RESEARCH ARTICLE

Heart Disease Prediction Using Support Vector Machine and Artificial Neural Network

Artificial Intelligence and Applications 2024, Vol. 2(1) 45–51 DOI: 10.47852/bonviewAIA3202823



Aishwarya Mondal¹, Banhishikha Mondal¹, Amaresh Chakraborty¹, Agnita Kar¹, Ayanty Biswas¹ and

Annwesha Banerjee Majumder^{1,*} 💿

¹Department of Information Technology, JIS College of Engineering Kalyani, India

Abstract: Heart-related illnesses, often known as cardiovascular diseases, have been the leading cause of mortality globally over the past several decades and are now recognized as the most major illness in both India and the rest of the globe. The severity out of the disease can be avoided with proper care at proper stage. This disease claims early and accurate prediction to avoid causalities. As proper medical support is not adequate, diseases are not being identified at the proper time and treatment cannot be started. Machine learning algorithms have shown promise in predicting heart disease risk based on patient data. In this study, a machine learning-based heart disease prediction model has been presented. The objective of the work is to build a machine learning-based model for early and adequate prediction of heart disease. The proposed model has utilized support vector machine and artificial intelligence with an accuracy of 81.6% and 86.6%, respectively. The findings show that the model predicts heart disease risk with excellent accuracy, sensitivity, and specificity, offering healthcare professionals a useful tool to pinpoint people who may be more at risk of developing heart disease.

Keywords: artificial intelligence, artificial neural network, machine learning, support vector machine

1. Introduction

Heart disease is one of the leading causes of death. There are many symptoms through which heart disease can be identified, such as headache, angina, swollen legs, fatigue, and many more. Lifestyle issues such as food habits, lack of physical activities, and presence of other diseases like high blood pressure play also a major role in developing heart disease. Number of efficient medical practitioners are not adequate over the globe, which is an alarming issue in the healthcare sector. Machine learning has emerged as a powerful tool in healthcare, particularly in the field of disease prediction. By leveraging large amounts of patient data and sophisticated algorithms, machine learning models can be trained to accurately predict the likelihood of various diseases in individuals. This has the potential to revolutionize healthcare by enabling earlier detection and more personalized treatment of diseases.

This issue has been addressed through the proposed model for identifying heat disease under artificial intelligence and machine learning.

The proposed model has been developed applying support vector machine (SVM) and artificial intelligence. The dataset has been collected from Kaggle. SVM has been considered as very effective in high-dimensional dataspace and memory efficient. Artificial neural network (ANN) has been considered as capable of working on incomplete knowledge set as well and fault tolerant also.

The major contribution through this work is developing an intelligent model for the prediction of heart disease at an early stage to accelerate the treatment process which will help the society. The major identification is that applying ANN the model has achieved 86.6% accuracy and by applying SVM the model has achieved 81.6% accuracy. Artificial intelligence has superseded the SVM. The novelty of the work is the combination of correlation coefficient-based feature selection and artificial intelligence and SVM as classifier which is capable of heart disease prediction at an early stage.

In Section 2, background study has been carried out. In Section 3, the proposed model has been represented followed by result analysis in Section 4.

2. Literature Review

Medical facilities have carried out numerous experiments on disease prediction systems using a variety of data mining techniques and machine learning algorithms. Multiple linear regression is suitable for forecasting the likelihood of developing heart disease, as shown by the suggested prediction of heart

^{*}Corresponding author: Annwesha Banerjee Majumder, Department of Information Technology, JIS College of Engineering Kalyani, India. Email: annwesha.banerjee@gmail.com

[©] The Author(s) 2023. Published by BON VIEW PUBLISHING PTE. LTD. This is an open access article under the CC BY License (https://creativecommons.org/licenses/by/4.0/).

disease using multiple regression model (Polaraju & Durga Prasad, 2017). Seema & Deepika (2016) suggested a model to predict chronic disease utilizing naive bayes (NB), decision trees, SVMs, and ANNs to mine the data from past health records (ANN). SVM exhibits the highest accuracy rate in this experiment. In Dwivedi (2016), various algorithms, including NB, classification tree, k nearest neighbor (KNN), logistic regression, SVM, and ANN, were advised. In comparison to other algorithms, the logistic regression provides superior accuracy. Shetty & Naik (2016) suggested a system that uses patient medical data to identify cardiac problems. The system has been built with consideration for 13 input attribute risk factors. Data integration and cleaning were done following the study of the dataset's data.

Banerjee Majumder et al. (2021) proposed an ensemble machine learning model for heart disease prediction. The model utilized bagging technique with base learners logistic regression, NB, and KNN. Boukhatem et al. (2022) proposed a model for heart disease prediction using multilayer perceptron (MLP), SVM, random forest (RF), and NB. A heart disease prediction model was proposed by Jindal et al. (2020) where KNN and logistic regression. Utilizing RF, SVM, NB, and decision tree, a model was proposed where highest accuracy was achieved by RF (Sharma et al., 2020). Rajdhan et al. (2020) proposed a model for heart disease prediction applying NB, Decision Tree, logistic regression, and RF. RF achieved highest accuracy of 90.16%. Yazdani et al. (2021) proposed a novel model of heart disease prediction applying weighted associative rule mining. Bharti et al. (2021) proposed a deep learning-based model. Isolated Forest was used for feature selection and the model achieved 94.2% accuracy. Applying SVM, convolutional neural network (CNN), and booting methodology, a model was proposed where SVM outperformed over other methods (Krittanawong et al., 2020).

