2020 International Conference on Electrical Engineering (ICEE)
September 25-27, 2020, Istanbul, Turkey

Heart Disease prediction using MLP and LSTM models

Mohamed Djerioui
LASS Laboratory, Department of Electronics
Faculty of Technology, University of M'sila
M'sila, Algeria
mohamed.djerioui@univ-msila.dz

Youcef Brik
LASS Laboratory, Department of Electronics
Faculty of Technology, University of M'sila
M'sila, Algeria
youcef.brik@univ-msila.dz

Mohamed Ladjal
LASS Laboratory, Department of Electronics
Faculty of Technology, University of M'sila
M'sila, Algeria
mohamed.ladjal@univ-msila.dz

Bilal Attallah
LASS Laboratory, Department of Electronics
Faculty of Technology, University of M'sila
M'sila, Algeria
bilal.attallah@univ-msila.dz

Abstract— **One of the key causes of premature disability and mortality in the world today is heart disease, which makes its prediction a vital problem in the field of healthcare systems. This work provides a contribution to the study and creation an intelligent system based on LSTM technique for heart disease prediction. A comparative study is presented between Multi Layer Perceptron (MLP) and Long Short Term Memory (LSTM) techniques in terms of accuracy and other predictive parameters for heart disease. The main aim is to develop an intelligent system based on LSTM technique for predicting heart disease in order to make an adapted decision to prevent and monitor heart disease and stroke. As it has better characteristics than those of the MLP technique, LSTM is shown to be the most effective technique for solving the aforementioned problems.**

Keywords-Heart disease; prediction; Multi Layer Perceptron; Long Short Term Memory

I. INTRODUCTION

Heart disease is one of the world's biggest causes of people's disability and premature death. An estimated 17.9 million deaths have occurred worldwide due to heart disease in 2016 [1], according to the World Health Organization. However, some main factors help us minimize the risk of heart disease as the regulation of blood pressure and cholesterol [2]. Indeed, whether there is a heart attack, angina, stroke, or heart failure, heart disease should not be diagnosed. Hence heart disease prediction is a delicate, dangerous, and very critical factor [3]. If done correctly, the medical personnel will do this to save lives. This process can be done by analyzing patient-reported data. Current healthcare programs usually use electronic health records to store this data. Advances in computer and information technology will solve these routine data for essential medical decision making [4].

Machine learning methods have also been thoroughly studied in healthcare systems to make decisions based on clinical evidence. Many researchers have used these methods to predict heart disease in this sense.

Actually, several types of research have addressed the development of a system for the prediction of intelligent heart disease using Machine learning methods. In [05], the authors contributed to the study of heart failure in a clinical decision support system. They compared the performance of various classifiers for machine learning, such as a neural network (NN), support vector machine (SVM), a system with fuzzy rules using the classification and regression tree (CART), and random forests (RF). The CART model and RF achieved the highest results with 87.6 percent accuracy. In [6], the authors proposed a method for improving prediction accuracy by defining key features and classifying them using a hybrid random forest. Abushariah et al. [7] developed two systems based on the Artificial Neural Network (ANN) and Neuro-Fuzzy methods to develop an automated method for diagnosing heart disease. Bouali and Akaichi [8] have used many machine learning techniques, such as: Baysian Network, Decision tree, Artificial Neural Network, Fuzzy pattern tree and SVM, to classify the Cleavland heart disease dataset using 10-fold-cross validation. Compared with other classifiers, SVM obtained the highest predictive precision. Otoom et al. [9] proposed a method for the identification and control of coronary artery disease, where three machine learning strategies such as: Bayes Net, SVM and Functional Trees are performed. The authors used WEKA tool for detection and SVM given the best technique with 85.1 per cent accuracy by applying feature selection. Therefore, RNNs are commonly used for sequence learning of nonlinear features. The Long Short-Term Memory Network (LSTM) is the most common model for neural network variant RNN at present. LSTM based on RNN solves the problem of gradient disappearance and RNN gradient eruption, which significantly enhances predictive accuracy [10]. At present, LSTM has been commonly used in many fields for prediction, such as artificial emotion-sensitive listening [11], aquaculture [12], speech recognition [13], machine translation [14], industrial processes [15], etc.

