A comparative study by using machine learning classifiers to enhance classification and prediction of heart failure disease

Shivanshi Shukla
Infosys Ltd
Usha Chouhan
MANIT: Maulana Azad National Institute of Technology
Sunil Kumar Suryawanshi
Atal Bihari Vajpayee Institute of Good Governace
Jyoti Kant Choudhari (jtchoudhary27@gmail.com)
Maulana Azad National Institute of Technology https://orcid.org/0000-0001-5903-3366

Research Article

Keywords: Classification, prediction, Heart Failure, Random Forest, Support vector machine

Posted Date: August 22nd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1773610/v1

A comparative study by using machine learning classifiers to enhance classification and prediction of heart failure disease

Shivanshi Shukla[1] · Usha Chouhan[2] · Sunil Kumar Suryawanshi[3] Jyoti Kant Choudhari*[2]
[1]Infosys Technologies Private Limited, Bangalore, India – 560100
[2]Department of Bioinformatics, Mathematics and Computer Applications, MANIT, Bhopal, MP, India – 462003
[3]Center for Knowledge Management and e-Governance Atal Bihari Vajpayee Institute of Good Governance and Policy Analysis, Bhopal, M.P., India – 462003
*Corresponding Email - jtchoudhary27@gmail.com

Abstract

Purpose

Heart failure is a complex clinical condition when the heart cannot provide blood with enough flow for the body's needs. It is a major clinical and public health problem. Even if heart failure is not yet diagnosed, it is important to get your health checked every three to six months. This study aims to improve the accuracy of diagnosing heart failure by using machine learning classifiers such as Recursive Feature Elimination (RFE) and Synthetic Minority Oversampling Technique (SMOTE).

Methods

Heart failure data has been acquired from the University of California, Irvine (UCI) repository. To improve the accuracy of diagnosing heart failure, we employed the following methods for this study: k-Nearest Neighbor, Naive Bayes (NB), Random Forest, XGBoost, Decision Tree (DT), Logistic Regression (LR), and Support Vector Machines (SVM). The model was validated using the F-measure and ROC-AUC (Receiver Characteristic Area Under Curve) methods.

Results

Support vector machines employing logistic regression as a feature selection strategy produced the most significant classification accuracy of 90%, while support vector machines utilising RF as a feature selection strategy showed an accuracy of 83%. We also have an accuracy of 90% in the random forest as a machine learning methodology with all of our features.

Conclusion

The small dataset size of the current research presents a challenge to everyone's ability to gain more accurate findings. Improved diagnostics for heart failure may be possible in the future

using our machine-learning classifier-based classification system. To accurately forecast heart failure, this is the easiest way to use and the most accurate.

Keywords Classification · prediction · Heart Failure · Random Forest · Support vector machine

1. Introduction

CVDs are the leading reason of death worldwide, claiming nearly 17.9 million persons yearly to World Health Organization (WHO). The conditions known as CVDs cardiac and blood vessel difficulties include coronary heart condition, brain-vascular disease, and rheumatic heart illness (Dahal and Gautam, 2020). Heart failure means the heart fails to circulate enough blood for other bodily organ requirements (Kemp and Conte, 2012). That will happen if the heart cannot obtain enough blood to fill it up. This is also possible if the heart is not strong enough to pump blood effectively. According to the Centers for Disease Control and Prevention, more than 6 million individuals in the United States have heart problems (Alonso et al., 2021). Machine learning evaluates and retrieves data from many perspectives (Jordan and Mitchell, 2015). The information gathered might be expanded to include the health industry in various ways. Today's healthcare industry generates massive amounts of data regarding patients, illnesses, and other issues(Dash et al., 2019). Machine learning is a set of methods that enables the discovery of hidden data patterns(Kourou et al., 2015). Many academics tried to uncover the most accurate ways for learning machines to establish the links between different cardiovascular illnesses and patient characteristics(Dwivedi, 2018; Pouriyeh et al., 2017) by staking, bagging and stacking on the heart failure Data-set. This study analyses multiple machine learning classification techniques, such as SVM, KNN, LR, NB, RF, Xgboost and DT, as well as their usage in combination, utilising five and ten-fold cross-validation as the testing method. The UCI archive of heart failure health reports dataset was used. To increase the accuracy of previous works by using the RFE process feature selection methodology and grappling with data imbalance in classification data using SMOTE. This paper mentioned a medical history dataset released by Ahmad and colleagues of cardiac failure patients(Ahmad et al., 2017). Ahmad and collaborators used the 2017 standard biostatistics-driven models to assess mortality and classify important variables from their medical history for patients with heart disease. Zahid and his colleagues are building two rival sex-based survival models in 2019, one for males and another for women. This inspired Jurman and Chicco to apply ML approaches to evaluate the viability of patients with heart failure who determined that is enough to predict survival of serum creatinine and ejection fraction on their own (Chicco and Jurman, 2020; Oladimeji and Oladimeji, 2020). The major goal of this work is to employ machine learning algorithms to anticipate the

survival of patients with heart failure by confronting the problem of unbalanced datasets and authorised characteristics to gain greater accuracy compared to previous studies.

2. Materials and Techniques

2.1 Datasets

The dataset for this research's heart failure medical data was received from the UCI. We examined the medical record of 299 sufferers with heart failure, including 105 women and 194 men, collected from Faisalabad Cardiology Institute and Allied Hospital in Faisalabad from Apr to Dec 2015. In their 40s and 50s, all the patients were present.

