PROCEEDINGS OF
INTERNATIONAL CONFERENCE ON ADVANCED TECHNOLOGIES
https://proceedings.icatsconf.org/
11th International Conference on Advanced Technologies (ICAT'23), Istanbul-Turkiye, August 17-19, 2023.

A System Architecture Based on The RNN Classifier for Heart Disease Detection

Volkan Göreke[1]

[1]*Sivas VSTS, Sivas Cumhuriyet University , Sivas, Türkiye*
vgoreke@.cumhuriyet.edu.tr, ORCID: 0000-0002-2418-8373

***Abstract*— Diagnosing heart disease is a challenging process for physicians. Insufficient number of experts, late diagnosis and misdiagnosis are the difficulties in this process. To overcome these difficulties, systems based on artificial intelligence are used today. Appropriate system selection and obtaining sufficient data sets are a challenge for researchers. In this study, a high-performance CAD architecture was proposed for the detection of heart disease. The proposed architecture has shown a higher performance than the studies carried out using the UCI dataset in the literature.**

***Keywords*— CAD, heart disease, RNN, deep learning, UCI dataset**

I. Introduction

Around 30% of deaths worldwide are due to heart disease [1]. Diagnosing heart disease is a challenging process for physicians. Insufficient number of experts, late diagnosis and misdiagnosis are the difficulties involved in this process. At the present time to overcome these difficulties, computer-aided decision support systems (CAD) are a popular tool [2]. These systems commonly use classifier structures such as support vector machine (SVM), K-Nearest-Neighbour algorithms (KNN), Naive-Bayes Random Forest (NBRF), and artificial neural network (ANN). these structures are optimized by incorporating some optimization methods such as ant colony optimization (ACO) and particle swarm optimization (PSO) algorithm [3]. In these systems commonly are used classifier structures such as support vector machine (SVM), K-Nearest-Neighbour algorithms (KNN), Naive-Bayes Random Forest (NBRF), and artificial neural network (ANN). these structures are sometimes optimized by incorporating some optimization methods such as ant colony optimization (ACO) and particle swarm optimization (PSO) algorithm [3].

In researches based on deep learning, the low number of samples causes memorization and low performance problems of the system. In order to overcome all these negativities, the use of data augmentation techniques such as Variational AutoEncoders (VAE) is mandatory [4].

In the literature, the UCI Cleveland data set has been widely used in studies for the diagnosis of heart diseases. The studies carried out on this data set was summarized below.

Zameer Khan et al. investigated various methods such as logistic regression, KNN Classifier, RF, SVM, Decision Tree, Gaussian Naïve Bayesian in classification of heart diseases. They achieved the highest 87.91% performance using the SVM classifier [5]. According to Saba Bashir et al. obtained the following accuracy on the logistic regression (LR) support vector machine: 84.85% [6]. Dinesh Kumar et al. used various types of classifiers and Logistic Regression achieved an accuracy of 86.51% [7]. Ganesan M et al., using the decision tree classification algorithm based on Iterative Dichotomiser (j48) technique, obtained 91.48% accuracy [8]. Kannan, et al., using Logistic regression, achieved the best accuracy of 86.51% [9]. Khashei and Bakhtiarvand proposed the discrete learning-based MLP (DIMLP) classifier model for the detection of heart disease. The researchers achieved 94.27% success with this classifier they proposed [10].

II. Material and Method

The dataset was used in this study includes 303 samples and 13 features. The dataset was obtained from the University of California Irvine (UCI) Machine Learning Repository heart disease dataset. 6 incorrect data in the data set were removed from the data set [11]. The software of the proposed architecture was developed using the python language on the Visual Studio Community 2017 platform. The diagram of the proposed method in this study was given in Figure 1.

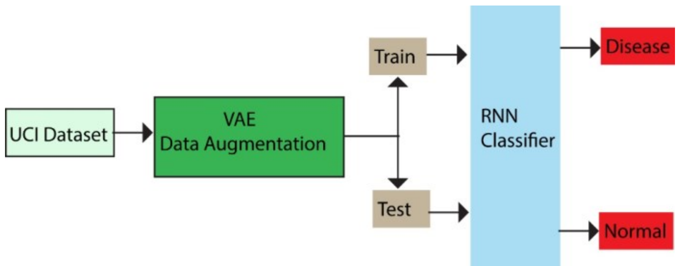

Fig 1 Diagram of the proposed method

A. Variational Autoencoder and Dataset Augmentation

Variational Autoencoder (VAE) consists of two basic components, an encoder and a decoder. The encoder transforms the data into latent variables that represent a low-dimensional latent space [12]. This transformation is performed by modelling the probability distribution of the data [13]. The decoder regenerates the data using latent variables [12].

In this study, the data set was increased and balanced approximately 2 times using the VAE technique. The resulting augmented dataset contains 600 samples.

B. Recurrent Neural Network Classifier

A Recurrent Neural Network (RNN) is an advanced artificial neural network and has internal memory. Hidden states of RNN are given in Equation 1 and 2 [14].

$$S_t = \tanh(Wx_t + US_{t-1} + b) \quad (1)$$

$$o_t = c + VS_t \quad (2)$$

where x_t the input vector at time t, b and c are bias values. Weight matrices are U, V and W.