Karthick et al. (2022) proposed a model for heart disease prediction applying SVM, Gaussian naïve bayes light GBM, RF, and XGBoost.

Latha & Jeeva (2019) proposed a model for heart disease risk applying ensemble mechanism.

Sarra et al. (2022) proposed a model using SVM and applied X2 for feature selection.

A work on heart disease prediction was presented by Raut et al. (2020) to find a suitable and computational efficient model working on UCI dataset.

Pathak et al. (2022) proposed a model for COVID-19 prediction using deep transfer learning method.

Soni et al. (2022) proposed a model for lung disease prediction applying integrated space transfer network and CNN.

Tuli et al. (2020) proposed a healthcare support system for heart disease prediction utilizing ensemble mechanism.

Ahmed et al. (2022) proposed a prediction model of cardiovascular disease using many machine learning mechanisms such as KNN, NB, RF, and Light Gradient Boosting Mechanism (LGBM). LGBM achieved highest accuracy.

Gupta & Banerjee (2015) proposed the level of risk of cardiovascular disease under bioinformatics framework considering other different factors such as lifestyle and food habit.

Gudadhe et al. (2010) proposed a model for heart disease prediction using SVM and MLP neural network.

Shah et al. (2020) undertook a study with the goal of using machine learning to create a model for the prediction of cardiovascular illness. The Cleveland heart disease dataset, which contained 303 cases and 17 attributes, was the source of the information used for this project. It was located in the UCI



machine learning repository. NB, decision trees, RFs, and KNN are just a few of the supervised categorization techniques used by the authors. The study's findings showed that, with a degree of accuracy of 90.8%, the KKN model performed the best.

The major goal of the study conducted by Hasan & Bao (2020) was to determine the best effective feature selection strategy for predicting cardiovascular disease. The results showed that the XGBoost classifier in combination with the wrapper strategy offered the most accurate prediction results for cardiovascular sickness. Accuracy was delivered by XGBoost at 73.74%, SVC at 73.18%, and ANN at 73.20%.

3. Proposed Model

This is a model for heart disease prediction carried out on the dataset collected from Kaggle. The proposed model has been built using SVM and ANN. Model block diagram is shown in Figure 1.

3.1. Heart failure dataset

The work is based on the dataset collected from Kaggle. There are 13 columns in total. Among those, 12 are independent features using which prediction will be done. Here, patients from age 40 to 95 have been selected in this dataset. Male patients are denoted by a gender value 1 and female patients are denoted by a gender value 0. The dataset summary is shown in Figure 2.

Before training the model, dataset has to be prepressed properly. It has been observed that there is an absence of null values in the dataset which is shown in Figure 3. Outliers from the dataset have also been removed in this phase.

	age	anaemia	creatinie_phosphokinase	diabetes	ejection_fraction	high_blood_pressure
0	75.0	0	582	0	20	1
1	55.0	0	7861	0	38	0
2	65.0	0	146	0	20	0
3	50.0	1	111	0	20	0
4	65.0	1	160	1	20	0

Figure 2 Dataset description

Figure 3 Null value checking

In [7]:	<pre>dataset.isnull().sum()</pre>				
Out[7]:	age	0			
	anaemia	0			
	creatinine_phosphokinase	0			
	diabetes	0			
	ejection_fraction	0			
	high_blood_pressure	0			
	platelets	0			
	serum_creatinine	0			
	serum_sodium	0			
	sex	0			
	smoking	0			
	time	0			
	DEATH_EVENT	0			
	dtype: int64				

3.2. Applying feature engineering and perform feature selection

All features in any dataset are not equally important and even leading to negative influence in decision making. So feature selection is an important step in any decision-making system. Heat map has been applied for the selection of relevant features in this suggested model.

3.2.1. Correlation heat map:

Heat map represents a two-dimensional information with the help of colors to identify the correlation of different features. Highly correlated features should be removed from the dataset for better performance. The heat map produced on the dataset is shown in Figure 4.

We have checked the correlation using heat map. If two variables are highly correlated, we have dropped one of them. Here, it is checked which features are having correlation > 0.1. Rest of the features are dropped. The sample code snippet for feature selection is shown in Figure 5.

3.3. Model development

In this proposed work, SVM and ANN have been applied.

3.3.1. Support vector machine:

One of the most well-known methods for supervised learning is the SVM. This is used in machine learning to solve issues involving both classification and regression.

The SVM algorithm's objective is to establish the best line of decision boundary using which n-dimensional space can be divided into categories, so that we may conveniently classify additional data points in the future. This ideal decision boundary is referred to as a hyperplane. The extreme vectors or points that help create the hyperplane are chosen via SVM. These extreme circumstances are described by support vectors. The SVM algorithm is so named for this reason. After applying SVM in our model, the obtained accuracy is 81.6%. The developed model applying SVM is shown in Figure 6.

3.3.2. Artificial neural network:

A branch of artificial intelligence influenced by biology and modeled after the brain is referred to as ANNs. A computational network called an ANN is typically based on the biological neural network that created the structure of the human brain. The neurons in ANNs are also coupled to one another at different layers of the networks, just as the neurons in the human brain. These neurons are referred to as nodes. The developed ANN model is shown in Figure 7. The proposed ANN model has used rectified linear activation (ReLu) and sigmoid activation function for input layer and hidden layer, respectively. The initializer is used in the uniform initializer. Adam and binary_crossentropy have been used as optimizer and loss function, respectively.

