

Implementation of an incremental deep learning model for survival prediction of cardiovascular patients

Sanaa Elyassami¹, Achraf Ait Kaddour²

¹Department of Information Security and Engineering Technology, Abu Dhabi Polytechnic, United Arab Emirates

²Manchester University, United Kingdom

Article Info

Article history:

Received Aug 15, 2020

Revised Dec 28, 2020

Accepted Feb 5, 2021

Keywords:

Activation function
Binary classification
Deep learning
Machine learning
Neural network

ABSTRACT

Cardiovascular diseases remain the leading cause of death, taking an estimated 17.9 million lives each year and representing 31% of all global deaths. The patient records including blood reports, cardiac echo reports, and physician's notes can be used to perform feature analysis and to accurately classify heart disease patients. In this paper, an incremental deep learning model was developed and trained with stochastic gradient descent using feedforward neural networks. The chi-square test and the dropout regularization have been incorporated into the model to improve the generalization capabilities and the performance of the heart disease patients' classification model. The impact of the learning rate and the depth of neural networks on the performance were explored. The hyperbolic tangent, the rectifier linear unit, the Maxout, and the exponential rectifier linear unit were used as activation functions for the hidden and the output layer neurons. To avoid over-optimistic results, the performance of the proposed model was evaluated using balanced accuracy and the overall predictive value in addition to the accuracy, sensitivity, and specificity. The obtained results are promising, and the proposed model can be applied to a larger dataset and used by physicians to accurately classify heart disease patients.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Sanaa Elyassami
Department of Information Security and Engineering Technology
Abu Dhabi Polytechnic
Mohammed Bin Zayed City, Abu Dhabi, United Arab Emirates
Email: sanaa.elyassami@adpoly.ac.ae

1. INTRODUCTION

Cardiovascular diseases are the most common underlying cause of death in the world, and the morbidity and mortality are still on the rise [1]. It has been estimated that, by 2030, more than 40% of US adults or 116 million people will have one or more forms of cardiovascular diseases. The direct medical costs related to the cardiovascular diseases are expected to triple, from \$273 billion to \$818 billion, however, the indirect costs due to lost productivity are estimated to increase from \$172 billion to \$276 billion [2]. It is critical to develop preventive intervention strategies to limit the progression of cardiovascular disease and to minimize the associated direct and indirect costs.

Modeling survival patients with heart failure remains a constant problem nowadays in terms of identifying the significant factors along with achieving high classification accuracy. However, the increasing availability of electronic data presents a major opportunity to implement robust models. Machine learning provides computational intelligence techniques to tackle the issue of analysis and prediction within large complex datasets. Machine learning is attracting broad interest in healthcare [3]. When applied to medical

records, common predictive models, also known as health forecasting, can be an effective tool for leveraging data to make predictions and highlight patients most at risk. Deep learning is one of the most used machine learning techniques in the medical field. In a recent study, deep learning was used along with new features that were extracted from the x-ray images for tuberculosis detection. The results show that the proposed method produced an accuracy of 89.77%, a sensitivity of 90.91%, and a specificity of 88.64% [4]. Another study did use a deep learning model called AlexNet based on 9,000 single red blood cell images taken from 130 patients. The model was used for classifying the abnormalities present in the sickle cell anemia disease to give a better insight into managing the concerned patient's life and it achieved a high classification prediction accuracy of 95.92% [5]. Neural networks were applied to cancer disease to classify lymph, neck and head, and breast cancer that might help clinicians and oncologists in the prediction and prognosis of cancer [6]. For heart disease, machine learning techniques can be useful to predict risk at an early stage. Some of the techniques used for such prediction problems were the support vector machines (SVM), neural networks, decision trees, regression, and naïve bayes classifiers. SVM was identified as the best predictor with 92.1% accuracy, followed by neural networks with 91% accuracy, and decision trees showed a lesser accuracy of 89.6% [7].

Other studies based on neural networks and other machine learning methods used data on cardiovascular patients collected from the UCI Laboratory, and applying discovery pattern algorithms including decision tree, neural networks, rough set, SVM, naïve bayes, and compare their accuracy and prediction, and achieving an F-measure of 86.8% [8]. Although, other studies were presented in [9-10] that trained neural network-based model for classifying the heart disease and to predict accurately abnormalities in the heart or it's functioning. Another research in cardiovascular disease prediction used seven classification techniques: k-NN, decision tree, naïve bayes, logistic regression, support vector machine, neural network with vote. The results showed that the heart disease prediction model using neural network with vote achieved the best accuracy of 87.4% [11]. To improve models' effectiveness, recent published studies used hybrid models. In [12], the Cleveland database was selected and a hybrid random forest with a linear model called HRFLM was used to find significant features and to improve the prediction of cardiovascular disease that produced an accuracy of 88.7%.