The parameter values of the RNN architecture that was used in this study were given in Table I.

TABLE I.
PARAMETER VALUES OF THE RNN ARCHITECTURE

RNN unit	256
Full conn.layer	1-2-3
Full conn.unit	2048-1024-512
Momentum	0,3
Decay	0,5
Learning rate	0,3
Epoc	70
Optimizier	SGD

C. Result and Discussion

Cross-validation method may be preferred in artificial intelligence studies using a small data set. However, this method gives less clear results in medical applications [15]. In this study, train-split approach was used to obtain clearer results. Also, split ratio was determined as 20. For test criteria, metrics with mathematical expressions between Equation 3 – Equation 6 were used. The results obtained were given in Table II.

$$Accuracy = (TP+TN)/(TP+TN+FP+FN) \quad (3)$$

$$\mathrm{Re}call = TP/(TP+FN) \quad (4)$$

$$Precision = TP/(TP+FP) \quad (5)$$

$$F1\text{-}score = 2 \times pricison \times recall \,/\, precision + recall \quad (6)$$

Here *TP* (true positive) is the number of patients diagnosed as patient, *TN* (true negative) is the number of healthy individuals diagnosed as healthy, *FP* (false positive) is the number of healthy individuals diagnosed as patient and, *FN* (false negative) is the number of patients diagnosed as healthy [16]. The receiving operating characteristic (ROC) curve is used to calculate the *AUC* (area under the curve) value. *AUC* is interpreted as a higher probability of test measurement for a random patient than for a random healthy individual [17].

TABLE II.
PERFORMANCE METRICS OF THE PROPOSED METHOD

Accuracy	Recall	Precision	F1-score	AUC
0,9976	0,9977	0,9977	0,9977	1,0000

Performance graphics of the proposed system were given in Figure 2, Figure 3 and Figure 4.

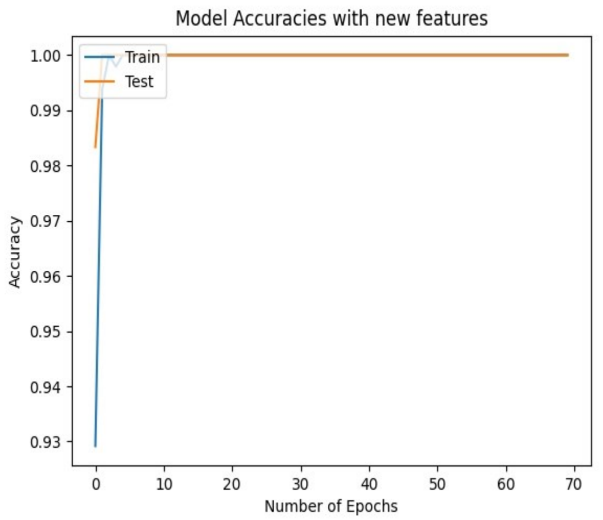

Fig. 2 RNN classifier accuracy graph

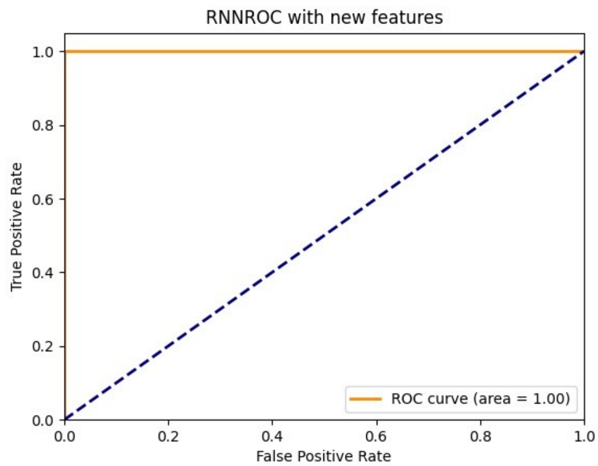

Fig. 3 RNN classifier ROC graph

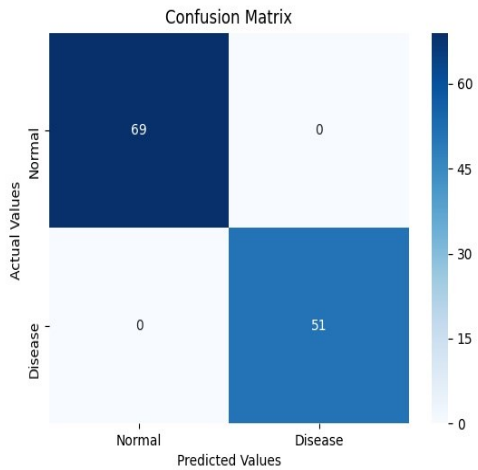

Fig. 4 Confusion matrix obtained from the RNN classifier

The results of comparing the classification accuracies obtained in the studies on the same data set in the literature with the classification accuracy obtained in this study were given in Table III.