Table 1 Detailed description of data set

Sr. no.	Attribute	Details	Quantification	Scale
1	Creatinine phosphokinase (CPK)	CPK enzyme concentration in blood	mcg/L	[23-7861]
2	Ejection fraction	The percent of blood that leaves the heart with each beat	Percentage	[14-80]
3	Serum Creatinine	Creatinine levels in the blood	mg/dL	[0.50-9.40]
4	Serum Sodium	Sodium levels in the blood	mEq/L	[114-148]
5	Age	Years of age of the sufferer.	Years	[40-95]
6	Platelets	Platelets are found in the blood	kiloplatelets/mL	[25.01-850.0]
7	Anaemia	RBC or haemoglobin deficiency.	Boolean	[0/1]
8	Diabetes	Whether or not the patient is diabetic	Boolean	[0/1]
9	High Blood Pressure	When a patient has blood pressure	Boolean	[0/1]
10	Sex	[man = 1, woman = 0]	Boolean	[0/1]

11	Smoking	Whether the patient is a cigarette smoker	Boolean	[0/1]
12	Death event (Target)	[0 = Alive, 1 = Died]	Boolean	[0/1]
13	Time	Follow-up period	Days	[4-285]

The average follow-up duration was 130 days and ranged between 4 and 285 days (Oladimeji and Oladimeji, 2020). The dataset 13 characteristics were chosen that provide clinical, physical, and lifestyle results. (Table1).

CPK, age, and serum sodium are continuous variables, while Ejection Fraction (EF), platelets, and serum creatinine are binary variables. Platelets were divided into three quartile groups, and EF was split into 3 categories. In addition, A serum creatinine level greater than 1.5 indicates renal impairment. The effect of creatinine >1.5 vs. creatinine ≤ 1.5 on mortality was investigated. The hematocrit has assessed the anaemia extent in patients. The hematocrit of patients was used to determine the severity of their anaemia. McClellan et al. classify patients as anaemic if their hematocrit is less than 36 (the minimal normal threshold)(Ahmad et al., 2017). CPK is a protein found in your skeletal muscles, heart, and brain that helps to trigger chemical changes in your body (Aujla and Patel, 2022). CPK enters the bloodstream when muscle tissue deteriorates. Increased CPK concentrations indicate pain or injury to the heart or other muscles. Higher CPK levels might cause heart failure or blood damage in a patient.

A creatinine blood test measures the concentration of creatinine in the blood. Serum creatinine is a byproduct of the breakdown of creatine, which is found in the muscle. The amount of creatinine in your blood will be used by your doctor to determine how well your kidneys are working (Narayanan and Appleton, 1980). A serum sodium test is a common procedure that evaluates salt concentration in blood. (Ackerman, 1990). Sodium is required for nerve and muscle function and maintenance of normal cellular homeostasis in the human body. Many mechanisms in the body work together to maintain sodium balance in the body. Sodium is present in various foods and drinks, and it enters and exits the body via urine, faeces, and sweat. It is critical for health to consume the proper quantity of salt. The blood pressure will increase if too much salt takes in the meal. (Strazzullo and Leclercq, 2014) .

Heart failure may be indicated by a low sodium level in the circulation. The ejection fraction is the quantity of blood expelled from the left ventricle during each pulse.(Abebe et al., 2018;

Hajouli and Ludhwani, 2022). The death case function we use in our binary classification analysis represents a patient's death or survival at the end of an average follow-up time of 130 days (Ahmad et al., 2017). There are 203 living patients and 96 dying patients in the dataset discrepancy.

Table 2 Ranking of features

Features	Score	
	3-Feature using RFE-LR	3-Feature using RFE-RF
Age	1	2
Anaemia	4	7
Creatinine phosphokinase	2	5
High blood pressure	5	3
Diabetes	7	8
Ejection fraction	1	1
Sex	3	6
Platelets	9	1
Serum createnine	1	1
Serum sodium	6	4
Smoking	8	9

2.2 Feature Selection and Imbalance Dataset

Feature selection is crucial in machine learning to remove redundant and less significant features and handle an imbalanced dataset (Cai et al., 2018; Mojrian et al., 2020; Qin et al., 2019). In this study, methods for dealing with the problem of data imbalance have been discussed. Another term for this is "feature selection," Which refers to choosing specific data from an enormous collection(Crisostomo et al., 2016). Because of the same features, irrelevant and unnecessary traits may be eliminated.

The RFE technique was employed in this study. We describe and use a novel RFE-wrapped machine learning technique to select our models' attributes. Filter-based feature selection, on the other hand, assigns a score to each feature before selecting only those with the highest (or lowest) scores (Brownlee, 2020). This method begins with the whole set of traits and removes the unnecessary ones. Traits may be eliminated after desired attributes are attained. Among

the different characteristics we selected were 3, 4, and 5(Turgut et al., 2018). Table 2 shows the features acquired using RFE feature selection in three categories.

The dataset is unbalanced if the groups are not represented equally. In the targeted variable there is an imbalance in the dataset, with 203 cases that survive denoted as death = 0 and 96 cases indicated as death = 1. Predictive accuracy is a prominent measure to evaluate machine learning algorithms. However, this is insufficient when the data is uneven and/or the prices of several mistakes fluctuate considerably. The machine learning community dealt with the problem of imbalanced data in two ways. One way is to allocate various charges to various training scenarios. Another option is to resolve the training dataset by under-sampling the dominant category sample or oversampling the minor category (Mansourifar and Shi, 2020). Resampling the files is one of the most popular approaches to dealing with an unbalanced dataset. The two most common strategies for this are under-sampling and over-sampling. In most circumstances, over-sampling procedures are recommended over under-sampling strategies since under-sampling data destroys instances containing critical information. In this work, we used SMOTE to balance an imbalanced dataset. SMOTE is an over-sampling technique that creates false data for a small class. This strategy helps to avoid over-fitting, which is caused by irregular oversampling. It focuses on the feature space to create new instances by closely matching positive examples (Demidova and Klyueva, 2017).

2.3 Logistic Regression (LR)

A sort of classification algorithm which employs regression modeling to understand and estimate features inside a dataset is known as logistic regression. Logistic regression is a discriminatory classification approach for a real-value input vector. Calculating binary classification likelihood is the centre of the learning and estimation operations. The Bernoulli trial may calculate the likelihood P of a dichotomous occurrence, which can then be connected to the event under inquiry in logistic regression (Alotaibi, 2019). Logistic regression is utilised when the dependent variable (target) is categorical. The independent variables may be binary, nominal or polynomial (Alotaibi, 2019; Hosmer Jr et al., 2013).

$$Logit(p) = \ln \frac{p}{1-p} = \frac{(\text{prob:o f presenceo f characteristics})}{(\text{prob:o f absenceo f characteristics})} \quad (1)$$

2.4 Support Vector Machine (SVM)

SVMs have supervised learning algorithms for classification and regression analysis that use linked learning techniques to analyse data(Alotaibi, 2019; Suthaharan, 2016). The SVM methodology seeks to locate a hyperplane in N-dimensional space that distinguishes between variables (N= the number of characteristics)(Gandhi, 2019) . In recent years, machine learning techniques like SVM have been used extensively in clinical and biological research. The new practitioners as well as the researchers will benefit from this strategy(Choudhari et al., n.d.).