In the current study, we developed and fine-tune a machine learning model using different techniques. First, we used a multilayer feedforward artificial neural network to build the model, then we employed a deep feedforward neural network to improve it. After that, we trained and utilized machine learning binary classifiers to build different models using several activation functions. Hyperparameters that affect both the regularization and the optimization during the training phase were considered. Different evaluation metrics based on confusion matrices were applied to evaluate the performance of the models, and additional metrics were suggested to get more accurate classifiers when dealing with an imbalanced dataset. To improve classification performance, features selection was applied by using the Chi-squared test to select the most pertinent factors. And to avoid overfitting, the dropout regularization technique was used to improve the model generalization.

2. RESEARCH METHOD

2.1. Dataset description

The current study is based on a dataset containing the medical records of 299 heart failure patients [13]. The patients' age ranged between 40 and 95 years old, and they all suffered from a left ventricular systolic dysfunction and had previous heart failures that categorize them in class III or class IV of the New York Heart Association classification of heart failure stages. The records were collected during the follow-up at the Allied Hospital in Faisalabad and at the Faisalabad Institute of Cardiology in Pakistan in 2015 based on blood reports, cardiac echo reports, and physician's notes. The dataset contains 299 records, each record is characterized by 13 clinical features as presented in Table 1. The death event feature is a binary attribute and is the target in our study which indicates if the patient died or survived before the end of the follow-up period. The follow-up period was between 4 and 285 days with an average of 130 days. The dead patients represent 32.11% (96 patients) and the survived patient represents 67.89% (203 patients).

The dataset is composed of six dichotomous binary variables: smoking, anemia, sex, high blood pressure, diabetes, and the dead event. It also includes seven continuous quantitative variables: creatinine phosphokinase, age, serum sodium, ejection fraction, serum creatinine, platelets, and time. The creatinine phosphokinase states the level of the creatinine phosphokinase enzyme in the blood. A high level of creatinine phosphokinase is indicative of stress or injury to the heart or other muscles. The creatinine phosphokinase normal values are 10 to 120 micrograms per liter (mcg/L) [14]. While the serum creatinine measures the level of creatinine in the blood and provides an estimate of how well the kidneys function, a high level of serum creatinine is indicative of renal dysfunction. The serum creatinine normal values are 0.9

to 1.3 milligrams per deciliter (mg/dL) for adult males, and 0.6 to 1.1 mg/dL for adult females [15]. Anemia is a condition in which the patient does not have enough healthy red blood cells to carry adequate oxygen to the body's tissues. The hospital physician considered a patient having anemia if the hematocrit level is lower than 36%. Platelets are blood cells that help the body form clots to stop bleeding. A normal platelet count ranges from 150,000 to 450,000 platelets per microliter of blood [16]. Ejection fraction is a measurement of the percentage of blood leaving the heart each contraction. An ejection fraction of 55% or higher is considered normal [17]. The serum sodium states if a patient has normal levels of sodium in the blood. A low sodium level has many causes, including kidney failure and heart failure. A normal sodium level is between 135 and 145 milliequivalents per liter (mEq/L) [18].

Table 1. Heart failure patients' dataset description

Clinical Feature	Description	Unit	Min value	Max value
Creatinine phosphokinase	Level of the CPK enzyme in the blood	mcg/L	23	7861
Serum creatinine	Level of serum creatinine in the blood	mg/dL	0.5	9.4
Serum sodium	Level of serum sodium in the blood	mEq/L	113	148
Ejection fraction	Percentage of blood leaving the heart at each contraction	Percentage	14	80
Platelets	Platelets in the blood	kiloplatelets/mL	25.1	850
Age	Patient' age	Year	40	95
Time	Follow-up period	Day	4	285
Diabetes	If the patient has diabetes	Boolean	0	1
Sex	Woman or man	Boolean	0	1
Anemia	Decrease of red blood cells or hemoglobin	Boolean	0	1
High blood pressure	If the patient has hypertension	Boolean	0	1
Smoking	If the patient smokes or not	Boolean	0	1
[target] Death event	If the patient deceased during the follow-up period	Boolean	0	1

2.2. Feed-forward neural network models

Classification is a task that requires the use of machine learning algorithms that learn how to assign a class label to examples from the problem domain. Binary classification predictive modeling involves assigning one of two classes to input examples. In the current study, we employed neural network-based models for binary classification. A neural network is comprised of an input layer, one or more hidden layers, and an output layer. The input nodes correspond to data sources, the output nodes correspond to the desired classes, whereas hidden layers are required for computational purposes. The values at each node are estimated through the summation of the multiplications between previous node values and weights of the links connected to that node. This value is referred to as the summed activation of the node which is then transformed via an activation function and defines the output as $h(x) = f(b + \sum w_i x_i)$ where $h(x)$ is the result of the neuron, x is the input, w is the weight, and b is the bias.