The fitting of the model is shown in Figure 8.

After applying ANN in our model, 86.6% accuracy has achieved which is shown in Figure 9. Along with accuracy, other metrics considered are specificity, sensitivity, and precision. The model has achieved 93.54% precision, 87.87% sensitivity, and 83.33% specificity.

4. Result Analysis

The proposed model has been implemented applying SVM and ANN. ANN has achieved better accuracy over SVM.

4.1. Support vector machine:

The SVM classifier is fitted with the training set. Sklearn.svm package has been used to develop the SVM classifier. Since the SVM can be separated linearly Kernel='Linear' has been chosen. The model performance has been evaluated through generated confusion matrix, and accuracy metric has been calculated based on that. To create the confusion matrix, confusion matrix function of sklearn is imported. The confusion matrix based on the classification of SVM is shown in Figure 10.

The calculated accuracy for our project is 81.6%. Achieved precision, sensitivity, and specificity scores are 97.14%, 77.27%, and 93.75%, respectively. The R2 score is 0.25.

4.2. Artificial neural network:

In ANN, sequential model type is used to build a model layer by layer. Each layer consists of weights that correspond to the layer that allows it. Layers are added with activation function ReLu and sigmoid. Units are mentioned as per need. Lastly, the model is compiled with optimizer "adam" and loss function = "binary_crossentopy." The model is trained with train set where epochs = 100 and batch_size is 8. The performance of the proposed ANN model has been justified through confusion matrix and then accuracy value has been shown. The confusion matrix generated through the ANN is shown in Figure 11.

The calculated accuracy for our project is 86.66% which is shown in Figure 12. Achieved precision, sensitivity, and

Figure 4 Heat map																
In [29]:	plt.figure(figsize=(12,6)) sns.heatmap(dataset.corr(),cmap=' <mark>seismic'</mark> , annot = True)															
Out[29]:	<matplotlib.axes< td=""><td>subplo</td><td>ts.Ax</td><td>esSubp</td><td>lot a</td><td>t 0×1</td><td>bba186</td><td>3dc8></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></matplotlib.axes<>	subplo	ts.Ax	esSubp	lot a	t 0×1	bba186	3dc8>								
	age	1	0.088	-0.082	-0.1	0.06	0.093	-0.052	0.16	-0.046	0.065	0.019	-0.22	0.25	-1.0	
	anaemia	0.088	1	-0.19	-0.013	0.032	0.038	-0.044	0.052	0.042	-0.095	-0.11	-0.14	-0.066	-0.8	
	creatinine_phosphokinase	-0.082	-0.19	1	-0.0096	-0.044	-0.071	0.024	-0.016	0.06	0.08	0.0024	-0.0093	0.063		
	diabetes	-0.1	-0.013	-0.0096	1	-0.0049	-0.013	0.092	-0.047	-0.09	-0.16	-0.15	0.034	-0.0019	-0.6	
	ejection_fraction	0.06	0.032	-0.044	-0.0049	1	0.024	0.072	-0.011	0.18	-0.15	-0.067	0.042	-0.27	-0.4	
	high_blood_pressure	0.093	0.038	-0.071	-0.013	0.024	1	0.05	-0.0049	0.037	-0.1	-0.056	-0.2	0.079		
	platelets	0.052	0.044	-0.024	-0.092	-0.012	-0.05	-0.041	-0.041	-0.19	-0.13	-0.028	-0.011	0.29	- 0.2	
	serum sodium	-0.046	0.042	0.06	-0.09	0.18	0.037	0.062	-0.19	1	-0.028	0.0048	0.088	-0.2	- 0.0	
	sex	0.065	-0.095	0.08	-0.16	-0.15	-0.1	-0.13	0.007	-0.028	1	0.45	-0.016	-0.0043		
	smoking	0.019	-0.11	0.0024	-0.15	-0.067	-0.056	-0.028	-0.027	0.0048	0.45	1	-0.023	-0.013	0.2	
	time	-0.22	-0.14	-0.0093	0.034	0.042	-0.2	-0.011	-0.15	0.088	-0.016	-0.023	1	-0.53	04	
	DEATH_EVENT	0.25	-0.066	0.063	-0.0019	-0.27	0.079	-0.049	0.29	-0.2	-0.0043	-0.013	-0.53	1	0.4	
		age	anaemia	creatinine_phosphokinase	diabetes	ejection_fraction	high_blood_pressure	platelets	serum_creatinine	serum_sodium	sex	smoking	time	DEATH_EVENT		

Figure 5 Feature selection

			SUDDORT VE
In [30]:	<pre>datacor = dataset.corr() data to react a she (data corr["DEATH ENGNT"])</pre>		SOFFORT VE
	<pre>Data_target = abs(datacor[DEAIn_tveni]) relevant_features relevant_features</pre>	In [35]:	<pre>from sklearn.sv from sklearn.me from sklearn.me</pre>
Out[30]:	age 0.253729		
	ejection_fraction 0.268603		# Burlaing a Su
	serum_creatinine 0.294278		svc_model = SVC
	serum_sodium 0.195204		svc_model.fit(x
	time 0.526964		anadiation - au
	DEATH_EVENT 1.000000		prediction = sv
	Name: DEATH_EVENT, dtype: float64		confusion_machi
			confusion_matri
In [31]:	<pre>dataset.drop('anaemia',axis=1,inplace=True)</pre>		
	<pre>dataset.drop('diabetes',axis=1,inplace=True) dataset.drop('high_blood_pressure',axis=1,inplace=True) dataset.drop('sex',axis=1,inplace=True) dataset.drop('sex',axis=1,inplace=True)</pre>	Out[35]:	array([[34, 1] [10, 15]
	dataset.urop(creatinine_phosphokinase ,axis=1,inplace=frue)		

dataset.drop('platelets',axis=1,inplace=True)
dataset.drop('smoking',axis=1,inplace=True)