2.5 Random Forests (RF)

RF is an ensemble learning system that trains a large number of decision-making trees, followed by the output of classes (classification) or medium/average predictor (regression) in the individual tree, and is also known as the Random Decision Forest(Biau and Scornet, 2016). RF is an ensemble learner, a mechanism for producing and integrating the effects of a large number of classifiers. RF creates several CART trees trained on the first training data sample and searched via a randomly-chosen subset of input variables to assess separation. To measure separation RF generates many CART trees(Ray et al., 2020).

2.6 K-Nearest Neighbour (KNN)

The KNN methodology is a sort of instance-based learning that does not begin with a thorough theoretical model. It uses a distance or similarity function to compare unknown occurrences with known ones. Calculating the closeness between the unaccounted class entry and the knowledged training set and picking the class from the lowest level is the nearest neighbour in the instance space. The Kth minimum size is used to calculate the nearest neighbours. The class commonly exists with a plurality vote as the given input class of the nearest neighbour collection(Ray et al., 2020).

$$Distance\ (X, Z) = \sqrt{\sum_{i=1}^{n}(xi - zi)(xi - zi)^2} \qquad (2)$$

where xi denotes training data variables, and zi denotes input data variables.

2.7 Decision Tree (DT)

A DT is a data classification system that divides instance space recursively. A decision tree's internal nodes split the case space into various sub depending on a distinct element of the input attribute values. The decision tree inductors automatically build a decision tree for a given dataset (Maimon and Rokach, 2005). The decision tree is used to construct a model tree structure. It can deal with numerical and categorical results. It subdivides the data into

smaller chunks. This is a straightforward method that is frequently applied. In the decision tree, the root node is the most crucial indicator. A leaf node represents a grouping or judgement.

2.8 Naive Bayes (NB)

A probabilistic classifier, the Naive Bayes (NB), posits that the outcome of a predictor's value on a given class is unaffected by other predictors' values. Use the Naive Bayes technique to calculate the posterior likelihood of each class. The posterior likelihood of the given input class having the maximum posterior probability will be picked(Metsis et al., 2006; Mitchell, 2006; Ray et al., 2020). Maximum posteriori hypothesis:

$$P(c|x) = \frac{(P(x|c*P(c)}{P(x)} \qquad (3)$$

$P(c)$ is the hypothesis's prior likelihood. $P(x)$ is the likelihood that the evidence is true. $P(x|c)$ is the likelihood that the hypothesis is correct based on the evidence(Ray et al., 2020).

2.9 Extreme Gradient Boosting (XGBoost)

XG-Boost is used for supervised learning tasks. Training data with several attributes are necessary to estimate the target attribute. It is a machine learning methodology based on the decisions tree, which employs a boosters framework for gradients. Tianqi Chen started XGBoost as a Distributed (Deep) Machine Learning Community (DMLC) research initiative (Chen and Guestrin, 2016).

3 **Classification Performance Measurement**

Seven supervised machine learning algorithms were employed to classify heart disease samples. Using 5- and 10-fold cross-validation, the classification accuracy was assessed. The cross-validation technology was applied to measure the model's efficiency and strength(Dwivedi, 2018) [4]. The data are divided into five or ten folds, with one fold used during testing and the remainder for training (Ray et al., 2020). For classification models, the 8-consistency metrics were utilized. The samples that did not have heart failure were labelled as negative, while those with heart failure were labelled as positive(Dwivedi, 2018).

3.1 Confusion Matrix

The confusion matrix is a matrix that contains details about the algorithm's accurate and wrong predictions, as well as the actual condition(Turgut et al., 2018) [17]. Table 3 shows an error matrix. True Positive (TP) refers to the number of people who have died due to heart

failure after being predicted to do so. False Positive (FP) refers to the number of alive people anticipated to die from heart failure. True Negative (TN) refers to the number of people expected to be alive but not. False Negative (FN) refers to the number of people expected to live but die from heart failure.

Table 3 Error (Confusion) Matrix Table

Actual Result	Predicted Result	
	P	N
P	TP	FN
N	FP	TN

3.2 Recall/Sensitivity

The ratio between the accurately projected positive and the actual positive in a given set of data is known as sensitivity.

$$Sensitivity = \frac{TP}{(TP+FN)} \quad (4)$$

3.3 Classification Accuracy

The accuracy of our model is obtained by dividing the number of correct estimations by the total number of samples.

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (5)$$

3.4 Precision

The percentage of correctly identified positive results percent by the total number of positive results is known as precision.

$$Precision = \frac{TP}{(TP+FP)} \quad (6)$$

3.5 F1 Score

The F-score is calculated using the Pythagorean means of precision and recall, with the most significant value 1 and the poorest 0. Precision and recall both contribute equally to the F1 score. The F1 score equation is given below:

$$F1score = \frac{2}{Recall-1+precision-1} \quad (7)$$

3.6Specificity

Specificity is the ratio of accurately anticipated negative to actual negative in a given set of data.

$$Specificity = \frac{TN}{(TN+FP)} \qquad (8)$$

4 **Results**

In the methods and materials portion, the performances of seven machine learning methods were assessed with 11, 3, 4 and 5 Parameters to predict heart diseases. The number of live patients is 203, with a maximum of 299 samples and 96 deaths. Data samples were partitioned in 10 or 5 folds, each fold tested, and the remaining folds used during the cross-validation as training courses (Dwivedi, 2018). Classification results of all techniques with different feature selection and cross-validation are given in Table 4.