This work presents a comparative study between MLP and LSTM techniques in term of accuracy and other parameters for prediction in heart disease; the main objective is to set up an intelligent system based on LSTM technique for heart disease prediction in order to make an adapted decision to prevent and monitor heart disease and stroke. The LSTM is used for predicting the selected parameters in order to identify the presence/absence of heart disease. In fact, LSTM can perfectly classify data in binary problem. This makes it very adequate for heart disease prediction system that contains data of patient records with binary target, i.e. referring to the presence or absence of heart disease.

The rest of the paper is organized as follows: section two deals with the materiel and an overview of the used techniques including MLP and LSTM respectively. In section three, the experimental results are presented with details. In Section four, a comparison of the state-of-the-art methods is given. Finally, the conclusions drawn from this work are in Section five.

II. Material and Methods

A. Dataset description

The performance evaluation of this work is conducted on the famous cardiovascular dataset called the Heart UCI disease (University of California, Irvine, CA, USA). It has been collected from UCI machine learning repository [16]. This dataset contains in total 303 patient records with 76 attributes for each one, but only 14 of them are used for our evaluation to make our scores compared to previous works. Table I provides a brief description about the selected attributes and their proprieties. The last attribute serves as the prediction target that indicates the absence or presence of heart disease in a patient (0 or 1 value, respectively). Of the 303 records, 165 records with Target 1 and the rest for patients with Target 0.

In this dataset, patients from ages between 29 and 79 have been selected. Male patients are represented by 1 and female ones are represented by 0. Also, four types of chest pain 'Cp' have been considered such as typical type Angina, atypical type Angina, non-angina pain, and the Asymptomatic types. Each type of them is described by a value from 1 to 4, respectively. The next attribute 'trestbps' is the resting blood pressure measured at the hospital admission. 'Chol' is the blood cholesterol level. The parameter 'Fbs' is the fasting blood sugar level which is represented by 0 if the fasting blood sugar is less than 120 mg/dl and 1 if it is higher. 'Restecg' is the resting electrocardiographic results classified in three levels 0, 1, and 2. Besides, the attribute 'thalach' describes the maximum heart rate achieved. 'exang' is exercise-induced angina which is recorded as 1 if there is pain and 0 else. The attribute 'oldpeak' represents the ST depression induced by exercise relative to rest. Furthermore, the slope of the peak exercise ST segment has been recorded with the values 0, 1, and 2. The attribute 'ca' is the number of major vessels colored by fluoroscopy varying from 0 to 4. Finally, the attribute 'thal' represents the nature of the defect which can take an integer value from 0 to 3.

TABLE I. DATASET PARAMETER DISCRIPTION

N°.	Attributes	Description	Values
1	age	Patient age in years	29 to 79
2	sex	Patient gender	0=F, 1=M
3	cp	Chest pain type	1, 2, 3, 4
4	trestbps	Resting blood pressure (mm/Hg)	94 to 200
5	chol	Serum cholesterol (mg/dl)	126 to 564
6	fbs	Fasting blood sugar (mg/dl)	0, 1
7	restecg	Resting electrocardiographic results	0,1, 2
8	thalach	Maximum heart rate achieved	71 to 202
9	exang	Exercise induced angina	0, 1
10	oldpeak	ST depression induced by resting exercise	1 to 3
11	slope	The slope of peak exercise ST segment	1, 2, 3
12	ca	Major vessels colored by flourosopy	0 to 3
13	thal	Represents heart rate of the patient	1, 2, 3
14	num	Presence or absence of heart disease	0, 1

B. Multi-Layer Perceptron (MLP)

The most widely used predictive neural model is MLP. An MLP's basic structure contains an input layer, one or more hidden layer(s), and an output layer. Each layer consists of one or more neurons. The input is transformed to output through the activation function; hence, it is important to pay attention to transition functions such as activation function. It can be either a linear or nonlinear transition function. A transfer function is chosen based on unique problem needs [17]. The example of this model used in this application is shown in figure 1.

In such networks, a layer input is the output of the previous layer that is accomplished by the activation function. Therefore, the equation for calculating the output is as follows in each state.

$$a^{L+1} = \mathrm{f}^{L+1}(\mathrm{w}^{L+1} a^{L} + b^{L+1}) \tag{1}$$

Where L is the number of network's layers and a is the output vector in each layer. w denotes the weight of each synapse; b denotes a bias vector of the hidden layers and f represents an activation function.