Figure 6 SVM model

SUPPORT VECTOR MACHINE

In [35]:	<pre>from sklearn.svm import SVC from sklearn.metrics import confusion_matrix from sklearn.metrics import plot_confusion_matrix</pre>
	<pre># Building a Support Vector Machine on train data svc_model = SVC(C= .1, kernel='linear', gamma= 1) svc_model.fit(X_train, Y_train)</pre>
	<pre>prediction = svc_model .predict(X_test) confusion_matrix = confusion_matrix(Y_test,prediction)</pre>
	confusion_matrix
Out[35]:	array([[34, 1], [10, 15]], dtype=int64)
In [36]:	<pre># check the accuracy on the training set print(svc_model.score(X_train, Y_train)) print(svc_model.score(X_test, Y_test))</pre>
	0.8661087866108786 0.8166666666666667

Figure 7 ANN implementation

has also been considered which is 0.27. In this proposed model, both the SVM and ANN have been applied to predict heart disease considering their different advantages. In the result analysis section, the performance of both the classifiers has been analyzed in details and from there it has been observed that ANN has outperformed over the SVM. The accuracy, sensitivity, specificity, and precision scores of these two classifiers used in the proposed model are shown in Figure 13.

specificity scores are 93.54%, 87.87%, and 83.33%. The R2 score

Comparative analysis of our proposed work with some existing work is presented in Table 1.

from sklearn.preprocessing import StandardScaler
sc=StandardScaler()
x_train=sc.fit_transform(x_train)
x_test=sc.transform(x_test)

classifier=Sequential()

classifier.add(Dense(activation="relu",input_dim=12,units=8,kernel_initializer="uniform"))
classifier.add(Dense(activation="relu",units=14,kernel_initializer="uniform"))
classifier.add(Dense(activation="sigmoid",units=1,kernel_initializer="uniform"))
classifier.compile(optimizer='adam',loss='binary_crossentropy',metrics=['accuracy'])

Figu	re 8
Fitting	model

In [04]: Classifier.fic(X_crain,y_crain,bacch_size=0,epochs=10	In	[64]:	classifier.fit(x	train,y train,batch	size=8,epochs=100)
--	----	-------	------------------	---------------------	--------------------