The activation function is a crucial component of learning that determines the accuracy and the computational efficiency of training a model. The simplest activation function is the linear one, where no transform is applied. A network comprised of only linear activation functions is very easy to train but cannot learn complex mapping functions. In our study, different neural network-based models have been implemented to predict survival patients. The hidden layers were trained using non-linear activation functions to allow the nodes to learn efficiently complex relationships in the data and provide accurate predictions. The four nonlinear activation functions: hyperbolic tangent [19], rectifier linear unit [20], maxout [21], and exponential rectifier linear unit [22] have been used to compute the output of the hidden nodes.

The hyperbolic tangent (tanH) is a continuous nonlinear function that produces outputs in the scale of $[-1, +1]$, where $f(x) = (e^x - e^{-x}) / (e^x + e^{-x})$. The rectified linear (ReLU) is a piecewise linear function. It is a linear function for values greater than zero and nonlinear for negative values. ReLU returns the input provided if the input is positive, otherwise, it returns zero where $f(x) = \max\{0, x\}$. Whereas, the exponential linear unit (ELU) is similar to ReLU except for negative values. ELU and ReLU are in identity function for positive inputs where $f(x) = x$. For negative values, ELU becomes smooth slowly until its output equal to $-\alpha$ as $f(x) = \alpha(e^x - 1)$. The maxout activation takes the maximum value over a set of units of the pre-activations and sends it forward to the output node.

In this paper, we developed a feedforward neural network model (FFNN) based on a multilayer feedforward artificial neural network. FFNN has an input layer of neurons, only one hidden layer that processes the inputs, and an output layer that provides the final output of the model. Each node in one layer is connected to every node on the next layer. Thus, information is continuously fed forward from one layer to the next layer, from the input nodes, through the hidden nodes, and to the output nodes. The pairs of input and output values are fed into the network for many cycles so that the network learns the relationship

between the input and output. Our second model is a deep feedforward neural network (DNN) based on a multilayer feedforward artificial neural network has an input layer of neurons, two hidden layers that process the inputs, and an output layer that provides the final output of the model. DNN is trained with stochastic gradient descent using the backpropagation algorithm. The stochastic gradient descent is based on a random probability and used to speed up learning by randomly picking out one sample from the dataset at each iteration to reduce the computations. stochastic gradient descent is an optimization technique that replaces the actual gradient computed from the entire dataset by an estimate thereof computed from a randomly selected subset of the dataset. The stochastic gradient descent recursively calculates the gradient of parameters starting at the network output layer and moving backward to other layers. The parameters are then updated and adjusted in order to reduce the loss function.

2.3. Hyperparameters selection

We trained and employed machine learning binary classifiers to build different models using several activation functions to the heart failure patients' data. The dataset contains 299 patients who suffered from a left ventricular systolic dysfunction, of which 203 survived and 96 died (32.11% negatives and 67.89% positives). Training neural networks requires setting hyperparameters that affect both the regularization and the optimization in the training phase. The hyperparameters affecting optimization are the learning rate η and the momentum coefficient μ . The standard value of $\mu = 0.9$ has been frequently observed to work well in practice [23] and was thus kept fixed throughout all experiments. Whereas, the learning rate value was explored by performing a grid search in the logarithmic scale between $\eta=1.0E-3$ and $\eta=1.0E-7$. In Figure 1, accuracy is plotted as a function of the learning rate. These experiments were carried out using tanH, ReLU, ELU, and Maxout activation functions throughout the feedforward neural network-based model. For very small learning rates ($\eta < 1.0E-5$), the accuracy is maximal. For values bigger than $1.0E-5$, the accuracy decreases sharply, especially with tanH and ELU. A learning rate of $\eta=1.0E-6$ was selected and kept fixed for all experiments. The optimum structure for a neural network should be large enough to learn the characteristics of the training set and small enough to generalize for the validation set [24]. To prevent overfitting, regularization methods should be used [24]. In the current study, the early stopping method has been used to stop model training when overfitting starts.

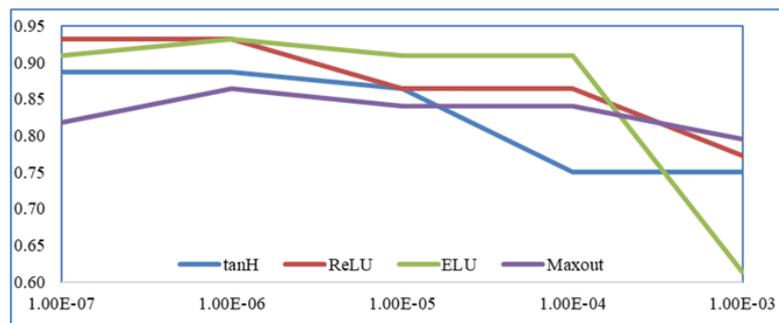


Figure 1. Optimum learning rate η based on models' accuracy

2.4. Evaluation metrics

The classification models predict the class of each instance of the dataset by assigning a predicted label to each sample. In our binary classification models (died, survived), each sample fall in one of four possibilities. True-positive (TP) where the model correctly predicts the positive class and thus, died people correctly identified as died. True-negative (TN) where the model correctly predicts the negative class and thus, survived people correctly identified as survived. False-negative (FN) where the model incorrectly predicts the positive class and thus, died people incorrectly identified as survived. False-positive (FP) where the model incorrectly predicts the negative class and thus, survived people incorrectly identified as dead. To evaluate the performance of our models, we employed several statistical measures based on confusion matrices. We measured the prediction results using accuracy, classification error, precision, sensitivity, and specificity [25].