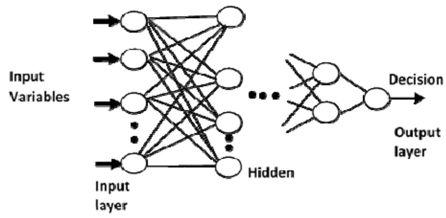

Figure 1. Example of Multilayer perceptron

A training method obtains the necessary weight matrices and the bias vectors of an MLP Neural Nets. The traditional training method relies on the backpropagation error framework[18]. The Mean Square Error (MSE), however, is commonly used as the objective function for model training [19]:

$$\text{MSE} = \min_{w_1, w_2, b_1, b_2} \frac{1}{M} \sum_{i=1}^{M} e_i^2 \tag{2}$$

where M is the number of data points in the training set; e_i denotes the deviation between the actual and the predict output variables.

C. *Long short-term memory (LSTM)*

The theoretical base theory of LSTM is influenced by the learning output of recurrent neural networks (RNN). The RNN is a neural network used to process data about the sequence [20]. Hochreiter's proposed LSTM method [10] has the potential to learn the long dependencies between variables. Long term time series is calculated using the LSTM algorithm. Faced with long time series results, it can dramatically boost the gradient explosion and the gradient disappearance of neural network algorithms.

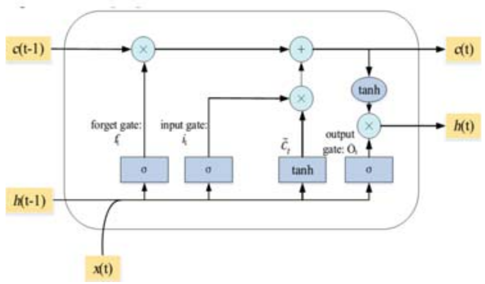

Figure 2. The structure of the LSTM unit

The LSTM structure diagram is shown in Figure 2. The LSTM unit is installed with three gates, namely the gates of input, forget, and output. They are used specifically to decide what information to keep. By switching gates, the LSTM network applies temporary memory to avoid gradient disappearance. Its external inputs for the basic LSTM unit are its preceding cell state $c_{(t-1)}$, previous hidden state $h_{(t-1)}$, and the current input vector $x_{(t)}$.

The entire computation can be described as follows by a series of equations [21-23]:

$$f_{(t)} = \sigma(W_{fx} x_{(t)} + W_{fh} h_{(t-1)} + b_f) \tag{3}$$

$$i_{(t)} = \sigma(W_{ix} x_{(t)} + W_{ih} h_{(t-1)} + b_i) \tag{4}$$

$$o_{(t)} = \sigma(W_{ox} x_{(t)} + W_{oh} h_{(t-1)} + b_o) \tag{5}$$

$$c_{(t)} = \tanh(W_{cx} x_{(t)} + W_{ch} h_{(t-1)} + b_c) \tag{6}$$

where W_x and W_h are the weights, b is the bias vector. The activation functions are designated by σ. The sigmoid function for the gates can generally be used as an activation function. The tanh represents the nonlinear activation function.

Then, the memory cell and the hidden state of this LSTM are updated as:

$$c_{(t)} = f_{(t)} \odot c_{(t)} + i_{(t)} \odot c_{(t)} \tag{7}$$

$$h_{(t)} = o_{(t)} \odot \tanh(c_{(t)}) \tag{8}$$

$\odot$ is used to denote pointwise multiplication operation for two vectors.

III. RESULTS AND DISCUSSION

Simulations of the proposed diagnosis of heart disease is carried out in this section. The new model aims to predict patients with high precision and less time-complexity of heart disease. The heart disease dataset developed by us in [16] is used to compare the effectiveness of MLP and LSTM in terms of accuracy rate and other parameters. The dataset was subdivided into two sets (training and test sets).

In MLP and LSTM models, the thirteen numbers of input variables are used.

It should be noted that the hardware and software used to perform our simulation experiments are as follows: We have used an Intel Core TM i5-4300U cores and 1.9 GHz CPU processor with 8 GB RAM memory.

A. *Heart diseaes prediction based on MLP*

Define To compare the various architectures for MLP technique. The performance indicator such as Hidden layers (HL), the Neurons in hidden layers (NHL), number of iteration (NI), Accuracy Training (ATr), Accuracy Testing (ATs), learning (t_{tr}), and execution times (t_{ex}) for training and testing datasets are calculated and given in Table II to determine the adequate number of neurons and hidden layers.