Epoch 1/100 209/209 [========] - 1s 3ms/step - loss: 0.6911 - accuracy: 0.6699 Epoch 2/100 209/209 [=======] - 0s 1ms/step - loss: 0.6861 - accuracy: 0.6746 Epoch 3/100 209/209 [=======] - 0s 1ms/step - loss: 0.6752 - accuracy: 0.6746 Epoch 4/100 209/209 [=======] - 0s 2ms/step - loss: 0.6492 - accuracy: 0.6746 Epoch 5/100 209/209 [=======] - 0s 1ms/step - loss: 0.6071 - accuracy: 0.6842 Epoch 6/100 209/209 [=======] - 0s 1ms/step - loss: 0.5525 - accuracy: 0.7799 Epoch 7/100 209/209 [=======] - 0s 1ms/step - loss: 0.4989 - accuracy: 0.8038: 0s - loss: 0.5228 - accuracy: 0. Epoch 8/100 209/209 [=======] - 0s 1ms/step - loss: 0.4547 - accuracy: 0.8373 Epoch 9/100 209/209 [=======] - 0s 1ms/step - loss: 0.4236 - accuracy: 0.8373 Epoch 10/100 209/209 [========] - 0s 1ms/step - loss: 0.4236 - accuracy: 0.8373 Epoch 10/100 209/209 [========] - 0s 1ms/step - loss: 0.4236 - accuracy: 0.8373 Epoch 10/100	
209/209 [========] - 1s 3ms/step - loss: 0.6911 - accuracy: 0.6699 Epoch 2/100 209/209 [=======] - 0s 1ms/step - loss: 0.6861 - accuracy: 0.6746 Epoch 3/100 209/209 [=======] - 0s 1ms/step - loss: 0.6492 - accuracy: 0.6746 Epoch 4/100 209/209 [=======] - 0s 1ms/step - loss: 0.6071 - accuracy: 0.6842 Epoch 5/100 209/209 [=======] - 0s 1ms/step - loss: 0.5525 - accuracy: 0.6842 Epoch 6/100 209/209 [=======] - 0s 1ms/step - loss: 0.5525 - accuracy: 0.7799 Epoch 7/100 209/209 [========] - 0s 1ms/step - loss: 0.4989 - accuracy: 0.8038: 0s - loss: 0.5228 - accuracy: 0. Epoch 8/100 209/209 [========] - 0s 1ms/step - loss: 0.4547 - accuracy: 0.8373 Epoch 9/100 209/209 [========] - 0s 1ms/step - loss: 0.4236 - accuracy: 0.8373 Epoch 10/100 209/209 [========] - 0s 1ms/step - loss: 0.4236 - accuracy: 0.8373 Epoch 10/100 209/209 [========] - 0s 1ms/step - loss: 0.4236 - accuracy: 0.8373 Epoch 10/100	Epoch 1/100
Epoch 2/100 209/209 [========] - 0s 1ms/step - loss: 0.6861 - accuracy: 0.6746 Epoch 3/100 209/209 [=======] - 0s 1ms/step - loss: 0.6752 - accuracy: 0.6746 Epoch 4/100 209/209 [=======] - 0s 2ms/step - loss: 0.6492 - accuracy: 0.6746 Epoch 5/100 209/209 [=======] - 0s 1ms/step - loss: 0.6071 - accuracy: 0.6842 Epoch 6/100 209/209 [=======] - 0s 1ms/step - loss: 0.5525 - accuracy: 0.7799 Epoch 7/100 209/209 [=======] - 0s 1ms/step - loss: 0.4989 - accuracy: 0.8038: 0s - loss: 0.5228 - accuracy: 0. Epoch 8/100 209/209 [========] - 0s 1ms/step - loss: 0.4547 - accuracy: 0.8373 Epoch 9/100 209/209 [========] - 0s 1ms/step - loss: 0.4236 - accuracy: 0.8373 Epoch 10/100 209/209 [========] - 0s 1ms/step - loss: 0.4236 - accuracy: 0.8230	209/209 [====================================
209/209 [========] - 0s 1ms/step - loss: 0.6861 - accuracy: 0.6746 Epoch 3/100 209/209 [========] - 0s 1ms/step - loss: 0.6752 - accuracy: 0.6746 Epoch 4/100 209/209 [========] - 0s 2ms/step - loss: 0.6492 - accuracy: 0.6746 Epoch 5/100 209/209 [========] - 0s 1ms/step - loss: 0.6071 - accuracy: 0.6842 Epoch 6/100 209/209 [========] - 0s 1ms/step - loss: 0.5525 - accuracy: 0.7799 Epoch 7/100 209/209 [========] - 0s 1ms/step - loss: 0.4989 - accuracy: 0.8038: 0s - loss: 0.5228 - accuracy: 0. Epoch 8/100 209/209 [========] - 0s 1ms/step - loss: 0.4547 - accuracy: 0.8373 Epoch 9/100 209/209 [========] - 0s 1ms/step - loss: 0.4236 - accuracy: 0.8373 Epoch 10/100 209/209 [========] - 0s 1ms/step - loss: 0.4236 - accuracy: 0.8230	Epoch 2/100
Epoch 3/100 209/209 [==========] - 0s 1ms/step - loss: 0.6752 - accuracy: 0.6746 Epoch 4/100 209/209 [========] - 0s 2ms/step - loss: 0.6492 - accuracy: 0.6746 Epoch 5/100 209/209 [=======] - 0s 1ms/step - loss: 0.6071 - accuracy: 0.6842 Epoch 6/100 209/209 [=======] - 0s 1ms/step - loss: 0.5525 - accuracy: 0.7799 Epoch 7/100 209/209 [========] - 0s 1ms/step - loss: 0.4989 - accuracy: 0.8038: 0s - loss: 0.5228 - accuracy: 0. Epoch 8/100 209/209 [=========] - 0s 1ms/step - loss: 0.4547 - accuracy: 0.8373 Epoch 9/100 209/209 [========] - 0s 1ms/step - loss: 0.4236 - accuracy: 0.8373 Epoch 10/100 209/209 [========] - 0s 1ms/step - loss: 0.4236 - accuracy: 0.8373 Epoch 10/100	209/209 [========================] - 0s 1ms/step - loss: 0.6861 - accuracy: 0.6746
209/209 [========] - 0s 1ms/step - loss: 0.6752 - accuracy: 0.6746 Epoch 4/100 209/209 [=======] - 0s 2ms/step - loss: 0.6492 - accuracy: 0.6746 Epoch 5/100 209/209 [=======] - 0s 1ms/step - loss: 0.6071 - accuracy: 0.6842 Epoch 6/100 209/209 [=======] - 0s 1ms/step - loss: 0.5525 - accuracy: 0.7799 Epoch 7/100 209/209 [=======] - 0s 1ms/step - loss: 0.4989 - accuracy: 0.8038: 0s - loss: 0.5228 - accuracy: 0. Epoch 8/100 209/209 [========] - 0s 1ms/step - loss: 0.4547 - accuracy: 0.8373 Epoch 9/100 209/209 [=======] - 0s 1ms/step - loss: 0.4236 - accuracy: 0.8373 Epoch 10/100 209/209 [=======] - 0s 1ms/step - loss: 0.4236 - accuracy: 0.8373 Epoch 10/100	Epoch 3/100