Figures 1, 2, and 3 show the confusion matrix of prediction results. The findings show that 3 features utilising RFE-LR (SVM) and all features (Random Forest) forecast the most significant number of TP (the number of records expected to die due to heart failure) and the most significant number of TN (the number of records classified as alive when they were genuinely alive) (Fig 1, Fig 3). The confusion matrix for three features using RFE-RF (Fig 2) yields the lowest TP in all models.

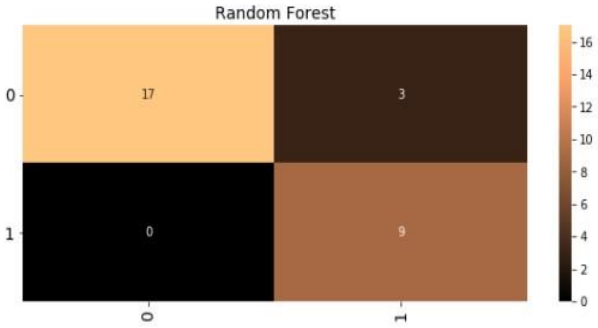

Fig. 1 confusion matrix for all features

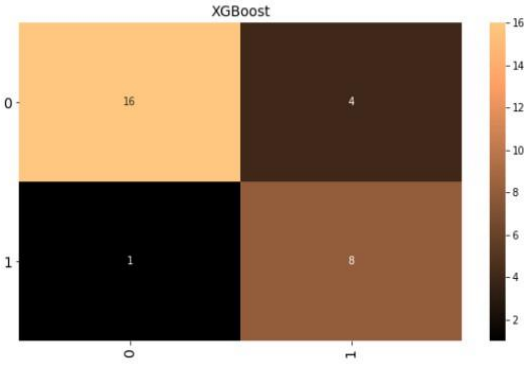

Fig. 2 confusion matrix for 3 features using RFE-RF

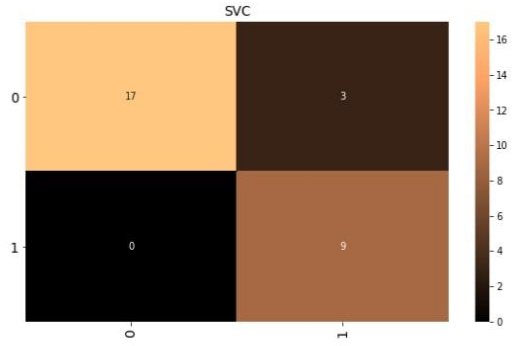

Fig. 3 confusion matrix for 3 features using RFE-LR

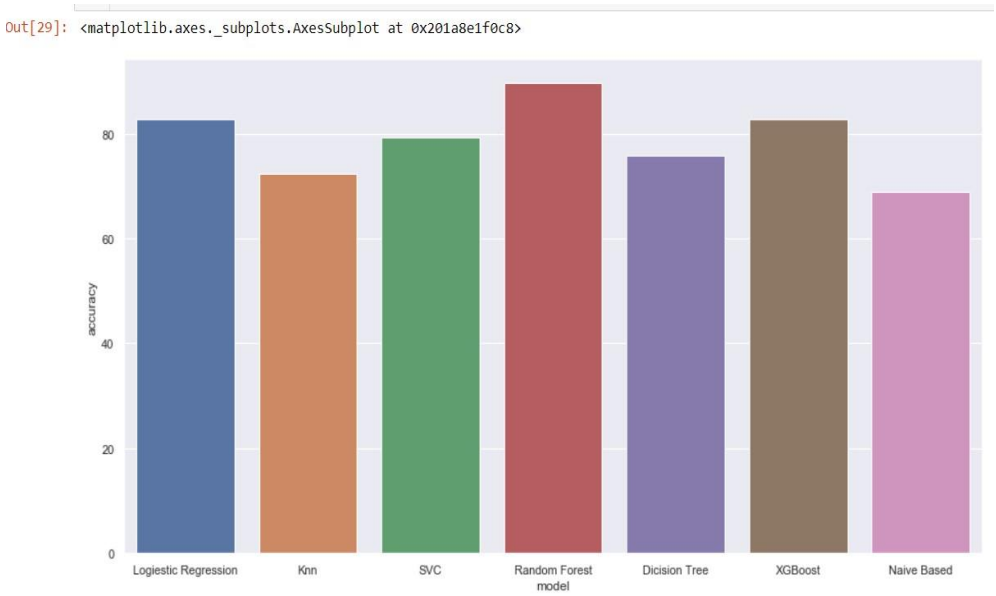

Fig. 4 Comparative result of different classification techniques using all features Table 4 All the accuracy obtained by different methods with 2 different cross validation

Feature Selection	Cross-Validation	No. of Features	KNN	SVC	RF	DT	XGBOOST	NB	LR
Logistic Regression	10	3	75.86	89.66	82.76	79.31	75.86	75.86	-
		4	79.66	88.14	71.19	59.32	69.49	72.88	-
		5	82.76	86.21	79.31	86.21	72.41	65.52	-

	5	3	74.58	88.14	72.88	77.97	66.1	77.97	-
		4	66.1	84.75	77.97	77.97	66.1	74.58	-
		5	79.66	83.05	76.27	74.58	67.8	76.27	-
Random Forest	10	3	65.52	79.31	-	75.86	82.76	68.97	79.31
		4	72.41	79.31	-	72.41	75.86	75.86	65.52
		5	79.31	72.41	-	82.76	65.52	58.62	68.97
	5	3	57.63	81.36	-	71.19	69.49	74.58	79.66
		4	74.58	83.05	-	61.02	62.71	77.97	81.36
		5	67.8	79.66	-	76.27	57.63	71.19	76.27
-	10	11(All)	72.41	79.31	89.66	75.86	82.76	68.97	82.76
-	5	11(All)	77.97	62.71	79.66	79.66	61.02	67.8	76.27