TABLE III.
COMPARISON OF THE RESULTS OF THIS STUDY WITH THE RESULTS OF THE STUDIES IN THE LITERATURE

Reference	**Method**	**Accuracy**
[5]	SVM	87.91%
[6]	LR-SVM	84.85%
[7]	LR	86.51%
[8]	J48	91.48%
[9]	LR	86.51%
[10]	DIMLP	94.27%
This study	**VAE-RNN**	**99.76%**

Obtaining a medical dataset containing enough samples is a challenge for researchers working in the biomedical field. In order to overcome this challenge, it is imperative for researchers to resort to data augmentation techniques. Data set augmentation is also a method used to improve classification performance. The use of the VAE method in heart disease detection studies both increases the dataset size appropriately and can be a solution to the balance problem between the classes. The VAE data augmentation technique proposed in this study and the use of RNN classifier based on deep learning achieved a superior performance compared to the studies in the literature.

REFERENCES

[1] World Health Organization and J. Dostupno, cardiovascular diseases: key facts, vol. 13, no. 2016, p. 6, 2016. [Online]. Available: https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases- (cvds).

[2] Baviskar V,Vermab M., Chatterjee P., Singal G. "Efficient Heart Disease Prediction Using Hybrid Deep Learning Classification Models". *IRBM* 44(2023) 100786.

[3] KhourdifiY, BahajM. "Heart disease prediction and classification using machine learning algorithms optimized by particle swarm optimization and ant colony optimization." *Int J Intell Eng Syst* 2019.

[4] Chadebec C., Allassonniere S., "Data Augmentation with Variational Autoencoders and Manifold Sampling." *arXiv*:2103.13751v3. https://doi.org/10.48550/arXiv.2103.13751

[5] Z. Khan, D.K. Mishra, V. Sharma, A. Sharma, "Empirical study of various classification techniques for heart disease prediction", in: *Proceedings of the IEEE 5th International Conference on Computing Communication and Automation (ICCCA),* 2020, pp. 57– 62, doi: 10.1109/ICCCA49541.2020.9250852

[6] S. Bashir, Z.S. Khan, F. Hassan Khan, A. Anjum, K. Bashir, "Improving heart disease prediction using feature selection approaches" in: *Proceedings of the 16th International Bhurban3 Conference on Applied Sciences and Technology (IBCAST),* 2019, pp. 619–623, doi: 10.1109/IBCAST.2019.8667106.

[7] K.G. Dinesh, K. Arumugaraj, K.D. Santhosh, V. Mareeswari, "Prediction of cardiovascular disease using machine learning algorithms" in: *Proceedings of the International Conference on Current Trends towards Converging Technologies (ICCTCT),* 2018, pp. 1–7, doi: 10.1109/ICCTCT.2018.8550857.

[8] M. Ganesan, N. Sivakumar, "IoT based heart disease prediction and diagnosis model for healthcare using machine learning models", in: *Proceedings of the IEEE (ICSCAN),* 2019, pp. 1–5, doi:10.1109/ICSCAN.2019.8878850.

[9] R. Kannan, V. Vasanthi, "Machine learning algorithms with ROC curve for predicting and diagnosing the heart disease." in: Springer Briefs in Applied Sciences and Technology, Springer Verlag, 2019, pp. 63–72, doi:10.1007/978-981-13-0059-2_8

[10] M. Khashei, N. Bakhtiarvand. "Discrete learning-based intelligent methodology for heart disease diagnosis." *Biomedical Signal Processing and Control.* 84 (2023) 104700. doi: 10.1016/j.bspc.2023.104700

[11] Uyar K, İhan A. "Diagnosis of heart disease using genetic algorithm based trained recurrent fuzzy neural networks." *Procedia Computer Science* 120 (2017) 588–593. Doi: 10.1016/j.procs.2017.11.283

[12] Doersch C., Mellon C. "Tutorial on Variational Autoencoders." *arXiv*:1606.05908v. https://doi.org/10.48550/arXiv.1606.05908

[13] Rezende D.J., Mohammed S. Variational Inference with Normalizing Flows.arXiv:1505.05770v6.(2015).https://doi.org/10.48550/arXiv.1505.05770.

[14] Asrav T.,Aydın E., "Physics-informed recurrent neural networks and hyper-parameter optimization for dynamic process systems." *Computers and Chemical Engineering* 173 (2023) 108195. https://doi.org/10.1016/j.compchemeng.2023.108195.

[15] T.B. Alakus, I. Turkoglu, "Comparison of deep learning approaches to predict COVID-19 infection." *Chaos Solitons Fractals* (140) (2020) 110120, http://dx.doi.org/10.1016/j.chaos.2020.110120.

[16] W. Mumtaz, A. Qayyum, "A deep learning framework for automatic diagnosis of unipolar depression." *Int. J. Med. Inform.* (132) (2019) 103983, http://dx.doi.org/10.1016/j.ijmedinf.2019.103983.

[17] S. Safari, A. Baratloo, M. Elfil, A. Negida, "Evidence based emergency medicine; part 5 receiver operating curve and area under the curve." Emergency 4 (2) (2016) 111–113, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4893763.