Epoch 4/100 209/209 [========] - 0s 2ms/step - loss: 0.6492 - accuracy: 0.6746 Epoch 5/100 209/209 [========] - 0s 1ms/step - loss: 0.6071 - accuracy: 0.6842 Epoch 6/100 209/209 [=======] - 0s 1ms/step - loss: 0.5525 - accuracy: 0.7799 Epoch 7/100 209/209 [=======] - 0s 1ms/step - loss: 0.4989 - accuracy: 0.8038: 0s - loss: 0.5228 - accuracy: 0. Epoch 8/100 209/209 [========] - 0s 1ms/step - loss: 0.4547 - accuracy: 0.8373 Epoch 9/100 209/209 [========] - 0s 1ms/step - loss: 0.4236 - accuracy: 0.8373 Epoch 10/100 209/209 [========] - 0s 1ms/step - loss: 0.4236 - accuracy: 0.8230	209/209 [====================================
209/209 [==========================] - 0s 2ms/step - loss: 0.6492 - accuracy: 0.6746 Epoch 5/100 209/209 [====================] - 0s 1ms/step - loss: 0.6071 - accuracy: 0.6842 Epoch 6/100 209/209 [==================] - 0s 1ms/step - loss: 0.5525 - accuracy: 0.7799 Epoch 7/100 209/209 [==================] - 0s 1ms/step - loss: 0.4989 - accuracy: 0.8038: 0s - loss: 0.5228 - accuracy: 0. Epoch 8/100 209/209 [=================] - 0s 1ms/step - loss: 0.4547 - accuracy: 0.8373 Epoch 9/100 209/209 [================] - 0s 1ms/step - loss: 0.4236 - accuracy: 0.8373 Epoch 10/100 209/209 [===================] - 0s 1ms/step - loss: 0.3988 - accuracy: 0.8230	Epoch 4/100
Epoch 5/100 209/209 [=========] - 0s 1ms/step - loss: 0.6071 - accuracy: 0.6842 Epoch 6/100 209/209 [========] - 0s 1ms/step - loss: 0.5525 - accuracy: 0.7799 Epoch 7/100 209/209 [========] - 0s 1ms/step - loss: 0.4989 - accuracy: 0.8038: 0s - loss: 0.5228 - accuracy: 0. Epoch 8/100 209/209 [========] - 0s 1ms/step - loss: 0.4547 - accuracy: 0.8373 Epoch 9/100 209/209 [========] - 0s 1ms/step - loss: 0.4236 - accuracy: 0.8373 Epoch 10/100 209/209 [=========] - 0s 1ms/step - loss: 0.4236 - accuracy: 0.8373 Epoch 10/100	209/209 [====================================
209/209 [=========================] - 0s 1ms/step - loss: 0.6071 - accuracy: 0.6842 Epoch 6/100 209/209 [====================================	Epoch 5/100
Epoch 6/100 209/209 [=========] - 0s 1ms/step - loss: 0.5525 - accuracy: 0.7799 Epoch 7/100 209/209 [========] - 0s 1ms/step - loss: 0.4989 - accuracy: 0.8038: 0s - loss: 0.5228 - accuracy: 0. Epoch 8/100 209/209 [=========] - 0s 1ms/step - loss: 0.4547 - accuracy: 0.8373 Epoch 9/100 209/209 [========] - 0s 1ms/step - loss: 0.4236 - accuracy: 0.8373 Epoch 10/100 209/209 [=========] - 0s 1ms/step - loss: 0.4988 - accuracy: 0.8230	209/209 [====================================
209/209 [=========================] - 0s 1ms/step - loss: 0.5525 - accuracy: 0.7799 Epoch 7/100 209/209 [===================] - 0s 1ms/step - loss: 0.4989 - accuracy: 0.8038: 0s - loss: 0.5228 - accuracy: 0. Epoch 8/100 209/209 [=================] - 0s 1ms/step - loss: 0.4547 - accuracy: 0.8373 Epoch 9/100 209/209 [===============] - 0s 1ms/step - loss: 0.4236 - accuracy: 0.8373 Epoch 10/100 209/209 [================] - 0s 1ms/step - loss: 0.3988 - accuracy: 0.8230	Epoch 6/100
Epoch 7/100 209/209 [====================================	209/209 [====================================
209/209 [===========================] - 0s 1ms/step - loss: 0.4989 - accuracy: 0.8038: 0s - loss: 0.5228 - accuracy: 0. Epoch 8/100 209/209 [=====================] - 0s 1ms/step - loss: 0.4547 - accuracy: 0.8373 Epoch 9/100 209/209 [===================] - 0s 1ms/step - loss: 0.4236 - accuracy: 0.8373 Epoch 10/100 209/209 [=================] - 0s 1ms/step - loss: 0.3988 - accuracy: 0.8230	Epoch 7/100
Epoch 8/100 209/209 [====================================	209/209 [====================================
209/209 [===================] - 0s 1ms/step - loss: 0.4547 - accuracy: 0.8373 Epoch 9/100 209/209 [====================] - 0s 1ms/step - loss: 0.4236 - accuracy: 0.8373 Epoch 10/100 209/209 [====================] - 0s 1ms/step - loss: 0.3988 - accuracy: 0.8230	Epoch 8/100
Epoch 9/100 209/209 [====================] - 0s 1ms/step - loss: 0.4236 - accuracy: 0.8373 Epoch 10/100 209/209 [========================] - 0s 1ms/step - loss: 0.3988 - accuracy: 0.8230	209/209 [====================================
209/209 [==================] - 0s 1ms/step - loss: 0.4236 - accuracy: 0.8373 Epoch 10/100 209/209 [=====================] - 0s 1ms/step - loss: 0.3988 - accuracy: 0.8230	Epoch 9/100
Epoch 10/100 209/209 [=======================] - 0s 1ms/step - loss: 0.3988 - accuracy: 0.8230	209/209 [====================================
209/209 [====================================	Epoch 10/100
	 209/209 [====================================