Accuracy (Acc) is the ratio between the number of correctly classified samples and the overall number of samples. Acc is calculated as $Acc = \frac{\sum \text{True positive} + \sum \text{True negative}}{\sum \text{Total number of samples}}$. Classification error (CE) is the ratio between the number of incorrectly classified sample cases and the

overall number of samples. CE is calculated as $CE = \frac{\sum \text{False positive} + \sum \text{False negative}}{\sum \text{Total number of samples}}$. Sensitivity is called also the true positive rate (TPR) and it measures the proportion of actual positives that are correctly identified as positives. TPR is calculated as $TPR = \frac{\sum TP}{\sum TP + \sum FN}$

Specificity is called also the true negative rate (TNR) and it measures the proportion of actual negatives that are correctly identified as negatives. TNR is calculated as $TNR = \frac{\sum TN}{\sum TN + \sum FP}$

The positive predictive values (PPV) called also precision and the negative predictive values (NPV) are respectively the proportions of positive and negative results. Where PPV is calculated as $PPV = \frac{\sum TP}{\sum \text{predicted condition positive}}$. And the predicted condition positive represents the sum of TP and FP. Whereas NPV is calculated as $NPV = \frac{\sum TN}{\sum \text{Predicted condition negative}}$. Where the predicted condition negative is the summation of TN and FN.

In the current study, we used an imbalanced dataset where the number of samples in the negative class is much larger than the number of samples in the positive class, with 67.89% negatives and 32.11% positives. However, when the dataset is imbalanced, some statistical rates can show overoptimistic and exaggerated results on the majority class, especially the accuracy. Thus, to overcome the class imbalanced dataset issue, we used additional metrics that produce a high rate only if the model was able to correctly predict both, positive samples and negative ones. The balanced accuracy (BACC) and the overall predictive value (OPV) provide useful insights into the classifier's behavior without being affected by the imbalanced dataset issue [26-27]. BACC is calculated as: $BACC = \frac{TPR + TNR}{2}$. Whereas OPV is calculated as $OPV = \frac{PPV + NPV}{2}$. Thus, a classification model with the highest balanced accuracy, the highest overall predictive value, and the lowest classification error is considered to be the most accurate classifier.

3. EXPERIMENT DESIGN AND RESULTS

In the current study, we employed two network architectures to build the models. The first model is based on a feedforward neural network (FFNN) and includes one input layer, one hidden layer, and one output layer. The second model is a deep feedforward neural network (DNN) that includes one input layer, two hidden layers, and one output layer and was trained with stochastic gradient descent using backpropagation. For both models, we trained the binary classifiers on a training set containing 80% of randomly selected data samples and test them on the testing set containing the remaining 20% data samples. Since activation functions can perform differently on different datasets the choice of function to use for the hidden neurons becomes challenging. For all the classifiers, we repeated the experiment execution using the four nonlinear activation functions (tanH, ReLU, ELU, Maxout) and recorded the results for accuracy, balanced accuracy, classification error, sensitivity, specificity, and the overall predictive value. We then make the choice to rank the results obtained on the testing sets based on the balanced accuracy first, then based on the overall predictive value. This choice will be discussed in the following paragraph. The overall adopted process in the current study is depicted in Figure 2.

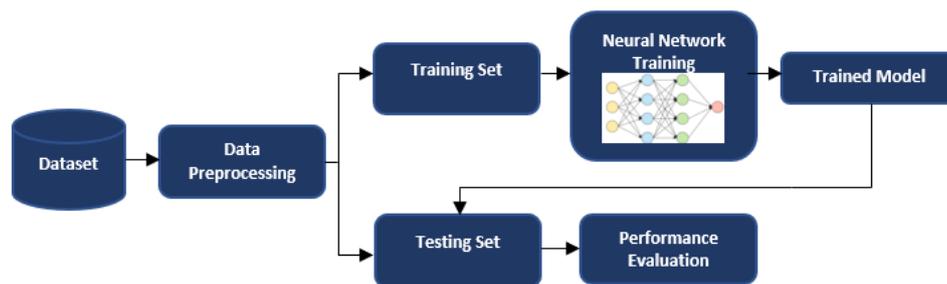


Figure 2. Adopted process

3.1. Results of feedforward neural network and deep neural network

After training the feedforward neural network (FFNN) model with different activation functions, the networks were finally evaluated on the testing data, obtaining the classification results displayed in Table 2. As mentioned earlier, we prefer to focus on the results obtained by the balanced accuracy and by the overall predictive value. These two metrics generate high scores only if the classifier was able to properly predict the positive data instances as well as the negative data instances. The two rankings we employed show interesting aspects. First, the top classifier changes when we consider the ranking based on balanced accuracy, or overall predictive value. In fact, the top-performing activation function based on the balanced

accuracy is tanH (82.62%), while based on the overall predictive value ranking the best classifier resulted in being Maxout (83.34%). ReLU is ranked fourth in the balanced accuracy ranking and in the overall predictive value ranking, whereas ELU is ranked third.