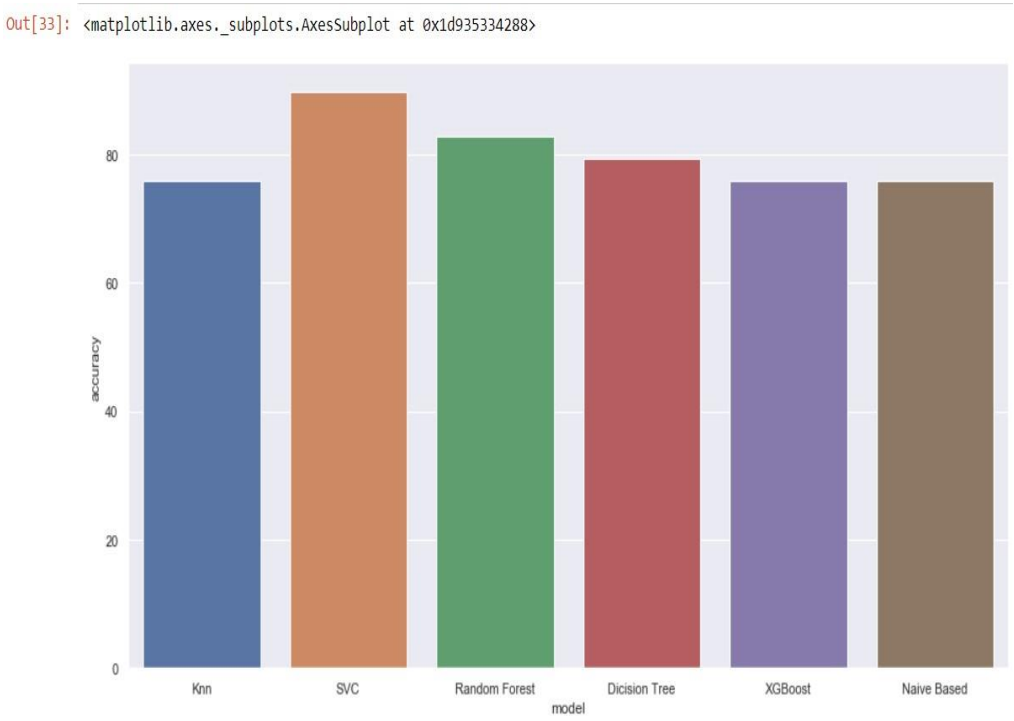

Fig. 5 Comparative result of different classification features using 3 features of logistic regression

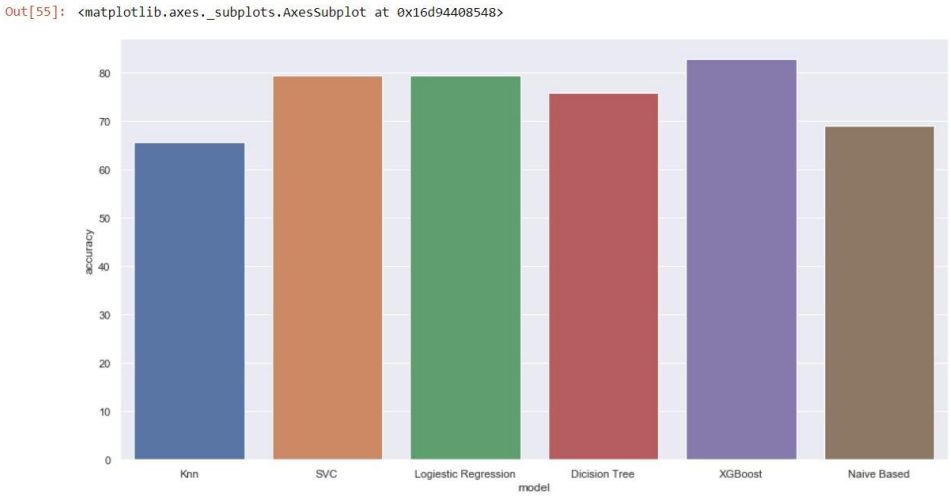

Fig. 6 Comparative result of different classification techniques using 3 features by random forest

Table 5 Comparisons of all accuracy measures for all features

Method			Accuracy	Precision
LR	0.827586	0.666667	Sensitivity	F-score
KNN	0.724138	0.545455	0.888889	0.761905
SVM	0.793103	0.615385	0.666667	0.600000

RF	0.896552	0.750000	0.888889	0.727273
DT	0.758621	0.562500	1.000000	0.857143
XGBOOST	0.827586	0.700000	1.000000	0.720000
NA¨IVE BASED	0.689655	0.500000	0.777778	0.736842

Table 6 Comparisons of all accuracy measures for 3 features using RFE-LR as feature selection

Method	Accuracy	Precision	Sensitivity	F-score
KNN	0.758621	0.625000	0.555556	0.588235
SVM	0.896552	0.750000	1.000000	0.857143
RF	0.827586	0.750000	0.666667	0.705882
DT	0.793103	0.714286	0.555556	0.625000
XGBOOST	0.758621	0.600000	0.666667	0.631579
NB	0.758621	0.625000	0.555556	0.588235

We will go through the survival prediction results we received on the entire data set before feature selection (Table 4) and then obtain the two most excellent accuracy results after feature selection (Tables 5, 7) using RFE LR and RFE RF. The primary measures employed in this investigation are confusion matrix, precision, recall, f1-score, accuracy and ROC. Precision is a statistic for determining the importance of data collection, whereas recall (or sensitivity) is a statistic for determining how frequently relevant results are returned. We evaluated our classifier using the F-measure. Three features utilising RFE LR had the highest accuracy, with all three features having a 90% accuracy with ten cross-validate. RFE RF is used to compare the outcomes with different feature selections and has a remarkable accuracy of 83%.

We compare many approaches to machine learning with graphs using seven machine learning approaches. As shown in Figure 4, the highest recall is obtained in all features and 3 feature selections in RFE LR, 1 (Table 5), 0.928 for 3 LR feature selection (Table 7), and 0.894 for 3 RF feature selection. Figure 5 compares the accuracy of three characteristics using logistic regression with different classification approaches, with SVC and RF obtaining above 80% accuracy. Figure 6 shows the accuracy of RF feature selection using various strategies in three features. With 10 cross-validations, the graph below displays the average greatest

precision attained by three features and all features using various machine learning approaches.