TABLE II. PERFORMANCE PARAMETERS OF MLP FOR TRAINING AND TESTING PHASES

				Training phase		**Testing phase**	
Parameters	HL	NHL	NI	ATr(%)	$t_{tr(s)}$	ATs(%)	$t_{ex(s)}$
Architecture of network	2	(24, 8)	1000	94.73	183	89.18	0.01
	3	(100, 20, 8)	1000	77.01	1230	71.64	0.04

As can be seen from Table II, the network with two hidden layers shows to be the best for this application. It is characterized by a training time of 183 seconds and an accuracy training error of 94.73%. Therefore, in the ensuing, the network with two hidden layers is considered for the study of the MLP technique.

B. *Heart diseaes prediction based on LSTM*

An appropriate prediction algorithm is chosen to predict heart disease, based on the analysis of the parameters for a case. Hence, this paper selects the LSTM neural network algorithm with a high predictive ability to predict a patient. There are therefore some parameters in the LSTM. When testing the parameters without the algorithm, the parameters are usually manually modified by detecting a change in the loss function.

In this section, the performance of two models have been compared with each other and found the appropriate model to predict exergetic performance of heart disease prediction. The comparison of the two models is shown in Table 3, which is based on the ATr, ATs, t_{tr} and t_{ex}.

TABLE III. PERFORMANCE PARAMETERS OF MLP AND LSTM FOR HEART DISEASE PREDICTION.

	Training phase		Testing phase	
Models	**ATr(%)**	**t_{tr}(s)**	**ATs(%)**	**t_{ex}(s)**
MLP	94.73	183	89.18	0.01
LSTM	99.8	3	**96.5**	1.96x10-3

Table III reports the prediction performances of all the models. It is observable that LSTM model has achieved the best outcomes in all metrics with ATr =99.8% and ATs =96.5%, followed by MLP Model (ATr =94.73% and ATs=89.18%). LSTM allows minimum time training than MLP for heart disease prediction, which is the learning time for LSTM is much faster than that obtained by MLP. LSTM model is therefore well suited to learning large datasets.

IV. COMPARISON OF THE STATE-OF-THE-ART METHODS

To offer an indication of where performance-wise ranks our heart disease prediction system, we compare works that used the same experimental protocol, the same performance metrics, and the same datasets. We also note that in contrast the execution time is not considered because of the lack of this knowledge in the works we compare with. The predictive accuracy of heart disease measured both in our proposed method and in other related works have been stated in Table IV.

TABLE IV. COMPARISON OF STATE-OF-THE-ART METHODS

Authors	Method	Accuracy (%)
Shouman et al. [24]	Decision tree	81.4
Nahar et al. [25]	Naive Bayes	69.11
Ismaeel et al. [26]	Extreme learning machine	80
Amin et al. [27]	Logistic regression	78.03
Our work	**LSTM**	**96.5**

V. CONCLUSION

In this work, the memory capable LSTM network is used to resolve the characteristics of time series high correlation problems and optimize the number of LSTM network hidden layer nodes in predicting heart disease. The proposed model has an excellent performance in the prediction of heart disease by comparing it with the multi-layer perceptron model. The results obtained showed that the LSTM enhanced the efficiency of the predictive system considerably. Applying this model can have a significant effect on the design and implementation of heart disease prediction systems in healthcare, and economic savings.

REFERENCES

[1] World health organization report on cardiovascular diseases. https://www.who.int/health-topics/cardiovascular-diseases/#tab=tab_1, accessed on Sep. 02, 2019.

[2] Raza, K, Improving the prediction accuracy of heart disease with ensemble learning and majority voting rule. U-Healthcare Monitoring Systems, 2019, pp. 179-196.

[3] Deng, M., Wang, C., Tang, M., Zheng, T, Extracting cardiac dynamics within ECG signal for human identification and cardiovascular diseases classification. Neural Networks, 2018, pp. 70-83.

[4] Amin, M.S., Chiam, Y.K., Varathan, K.D, Identification of significant features and data mining techniques in predicting heart disease. Telematics and Informatics, 2019, pp. 82-93.

[5] G. Guidi, M.C. Pettenati, P. Melillo, E. Iadanza, A machine learning system to improve heart failure patient Assistance, IEEE J Biomed Health Inf., 2014, pp. 1750-1756

[6] S. Mohan, C. Thirumalai, G. Srivastava, Effective heart disease prediction using hybrid machine learning techniques IEEE Access, 2019, pp. 81542-81554.