Figure 9 Artificial neural network

y_pred=classifier.predict(x_test)
y_pred=(y_pred>0.5)

cm=confusion_matrix(y_test,y_pred)

cm

array([[58, 4], [8, 20]], dtype=int64)

 $acc=(cm[0][0]+cm[1][1])/(cm[0][1]+cm[1][0]+cm[0][0]+cm[1][1]) \\ print(acc*100) \\ \label{eq:acc}$

86.666666666666

Figure 11 Confusion matrix of artificial neural network

y_pred=classifier.predict(x_test)
y_pred=(y_pred>0.5)

cm=confusion_matrix(y_test,y_pred)

cm array([[58, 4], [8, 20]], dtype=int64)

Figure 12 Accuracy of artificial neural network

acc=(cm[0][0]+cm[1][1])/(cm[0][1]+cm[1][0]+cm[0][0]+cm[1][1]) print(acc*100)

86.66666666666667



Figure 10 Confusion matrix of support vector machine

SUPPORT VECTOR MACHINE

In [35]: from sklearn.svm import SVC from sklearn.metrics import confusion_matrix from sklearn.metrics import plot_confusion_matrix

Building a Support Vector Machine on train data
svc_model = SVC(C= .1, kernel='linear', gamma= 1)
svc_model.fit(X_train, Y_train)

prediction = svc_model .predict(X_test)
confusion_matrix = confusion_matrix(Y_test,prediction)

confusion_matrix

 Out[35]:
 array([[34, 1],

 [10, 15]],
 dtype=int64)

Proposed methods	Observations
Polaraju & Durga Prasad (2017)	Used multiple linear regression
Seema & Deepika (2016)	Used naive bayes, decision trees, support vector machines (SVM), and artificial neural networks
Shetty & Naik (2016)	Used artificial neural network and hybrid method
	Artificial neural network achieved 84%; hybrid model achieved 89% accuracy
Banerjee Majumder et al. (2021)	Used bagging technique with logistic regression, naive bayes and K nearest neighbor
Boukhatem et al. (2022)	Used multilayer perceptron (MLP), support vector machine (SVM), random forest (RF), and naïve bayes (NB)
	Achieved 82.8%, 82.5%, and 83.2%, respectively for logistic regression, naïve bayes and K nearest neighbor
Jindal et al. (2020)	Used KNN and logistic regression
	Achieved accuracy 88.5%
Rajdhan et al. (2020)	naive bayes, decision tree, logistic regression and random forest
	Highest accuracy achieved was 90.16%
Yazdani et al. (2021)	Used weighted associative rule mining Achieved 98% accuracy
Krittanawong et al. (2020).	Used support vector machine, convolutional neural network, and booting methodology
	Achieved 94.2% accuracy
Karthick et al. (2022)	Used SVM, Gaussian naïve bayes light GBM, random forest, and XGBoost
	Achieved 78.77% average accuracy
Pathak et al. (2022)	Used deep transfer learning method
	Achieved 92% average accuracy
Ahmed et al. (2022)	Used KNN, naïve bayes, random forest, LGBM
	LGBM achieved highest accuracy
Shah et al. (2020)	Used naive bayes, decision trees, random forests, and K nearest neighbor
	KNN achieved highest accuracy of 90.08%
Hasan & Bao (2020)	Used XGBoost, support vector machine, and artificial neural network
	XGBoost achieved 73.74%, SVC achieved 73.18%, and ANN achieved 73.20%
Our proposed model	Used support vector machine and artificial network
	SVM achieved 81.6% accuracy and ANN achieved 86.6% accuracy

Table 1 Existing works and our proposed work comparison

5. Conclusion

In this work, we have trained the machine using SVM and ANN. Applying SVM, accuracy achieved is 81.6%. The ANN model consists of hidden layers, and output of each layer is carried on to the next input which ensures greater accuracy. ANN model is giving an accuracy of 86.66%. This model can be utilized as a classifier for early and accurate prediction of heart disease. In future, we would utilize different ensemble mechanism and deep learning approach.

Acknowledgement

We are thankful to our organization to provide support to complete the work.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

References

- Ahmed, S., Shaikh, S., Ikram, F., Fayaz, M., Alwageed, H. S., Khan, F., & Jaskani, F. H. (2022). Prediction of cardiovascular disease on self-augmented datasets of heart patients using multiple machine learning models. *Journal of Sensors*, 2022, 1–21. https://doi.org/10.1155/2022/3730303
- Bharti, R., Khamparia, A., Shabaz, M., Dhiman, G., Pande, S., & Singh, P. (2021). Prediction of heart disease using a combination of machine learning and deep learning. *Computational Intelligence and Neuroscience*, 2021, 1–11. https://doi.org/10.1155/2021/8387680
- Banerjee Majumder, A., Gupta, S., & Singh, D. (2021). An ensemble heart disease prediction model bagged with logistic regression, Naïve Bayes and K Nearest neighbour. *Journal of Physics: Conference Series*, 2286(1), 012017. http://doi.org/10.1088/ 1742-6596/2286/1/012017
- Boukhatem, C., Youssef, H. Y., & Nassif, A. B. (2022). Heart disease prediction using machine learning. *Advances in Science and Engineering Technology International Conferences*, 1–6, http://doi.org/10.1109/ASET53988.2022.9734880.
- Dwivedi, A. K. (2016). Performance evaluation of different machine learning techniques for prediction of heart disease. *Neural Computing and Applications*, 29, 685–693. https://doi.org/ 10.1007/s00521-016-2604-1
- Gudadhe, M., Wankhade, K., & Dongre, S. (2010). Decision support system for heart disease based on support vector machine and artificial neural network. In 2010 International Conference on

Computer and Communication Technology, 741–745. https://doi.org/10.1109/ICCCT.2010.5640377.