Table 7 Comparisons of all accuracy measures for 3 features using RFE-RF as feature selection

Method	Accuracy	Precision	Sensitivity	F-score
KNN	0.655172	0.470588	0.888889	0.615385
SVM	0.793103	0.615385	0.888889	0.727273
LR	0.793103	0.615385	0.888889	0.727273
DT	0.758621	0.571429	0.888889	0.695652
XGBOOST	0.827586	0.666667	0.888889	0.761905
NB	0.689655	0.500000	0.444444	0.470588

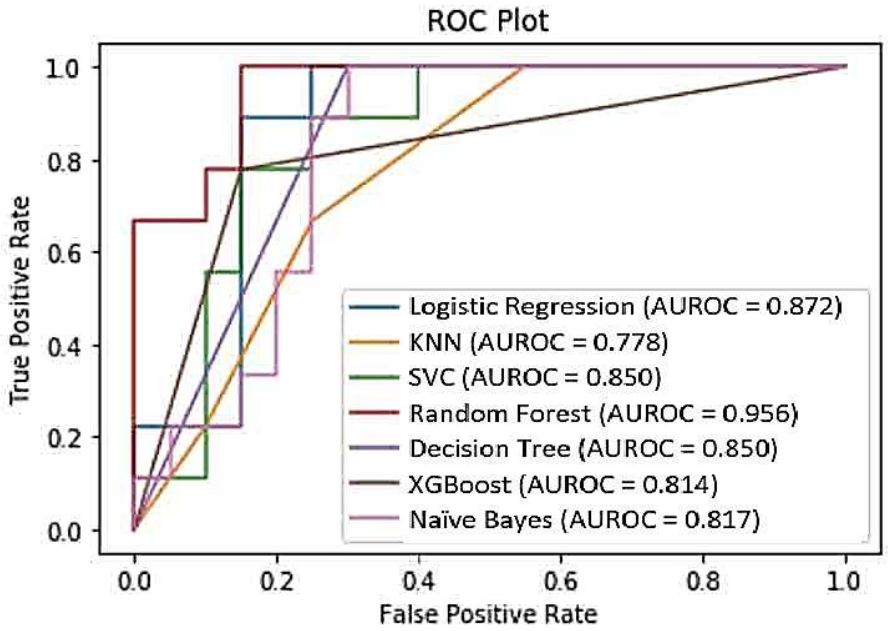

Fig. 7 ROC of All feature 10 cross validation

4.1 Performance Evaluation using ROC

The ROC curve is a graphical plot showing how a binary classification system's analytic capabilities change as the discrimination threshold changes. The ratio between sensitivity and specificity demonstrates the ROc plot. The sensitivity (also known as the true positive rate, TPR) is represented on the y-axis, while the false positive rate (FPR, also known as 1 specificity) is represented on the x-axis(Metz, 1978; Mitchell, 2006). The AUC varies from

0 to 1. While 1 represents a flawless approach, 0.5 represents an irregular approach, and 0.8 represents a satisfactory approach(Dwivedi and Chouhan, 2018; Ray et al., 2020).

The ROC curves of the machine learning technique on all features, 3 features of LR feature selection, and 3 features of Rf feature selection are compared in Figures 7, 8 and 9. It shows that the RF with all features has the highest AUC of 0.956 (Table 4),0.928 is the highest for 3 LR feature selection (Table 5), and 0.894 is the highest for 3 RF feature selection. The AUC scores of 3 main models seem to be greater than 0.8, indicating that the constructed models were implemented well. As a result, our data show that RF with all features generally performs better than alternative feature selection processes(Ray et al., 2020).