[7] Abushariah, M.A., Alqudah, A.A., Adwan, O.Y., Yousef,R.M, Automatic heart disease diagnosis system based on artificial neural network (ANN) and adaptive neuro-fuzzy inference systems (ANFIS) approaches. Journal of Software Engineering and Applications, 2014, 7(12):1055.

[8] Bouali, H., Akaichi, J, Comparative study of different classification techniques: Heart disease use case, 13th International Conference on Machine Learning and Applications, Detroit, MI, USA, 2014, pp. 482-486.

[9] Otoom, A.F., Abdallah, E.E., Kilani, Y., Kefaye, A., Ashour, M, Effective diagnosis and monitoring of heart disease. International Journal of Software Engineering and Its Applications, (2015). 9(1): pp. 143-156.

[10] S. Hochreiter, and J. Schmidhuber, Long Short-Term Memory, Neural. Comput., 1997, pp. 1735-1780.

[11] M. Wollmer, B. Schuller, F. Eyben, and G. Rigoll, Combining Long Short-Term Memory and Dynamic Bayesian Networks for Incremental Emotion-Sensitive Artificial Listening, IEEE Journal of Selected Topics in Signal Processing, , 2010, pp. 867-881.

[12] Juan Huan, et al., Prediction of dissolved oxygen in aquaculture based on gradient boosting decision tree and long short-term memory network: A study of Chang Zhou fishery demonstration base, China, Computers and Electronics in Agriculture 2020, 175, 105530

[13] M. Wöllmer, F. Weninger, J. Geiger, B. Schuller, and G. Rigoll, Noise robust ASR in reverberated multisource environments applying convolutive NMF and Long Short-Term Memory, Computer Speech & Language, 2013, pp. 780-797.

[14] N. Srivastava, E. Mansimov, and R. Salakhutdinov, Unsupervised Learning of Video Representations using LSTMs, 32nd International Conference on International Conference on Machine Learning, 2015, pp. 843–852.

[15] Xiaofeng Yuan, et al., Nonlinear dynamic soft sensor modeling with supervised long short-term memory network, IEEE transactions on industrial informatics, 2018, pp. 1551-3203.

[16] Heart disease dataset of University of California at Irvine (UCI). https://archive.ics.uci.edu/ml/datasets/Heart+Disease.accessed on Sept. 02, 2019.

[17] H. Hoglund, Detecting earnings management with neural networks, Expert Syst. Appl, 2012, pp.9564–9570.

[18] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning representations by back-propagating errors, Nature 323, 1986, pp.533–536.

[19] A.P. Piotrowski, Differential evolution algorithms applied to neural network training suffer from stagnation, Appl. Soft Comput, 2014, pp. 382–406.

[20] R. Pascanu, T. Mikolov, Y. Bengio, On the difficulty of training recurrent neural networks, Int. Conf. Mach. Learn, 2012, pp. 1298–1310.

[21] Xiaofeng Yuan, Nonlinear dynamic soft sensor modeling with supervised long short-term memory network, IEEE transactions on industrial informatics, 2019, PP(99):1-1.

[22] Farah Shahid, et all., A novel wavenets long short term memory paradigm for wind power prediction, Applied Energy, 2020, 115098.

[23] Qin Zhang, et all., Prediction of Sea Surface Temperature Using Long Short-Term Memory, IEEE geoscience and remote sensing letters, 2017, pp. 1745 - 1749.

[24] Shouman, M., Turner, T., Stocker, R, Using decision tree for diagnosing heart disease patients. In: 9thAustralasian Data Mining Conference, 2011,pp. 23-30.

[25] Nahar, J., Imam, T., Tickle, K.S., Chen, Y.P.P, Computational intelligence for heart disease diagnosis: Amedical knowledge driven approach. Expert Systems with Applications, 2013, pp. 96-104.

[26] Ismaeel, S., Miri, A., Chourishi, D, Using the extreme learning machine technique for heart disease diagnosis. In: International Humanitarian Technology Conference (IHTC2015), Ottawa, ON, Canada, 2015, pp. 1-3.

[27] Amin, M.S., Chiam, Y.K., Varathan, K.D, Identification of significant features and data mining techniques in predicting heart disease. Telematics and Informatics, 2019, pp. 82-93.