The classification results of the deep neural network (DNN) model measured in terms of a set of evaluation metrics are shown in Table 3. The network using Maxout as activation function did quite well both on the recall (TP rate=71.43%) and on the specificity (TN rate=86.67%) and was ranked first in terms of balanced accuracy (79.05%). In terms of overall predictive value, tanH classifier is top ranked (85.88%). ELU is the top performing in the accuracy ranking with an excellent score for specificity (TN rate=93.33%) but only a moderate score on recall (TP rate=64.29%). It is also noticed that ELU is performing much better than ReLU in terms of prediction and accuracy. This can be interpreted by the fact that ReLU for a set of inputs, the network cannot perform backpropagation and cannot learn anymore.

Table 2. FFNN model classification results on the testing data trained with different activation functions

Activation Function	Accuracy	Classification Error	Negative predictive value	Positive predictive value	Overall predictive value	TN rate	TP rate	Balanced accuracy
tanH	84.09%	15.91%	89.66%	73.33%	81.50%	86.67%	78.57%	82.62%
ReLU	77.27%	22.73%	79.41%	70.00%	74.71%	90.00%	50.00%	70.00%
Maxout	84.09%	15.91%	84.85%	81.82%	83.34%	93.33%	64.29%	78.81%
ELU	79.55%	20.45%	83.87%	69.23%	76.55%	86.67%	64.29%	75.48%

The results obtained from FFNN and DNN models showed that DNN outperformed FFNN for the classification of patients for most of the activation functions. Using deep learning, ELU-based network overall prediction and tanH-based network balanced overall prediction have been increased respectively by 6.79% and 4.38%. It can be noticed also that because of the class imbalance of the dataset (203 negative samples and 96 positive samples), prediction scores on the true negative rate are much better than the true positive rate. These results happen because the neural networks were well trained with large negative samples, and consequently, they can efficiently recognize them.

Table 3. DNN model classification results on the testing data trained with different activation functions

Activation Function	Accuracy	Classification Error	Negative predictive value	Positive predictive value	Overall predictive value	TN rate	TP rate	Balanced accuracy
tanH	84.09%	15.91%	82.86%	88.89%	85.88%	96.67%	57.14%	76.91%
ReLU	77.27%	22.73%	88.46%	61.11%	74.79%	76.67%	78.57%	77.62%
Maxout	81.82%	18.18%	86.67%	71.43%	79.05%	86.67%	71.43%	79.05%
ELU	84.09%	15.91%	84.85%	81.82%	83.34%	93.33%	64.29%	78.81%

3.2. Deep neural network model enhancement using feature selection

The motivation for applying feature selection is not only to reduce the dimension of the input layer but also to eliminate the least effective and correlated features, and to remove some interconnections or eliminate some hidden layer neurons to improve generalization capabilities, and thus achieve an improved performance. Feature selection is the process of identifying and extracting the most relevant attributes prior to applying any machine learning techniques on dataset samples. Applying machine learning algorithms on a large number of irrelevant attributes increases exponentially the training time and the risk of overfitting. The feature selection reduces the training time, so the models train faster, and with less redundant data that give a boost to the model performance. In our study, the Chi-squared test [28-29] has been used to select the most pertinent attributes. This metric determines if a distribution of observed frequencies differs from the theoretical expected frequencies. The chi-square score statistic is calculated as $X^2 = \sum [(OF - EF)^2 / EF]$ where X^2 is the chi-square statistic, OF is the observed frequency and EF is the expected frequency. This metric measures the weights of the dataset attributes with respect to the target attribute. We calculated Chi-square between each feature and the target died event, and we selected four attributes with the best Chi-square scores as shown in Figure 3. The attributes with higher weight are considered more relevant to predict survival patients. Thus, ejection fraction, serum creatinine, age, and serum sodium are the selected attributes.