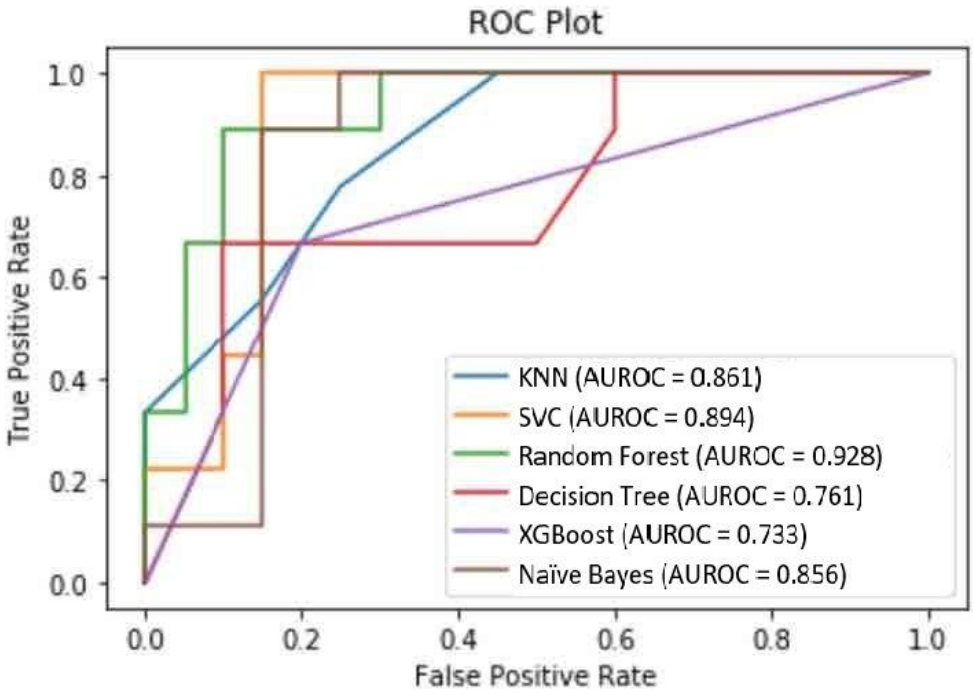

Fig. 8 ROC of 3 features LR 10 cross validation

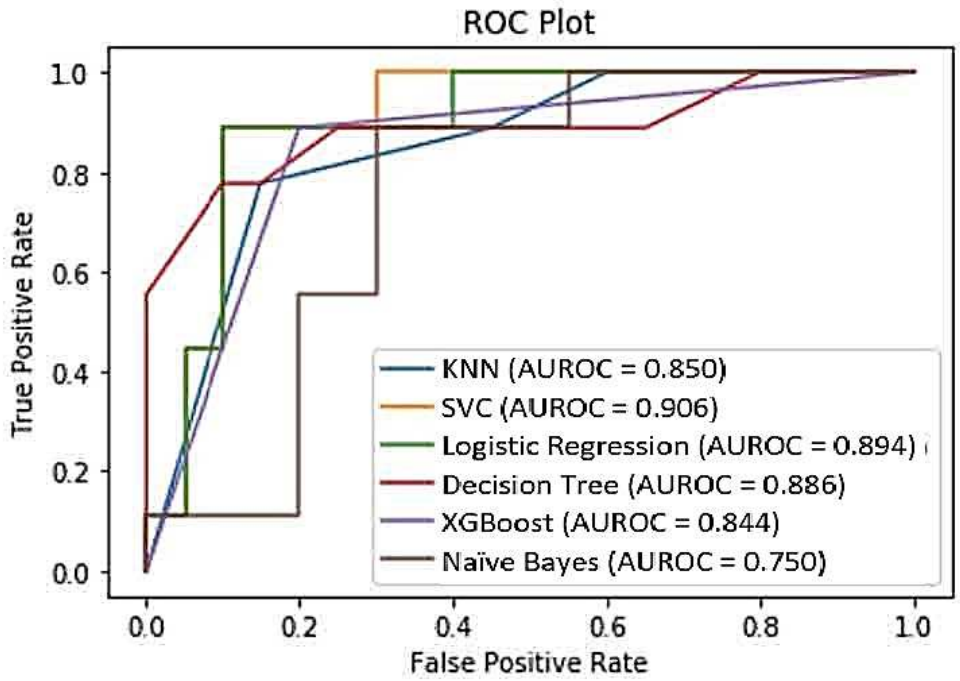

Fig. 9 ROC of 3 feature RF 10 cross-validation

5. Discussion

There are several machine learning algorithms to choose from, including neural networks, decision trees, Bayesian approaches, and support vector machines, all of which have a plethora of parameters that may be tailored to meet specific needs. Some trial and error is required when making a decision from a variety of options. Many researches worked deals with heart disease predictions with ML techniques. Among all the techniques combined feature model with Random Forest technique produces an accuracy of 81%. The proposed approach involves various machine learning techniques such as K-Nearest Neighbor, Naïve Bayes and Random Forest for classifying heart disease (Anbukkarasi et al., 2021). A study has suggested that random forest classifier algorithm can predict the heart disease with higher accuracy. The algorithm has an approximately 83% accuracy rate over training data(Chang et al., 2022). This study also suggested that random forest classifier algorithm can prediction the heart disease with higher accuracy. Ambale-Venkatesh et al. used random survival forest to predict six CV outcomes including new onset HF. While the model predicted incident HF with an AUC of 0.84, this was only a modest improvement compared to the MESA-HF score (0.8)(Ambale-Venkatesh et al., 2017). Frizzell et al. employed ML to predict 30-day all-cause readmissions in GWTG-HF patients. All models in this investigation had moderate discriminating power, with C statistics around 0.62. A study has shown that a model

developed by Golas et al. has the greatest AUC for 30-day readmission prediction for patients who are admitted to hospital with heart failure (HFM) rather than acute respiratory distress syndrome (FHS)(Golas et al., 2018). The model that we employed in this research has several potentially useful therapeutic applications. This model may be used by healthcare organisations and doctors to monitor patients with HF and assess the likelihood of readmission after hospitalisation for these patients.

ML algorithms, on the whole, displayed promising outcomes, despite the fact that there are considerable limits that need to be solved before ML algorithms can be used in clinical practise. In health informatics and medicine, ML algorithms are seeing widespread application and are producing positive outcomes(Sidey-Gibbons and Sidey-Gibbons, 2019; Verdonk et al., 2020). However, for the study to be interpreted in the appropriate clinical context, it is necessary to select the appropriate algorithms for the appropriate research questions, to compare the results to those of human experts, to validate the results using cohorts, and to report on all of the possible evaluation matrices. Several studies simply reported the technical features and none of the clinical aspects; this was probably owing to a lack of physician supervision.

5 Conclusion

In this study, a heart failure dataset is used to predict early onset of heart failure using seven classic machine learning approaches. These techniques include LR,SVM, key-value network, neural network, NB, DT, XGBOOST, and RF. We decided to employ the SMOTE method because of the skewed nature of the data. RFE selects features by using the Random Forest and Logiest regression techniques. The number of features used in feature selection ranges from three to five, with three offering superior performance over the other two. These procedures passed the 5- and 10-fold cross-validation tests to prove their reliability. The performance of RF with all characteristics is superior to that of RFE LR and RFE RF with just three characteristics. The results produced by SVM and random forest are the most accurate, coming in at 90 %. This is a broad topic with a number of different research resources, and it has the potential to provide the biological industry with an extra advantage or capability as a future exploration. The current research presents a challenge due to the small dataset size, which only includes 299 patients. A larger dataset would enable everyone to get more trustworthy conclusions. If our results had been made public, a supplementary external

data set that had the same geographical attributes as a replication cohort would have been employed to confirm our findings. This would have been the case only if our findings had been made public.

Conflicts of interest: the authors have no conflict of interest to declare.

References

Abebe, T. B., Gebreyohannes, E. A., Tefera, Y. G., Bhagavathula, A. S., Erku, D. A., Belachew, S. A., Gebresillassie, B. M., & Abegaz, T. M. (2018). The prognosis of heart failure patients: Does sodium level play a significant role? *PLoS ONE*, *13*(11), e0207242. https://doi.org/10.1371/journal.pone.0207242

Ackerman, G. L. (1990). Serum Sodium. In H. K. Walker, W. D. Hall, & J. W. Hurst (Eds.), *Clinical Methods: The History, Physical, and Laboratory Examinations* (3rd ed.). Butterworths. http://www.ncbi.nlm.nih.gov/books/NBK306/

Ahmad, T., Munir, A., Bhatti, S. H., Aftab, M., & Raza, M. A. (2017). Survival analysis of heart failure patients: A case study. *PloS One*, *12*(7), e0181001.