- Gupta, S., & Banerjee Majumder, A. (2015). Proposed intelligent system to identify the level of risk of cardiovascular diseasesunder the framework of bioinformatics. In Gupta, S., Bag, S., Ganguly, K., Sarkar, I., Biswas, P. (Eds.), Advancements of Medical Electronics. Lecture Notes in Bioengineering, (pp. 3–12), Springer. https://doi.org/10. 1007/978-81-322-2256-9_1.
- Hasan, N., & Bao, Y. (2020) Comparing different feature selection algorithms for cardiovascular disease prediction. *Health Technology*, 11, 49–62. https://doi.org/10.1007/s12553-020-00499-2
- Jindal, H., Agrawal, S., Khera, R., Jain, R., & Nagrath, P. (2020). Heart disease prediction using machine learning algorithms. In *IOP Conference Series: Materials Science and Engineering*, *1022*(1), 012072. http://doi.org/10.1088/1757-899X/1022/1/012072
- Karthick, K., Aruna, S. K., Samikannu, R., Kuppusamy, R., Teekaraman, Y., & Ramesh Thelkar, A. (2022). Implementation of a heart disease risk prediction model using machine learning. *Computational and Mathematical Methods* in Medicine, 2022, 1–14. https://doi.org/10.1155/2022/6517716.
- Krittanawong, C., Virk, H. U. H., Bangalore, S., Wang, Z., Johnson, K. W., Pinotti, R., ..., & Tang, W. W. (2020). Machine learning prediction in cardiovascular diseases: a meta-analysis. *Scientific Reports*, 10(1), 16057. https://doi.org/10.1038/s41598-020-72685-1.
- Latha, C. B. C., & Jeeva, S. C. (2019) Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. *Informatics in Medicine Unlocked*, 16, 100203. https://doi.org/10.1016/j.imu.2019.100203
- Magar, R., Mamane, R., Raut, S., & Rupnar, V. S. (2020). Heart disease prediction using machine learning. *International Journal* of Emerging Technologies and Innovative Research, 7(6), 2081–2085. http://www.jetir.org/papers/JETIR2006301.pdf.
- Pathak, Y., Shukla, P. K., Tiwari, A., Stalin, S., Singh, S., & Shukla, P. K. (2022). Deep transfer learning based classification model for COVID-19 disease. *Innovation and Research in BioMedical Engineering*, 43(2), 87–92. https://doi.org/10. 1016/j.irbm.2020.05.003
- Polaraju, K., & Prasad, D. D. (2017). Prediction of heart disease using multiple linear regression model. *International Journal* of Engineering Development and Research Development, 5(4), 1419–1425.

- Rajdhan, A., Agarwal, A., Sai, M., Ravi, D., & Ghuli, P. (2020). Heart disease prediction using machine learning. *International Journal* of Engineering Research & Technology, 9(4), 659–662. http:// doi.org/10.17577/IJERTV9IS040614
- Sarra, R. R., Dinar, A. M., Mohammed, M. A., & Abdulkareem, K. H., (2022). Enhanced heart disease prediction based on machine learning and χ2 statistical optimal feature selection model. *Designs*, 6(5), 87. https://doi.org/10.3390/designs6050087.
- Seema, S., & Deepika, K. (2016). Predictive analytics to prevent and control chronic disease. In 2nd International Conference on Applied and Theoretical Computing and Communication Technology, 381–386. http://doi.org/10.1109/ICATCCT.201607912028
- Shetty, A., & Naik, C. (2016). Different data mining approaches for predicting heart disease. *International Journal of Innovative in Science Engineering and Technology*, 5, 277–281.
- Shah, D., Patel, S. B., & Bharti, S. K. (2020). Heart disease prediction using machine learning techniques. SN Computer Science, 1, 345. https://doi.org/10.1007/s42979-020-00365-y.
- Sharma, V., Yadav, S., & Gupta, M. (2020). Heart disease prediction using machine learning techniques. In 2nd International Conference on Advances in Computing, Communication Control and Networking, 177–181. http:// doi.org/10.1109/ICACCCN51052.2020.9362842.
- Soni, M., Gomathi, S., Kumar, P., Churi, P. P., Mohammed, M. A., & Salman, A. O. (2022). Hybridizing convolutional neural network for classification of lung diseases. *International Journal of Swarm Intelligence Research*, 13(2), 1–15. http://doi.org/10.4018/IJSIR.287544.
- Tuli, S., Basumatary, N., Gill, S. S., Kahani, M., Arya, R. C., Wander, G. S., & Buyya, R. (2020) HealthFog: an ensemble deep learning based Smart Healthcare System for Automatic diagnosis of heart diseases in integrated IoT and fog computing environments. *Future Generation Computer Systems*, 104, 187–200.
- Yazdani, A., Varathan, K. D., Chiam, Y. K., Malik, A. W., & Wan Ahmad, W. A. (2021). A novel approach for heart disease prediction using strength scores with significant predictors. *BMC Medical Informatics and Decision Making*, 21(1), 194. https://doi.org/10.1186/s12911-021-01527-5.

How to Cite: Mondal, A., Mondal, B., Chakraborty, A., Kar, A., Biswas, A., & Banerjee Majumder, A. (2024). Heart Disease Prediction Using Support Vector Machine and Artificial Neural Network. *Artificial Intelligence and Applications*, 2(1), 45–51. https://doi.org/10.47852/bonviewAIA3202823