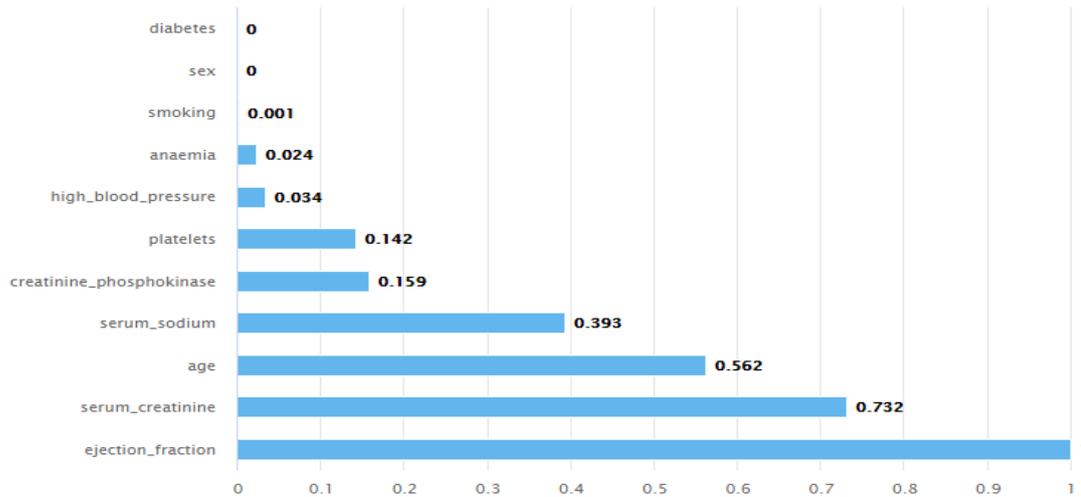


Figure 3. Normalized attribute weights using Chi-squared test with respect to the target feature

Incorporating the feature selection process in our deep neural network model (FS_DNN), allowed us to improve the prediction of survival and get better classification performance as shown in Table 4.

Table 4. FS_DNN Classification results on the testing data trained with different activation functions

Activation Function	Accuracy	Classification Error	Negative predictive value	Positive predictive value	Overall predictive value	TN rate	TP rate	Balanced accuracy
tanH	86.36%	13.64%	92.86%	75.00%	83.93%	86.67%	85.71%	86.19%
ReLU	88.64%	11.36%	90.32%	84.62%	87.47%	93.33%	78.57%	85.95%
Maxout	86.36%	13.64%	87.5%	83.33%	85.42%	93.33%	71.43%	82.38%
ELU	93.18%	6.82%	93.55%	92.31%	92.93%	96.67%	85.71%	91.19%

It has been shown that the exponential linear unit (ELU) outperformed other activation functions. Thus, the overall prediction value has reached a high score of 92.93% with a performance increase of 7% compared to the DNN model. And based on the balanced accuracy, FS_DNN scored 91.19% with a performance increase of 12%.

3.3. Deep neural network model enhancement using dropout regularization

Deep architecture networks are more severely affected by overfitting and benefits more from regularization. The dropout regularization technique was applied to the proposed model and it was achieved by frizzing each unit in the hidden layer of the network at each training iteration which expands the training process time, as a large number of the parameters are deactivated at each iteration. Dropout probability was set to the recommended value of 0.5 [30-31]. With dropout technique, the networks learned more slowly, since parameters are updated less frequently, and parameters receive smaller gradients. As shown in Table 5, the dropout technique did enhance the balanced accuracy scores for the three networks that used tanH (enhanced by 5.24%), ReLU (enhanced by 3.82%), and Maxout (enhanced by 2.14%), and achieved the highest score of 91.43% compared to all previously trained models. However, the ELU-based network balanced accuracy decreased by 5% when using dropout regularization. Regarding the overall predictive value, the dropout technique did improve slightly the tanH-based network and the ELU-based network with the highest score of 94.12%.

The results obtained from our models are more accurate and efficient than [32]. From the results published in [32], the top accuracy was achieved by Random Forests (74%), followed by Gradient Boosting (73.8%), followed by Decision Trees (73.7%), followed by Neural networks (68%). The classification results showed that our model outperformed all the other existing methods and achieve an overall predictive value of 94.12%.

Table 5. Classification results on the testing data for the FS_DNN model using dropout regularization

Activation Function	Accuracy	Classification Error	Negative predictive value	Positive predictive value	Overall predictive value	TN rate	TP rate	Balanced accuracy
tanH	90.91%	9.09%	96.43%	81.25%	88.84%	90.00%	92.86%	91.43%
ReLU	88.64%	11.36%	96.30%	76.47%	86.39%	86.67%	92.86%	89.77%
Maxout	84.09%	15.91%	92.59%	70.59%	81.59%	83.33%	85.71%	84.52%
ELU	90.91%	9.09%	88.24%	100%	94.12%	100%	71.43%	85.72%

4. CONCLUSION

The current research study investigates the performance of the classification of heart disease patients. The impact of the learning rate on the accuracy of shallow neural networks was explored, and different activation functions were investigated for the first time for heart disease classification problems. These functions are the hyperbolic tangent, the rectifier linear unit, the maxout, and the exponential rectifier linear unit. The impact of the depth of neural networks on the accuracy was investigated. A comparison between a feed-forward network classifier accuracy and a deep feed-forward network classifier accuracy was carried out. An intelligent deep learning model was developed and trained with stochastic gradient descent using the backpropagation algorithm. The dropout regularization and the chi-square test have been incorporated into the model to improve the classification accuracy of heart disease patients. The performance of the proposed deep neural network model was evaluated using the balanced accuracy and the overall predictive value metrics that provide useful insights into the classifier's behavior without being affected by the imbalanced dataset. We suggest all the researchers dealing with imbalanced datasets to evaluate their binary classification predictions through balanced accuracy and the overall prediction value in addition to the accuracy, sensitivity, and specificity.