Alonso, A., Aparicio, F. H. J., Benjamin, E. J., Bittencourt, M. S., Callaway, C. W., Carson, F. A. P., Chamberlain, F. A. M., Kissela, B. M., Knutson, F. K. L., & Lee, C. D. (2021). Heart disease and stroke statistics—2021 update. *Circulation*, *2021*(143), e00–e00.

Alotaibi, F. S. (2019). Implementation of machine learning model to predict heart failure disease. *International Journal of Advanced Computer Science and Applications*, *10*(6), 261–268.

Aujla, R. S., & Patel, R. (2022). Creatine Phosphokinase. In *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK546624/

Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test*, *25*(2), 197–227.

Brownlee, J. (2020). Recursive Feature Elimination (RFE) for feature selection in Python. *Machine Learning Mastery*, *27*.

Cai, J., Luo, J., Wang, S., & Yang, S. (2018). Feature selection in machine learning: A new perspective. *Neurocomputing*, *300*, 70–79.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, 785–794.

Chicco, D., & Jurman, G. (2020). Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. *BMC Medical Informatics and Decision Making*, *20*(1), 1–16.

Crisostomo, J., Matafome, P., Santos-Silva, D., Gomes, A. L., Gomes, M., Patrício, M., Letra, L., Sarmento-Ribeiro, A. B., Santos, L., & Seiça, R. (2016). Hyperresistinemia and

metabolic dysregulation: A risky crosstalk in obese breast cancer. *Endocrine*, *53*(2), 433–442.

Dahal, K. R., & Gautam, Y. (2020). Argumentative Comparative Analysis of Machine Learning on Coronary Artery Disease. *Open Journal of Statistics*, *10*(4), 694–705. https://doi.org/10.4236/ojs.2020.104043

Dash, S., Shakyawar, S. K., Sharma, M., & Kaushik, S. (2019). Big data in healthcare: Management, analysis and future prospects. *Journal of Big Data*, *6*(1), 1–25.

Demidova, L., & Klyueva, I. (2017). SVM classification: Optimization with the SMOTE algorithm for the class imbalance problem. *2017 6th Mediterranean Conference on Embedded Computing (MECO)*, 1–4.

Dwivedi, A. K. (2018). Performance evaluation of different machine learning techniques for prediction of heart disease. *Neural Computing and Applications*, *29*(10), 685–693.

Dwivedi, A. K., & Chouhan, U. (2018). Comparative study of artificial neural network for classification of hot and cold recombination regions in Saccharomyces cerevisiae. *Neural Computing and Applications*, *29*(2), 529–535.

Gandhi, R. (2019). Support Vector Machine—Introduction to Machine Learning Algorithms. 2018. *URl: Https://Towardsdatascience. Com/Supportvector-Machine-Introduction-to-Machine-Learning-Algorithms-934a444fca47*.

Hajouli, S., & Ludhwani, D. (2022). Heart Failure And Ejection Fraction. In *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK553115/

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398). John Wiley & Sons.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255–260.

Kemp, C. D., & Conte, J. V. (2012). The pathophysiology of heart failure. *Cardiovascular Pathology*, *21*(5), 365–371.

Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. *Computational and Structural Biotechnology Journal*, *13*, 8–17.

Maimon, O., & Rokach, L. (2005). *Data mining and knowledge discovery handbook*.

Mansourifar, H., & Shi, W. (2020). Deep synthetic minority over-sampling technique. *ArXiv Preprint ArXiv:2003.09788*.

Metsis, V., Androutsopoulos, I., & Paliouras, G. (2006). Spam filtering with naive bayes-which naive bayes? *CEAS*, *17*, 28–69.

Metz, C. E. (1978). Basic principles of ROC analysis. *Seminars in Nuclear Medicine*, *8*(4), 283–298.

Mitchell, T. M. (2006). *The discipline of machine learning* (Vol. 9). Carnegie Mellon University, School of Computer Science, Machine Learning ….

Mojrian, S., Pinter, G., Joloudari, J. H., Felde, I., Szabo-Gali, A., Nadai, L., & Mosavi, A. (2020). Hybrid machine learning model of extreme learning machine radial basis function for breast cancer detection and diagnosis; a multilayer fuzzy expert system. *2020 RIVF International Conference on Computing and Communication Technologies (RIVF)*, 1–7.

Narayanan, S., & Appleton, H. D. (1980). Creatinine: A review. *Clinical Chemistry*, *26*(8), 1119–1126.

Oladimeji, O. O., & Oladimeji, O. (2020). Predicting survival of heart failure patients using classification algorithms. *JITCE (Journal of Information Technology and Computer Engineering)*, *4*(02), 90–94.

Pouriyeh, S., Vahid, S., Sannino, G., De Pietro, G., Arabnia, H., & Gutierrez, J. (2017). A comprehensive investigation and comparison of machine learning techniques in the domain of heart disease. *2017 IEEE Symposium on Computers and Communications (ISCC)*, 204–207.

Qin, J., Chen, L., Liu, Y., Liu, C., Feng, C., & Chen, B. (2019). A machine learning methodology for diagnosing chronic kidney disease. *IEEE Access*, *8*, 20991–21002.

Ray, U., Chouhan, U., & Verma, N. (2020). Comparative study of machine learning approaches for classification and prediction of selective caspase-3 antagonist for Zika virus drugs. *Neural Computing and Applications*, *32*(15), 11311–11328.

Strazzullo, P., & Leclercq, C. (2014). Sodium1. *Advances in Nutrition*, *5*(2), 188–190. https://doi.org/10.3945/an.113.005215

Suthaharan, S. (2016). Support vector machine. In *Machine learning models and algorithms for big data classification* (pp. 207–235). Springer.

Turgut, S., Dağtekin, M., & Ensari, T. (2018). Microarray breast cancer data classification using machine learning methods. *2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT)*, 1–3.