1-4.

[12] Lee J, Mark RG. An investigation of patterns in hemodynamic data of impending hypotension in intensive care. Biomedical engineering online; 2010.

[13] Misiti M, Misiti M, Oppenheim M, Poggi JM. Wavelet Toolbox for Use with MATLAB; 1996.

[14] Graps A. An introduction to wavelets. IEEE Computational Science and Engineering 1995; 2: 50-61.

[15] Chavan MS, Mastorakis N. Studies on Implementation of Harr and daubechies Wavelet for Denoising of Speech Signal. International Journal of Circuits, Systems and Signal Processing 2010; 4: 83-96.

[16] Ababneh, Wadi SAL, Ismail MT. Haar and Daubechies Wavelet Methods in modeling Banking Sector. International Mathematical Forum 2013; 8: 551-566.

[17] Haykin S. Neural Networks a Comprehensive Foundation. McMaster university, Published by Pearson Education.

[18] Wang LX. A Course in fuzzy system and control. Prentice-Hall, Englewood Cliffs, NJ, 1996.

[19] Utans J, Moody, Rehfuss S, Siegelmann H. Input Variable Selection for Neural Networks: Application to Predicting the US business Cycle. Proceedings of the IEEE/IAFE Computational Intelligence for Financial Engineering; 1995; pp. 118-122.