Incorporating the feature selection process, allowed the proposed model to eliminate the least effective and the most correlated data and improved the model generalization capabilities. The overall prediction value was enhanced by 7%, and the balanced accuracy was enhanced by 12% compared to the deep neural network model. The performance was further slightly enhanced after integrating the dropout regularization technique that was used to prevent the model from overfitting and thus improve the classification performance especially for networks trained using tanH, ReLU, and Maxout activation functions. The proposed model achieves a balanced accuracy of 91.43% and a high overall predictive value of 94.12%. Therefore, the proposed model has the potential to generate a knowledge-rich environment that can significantly help to enhance the quality of clinical decisions by accurately predict the survival of cardiovascular patients. The obtained results are promising, and the proposed model can be applied to a larger dataset and used by physicians to accurately classify heart disease patients. Obviously, using deep feedforward neural networks for heart disease patient's classification is just one example of the successful applications of deep learning-based models to a real-world problem

REFERENCES

- [1] World Health Organization, "The top 10 causes of death fact sheet N 310," Geneva, Switzerland: World Health Organization, 2013.
- [2] Heidenreich, Paul A., *et al*, "Forecasting the future of cardiovascular disease in the United States: a policy statement from the American Heart Association," *Circulation*, vol.123, no. 8, pp. 933-944, 2011, doi: 10.1161/CIR.0b013e31820a55f5.
- [3] Wiens, J. and Shenoy, E.S., "Machine learning for healthcare: on the verge of a major shift in healthcare epidemiology," *Clinical Infectious Diseases*, vol. 66, no. 1, pp.149-153, 2018, <https://doi.org/10.1093/cid/cix731>.
- [4] Hijazi, M. H. A., Hwa, S. K. T., and Jeffree, M. S., "Ensemble deep learning for tuberculosis detection using chest X-Ray and canny edge detected images," *IAES International Journal of Artificial Intelligence*, vol. 8, no. 4, pp. 429-435, 2019, doi:10.11591/ijai.v8.i4.pp429-435.
- [5] Aliyu, H. A., Razak, M. A. A., Sudirman, R., and Ramli, N., "A deep learning AlexNet model for classification of red blood cells in sickle cell anemia," *Int J Artif Intell*, vol. 9, no. 2, pp. 221-228, 2020, doi:10.11591/ijai.v9.i2.pp221-228.
- [6] Mahmood, M., Al-Khateeb, B., and Alwash, W. M., "A review on neural networks approach on classifying cancers," *IAES International Journal of Artificial Intelligence*, vol. 9, no. 2, pp. 317-326, 2020, doi:10.11591/ijai.v9.i2.pp317-326.
- [7] Xing, Y., Wang, J. and Zhao, Z., "Combination data mining methods with new medical data to predicting outcome of coronary heart disease," *In 2007 International Conference on Convergence Information Technology (ICCIT 2007)*, *IEEE*, pp. 868-872, 2007, doi: 10.1109/ICCIT.2007.204.

- [8] H. A. Esfahani and M. Ghazanfari, "Cardiovascular disease detection using a new ensemble classifier," *2017 IEEE 4th International Conference on Knowledge-Based Engineering and Innovation (KBEI)*, Dec. 2017, pp. 1011-1014, doi: 10.1109/KBEI.2017.8324946.
- [9] D. K. Ravish, K. J. Shanthi, N. R. Shenoy and S. Nisargh, "Heart function monitoring prediction and prevention of heart attacks: Using artificial neural networks," *2014 International Conference on Contemporary Computing and Informatics (IC3I)*, pp. 1-6, Nov. 2014, doi: 10.1109/IC3I.2014.7019580.
- [10] W. Zhang and J. Han, "Towards heart sound classification without segmentation using convolutional neural network," *Proc. Comput. Cardiol. (CinC)*, vol. 44, pp. 1-4, Sep. 2017, doi: 10.22489/CinC.2017.254-164.
- [11] Latha, C.B.C. and Jeeva, S.C., "Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques," *Informatics in Medicine Unlocked*, vol. 16, 2019, <https://doi.org/10.1016/j.imu.2019.100203>.
- [12] Mohan, S., Thirumalai, C. and Srivastava, G., "Effective heart disease prediction using hybrid machine learning techniques," *IEEE Access*, vol. 7, pp. 81542-81554, 2019, doi: 10.1109/ACCESS.2019.2923707.
- [13] Ahmad T., Munir A., Bhatti S.H., Aftab M., Raza M. A., "Survival analysis of heart failure patients: a case study," *PLoS ONE*, vol. 12, no. 7, 2017, doi: 10.1371/journal.pone.0181001.
- [14] Aujla, Ravinder S., and Roshan Patel. "Creatine Phosphokinase. Treasure Island (FL): StatPearls Publishing; 2020 Jan-. PMID: 31536231. <https://www.ncbi.nlm.nih.gov/books/NBK546624/#article-20105.s7>
- [15] Brito, C., Esteves, M., and Peixoto, H., *et al.*, "A data mining approach to classify serum creatinine values in patients undergoing continuous ambulatory peritoneal dialysis," *Wireless Networks*, pp. 1-9, 2019. <https://doi.org/10.1007/s11276-018-01905-4>.
- [16] Chen, J. and Zeng, R., "Anemia," *In Handbook of Clinical Diagnostic*, Springer, Singapore, pp. 19-21, 2020, doi:10.1007/978-981-13-7677-1_6.
- [17] Butler J, Anker SD, Packer M., "Redefining Heart Failure with a Reduced Ejection Fraction," *JAMA*, vol. 322, no. 18, pp. 1761-1762, 2019, doi:10.1001/jama.2019.15600.
- [18] Vanessa A. Ravel, *et al.*, "Serum sodium and mortality in a national peritoneal dialysis cohort," *Nephrology Dialysis Transplantation*, vol. 32, no. 7, pp 1224-1233, July 2017, doi:10.1093/ndt/gfw254.
- [19] Osborn, G., "Mnemonic for hyperbolic formulae," *The Mathematical Gazette*, vol. 2, no. 34, pp. 189, July 1902, doi:10.2307/3602492.
- [20] Nair, Vinod, Hinton, Geoffrey E., "Rectified Linear Units Improve Restricted Boltzmann Machines," *27th International Conference on International Conference on Machine Learning, ICML'10, USA: Omnipress*, 2010, pp. 807-814.
- [21] Goodfellow, I. J, W. Farley, D. Mirza, *et al.*, "Maxout Networks," *JMLR Workshop and Conference Proceedings*, vol. 28, no. 3, pp. 1319-1327, 2013.
- [22] Clevert, D.A., Unterthiner, T. and Hochreiter, S., "Fast and accurate deep network learning by exponential linear units (elus)," *arXiv preprint arXiv:1511.07289*, 2015.
- [23] Moradi, R., Berangi, R. and Minaei, B., "A survey of regularization strategies for deep models," *Artif Intell Rev*, vol. 53, pp. 3947-3986. 2020. <https://doi.org/10.1007/s10462-019-09784-7>.
- [24] Faris, H., Mirjalili, S. and Aljarah, I., "Automatic selection of hidden neurons and weights in neural networks using grey wolf optimizer based on a hybrid encoding scheme," *Int. J. Mach. Learn. & Cyber*, vol. 10, no. 10, pp. 2901-2920, 2019, <https://doi.org/10.1007/s13042-018-00913-2>.
- [25] Lever, J., Krzywinski, M. and Altman, N., "Classification evaluation," *Nat Methods*, vol. 13, pp. 603-604, 2016, <https://doi.org/10.1038/nmeth.3945>.
- [26] K. H. Brodersen, C. S. Ong, K. E. Stephan and J. M. Buhmann, "The Balanced Accuracy and Its Posterior Distribution," *2010 20th International Conference on Pattern Recognition*, Istanbul, pp. 3121-3124, 2010, doi: 10.1109/ICPR.2010.764.
- [27] Eickholt, J., Cheng, J., "DNdisorder: predicting protein disorder using boosting and deep networks," *BMC Bioinformatics*, vol. 14, no. 88, 2013, <https://doi.org/10.1186/1471-2105-14-88>.
- [28] Mahmood, M., Al-Khateeb, B and Alwash, M., "A review on neural networks approach on classifying cancers," *International Journal of Artificial Intelligence*, vol. 9, pp. 317-326, 2020, doi:10.11591/ijai.v9.i2.pp317-326.
- [29] Jin, X., Xu, A., Bie, R. and Guo, P., "April. Machine learning techniques and chi-square feature selection for cancer classification using SAGE gene expression profiles," *In International Workshop on Data Mining for Biomedical Applications*, Springer, Berlin, Heidelberg, pp. 106-115, 2006, https://doi.org/10.1007/11691730_11.
- [30] I. Goodfellow, Y. Bengio, and A. Courville, "Deep learning", Book in preparation for MIT Press, 2016, [Online]. Available: <http://www.deeplearningbook.org>.
- [31] Lei Jimmy Ba, Brendan Frey., "Adaptive dropout for training deep neural networks," *In Proceedings of the 26th International Conference on Neural Information Processing Systems-Volume 2 (NIPS'13)*. Curran Associates Inc., Red Hook, NY, USA, pp. 3084-3092, 2013.
- [32] Chicco, D., Jurman, G., "Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone," *BMC medical informatics and decision making*, vol. 20, no.1, pp. 16, 2020.