International Journal of Computing and Artificial Intelligence

E-ISSN: 2707-658X
P-ISSN: 2707-6571
IJCAI 2021; 2(2): 29-35
Received: 16-05-2021
Accepted: 18-06-2021

Bhuvaneswari M
Department of Computer Science, SDHR College, Tirupati, Andhra Pradesh, India

Boyella Mala Konda Reddy
Assistant Professor, Department of Computer Science, SDHR College, Tirupati, Andhra Pradesh, India

Corresponding Author:
Bhuvaneswari M
Department of Computer Science, SDHR College, Tirupati, Andhra Pradesh, India

An efficient feature reduction approach for arrhythmia disease detection utilizing SVM

Bhuvaneswari M and Boyella Mala Konda Reddy

DOI: https://doi.org/10.33545/27076571.2021.v2.i2a.35

Abstract
Arrhythmia is viewed as a hazardous infection causing genuine medical problems in patients, when left untreated. An early determination of arrhythmias would be useful in saving lives. This examination is led to group patients into one of the sixteen subclasses, among which one class addresses nonappearance of sickness and the other fifteen classes address electrocardiogram records of different subtypes of arrhythmias. The exploration is done on the dataset taken from the University of California at Irvine Machine Learning Data Repository. The dataset contains a huge volume of highlight measurements which are decreased utilizing SVM-RFE based component determination strategy. The dataset contains a huge list of capabilities which is diminished utilizing an improved component choice strategy named as covering technique. The proposed covering technique is based on a SVM-RFE calculation to choose the main highlights from the given dataset. The chose subset of highlights at that point goes through a preprocessing step to present a consistency in the dispersion of information. Since help vector machine (SVM) is perceived to have the advantage of giving an eminent execution in grouping stage.

Keywords: SVM, SVM-RFE, feature selection, arrhythmia and EEG

1. Introduction
In this day and age, individuals are experiencing different constant illnesses. Among them heart illnesses are found to influence a huge size of populace. An early identification and exact clinical help to coronary illness patients can save human lives as heart sicknesses can be dangerous causing abrupt demise. The most generally utilized instrument for diagnosing the capacity of heart is the electrocardiogram (ECG) recorded utilizing terminals puts on the body, which delivers a graphical example of the electrical driving forces of heart [1].

Arrhythmia is a type of inconsistency in heart rhythms and, sometimes, brings about coronary illness, which presents genuine dangers to living souls. It ought to be analyzed and treated as right on time as conceivable to diminish the danger of abrupt passing, since whenever left untreated it can likewise prompt a coronary episode. Arrhythmia is a sort of infection that upsets the smooth mood of heart's electrical framework and makes the heart beat either excessively lethargic or excessively quick, to race, and to skip pulsates and causes non sequential development of heart signals [13]. For the most part, arrhythmia is recognized and dissected from an ECG recording alongside the indications, for example, lacking siphoning of blood from heart, windedness, weakness, chest agony, and obviousness. In this manner, arrhythmia shows a sudden and abnormal ECG signal [2].

For building up an exact indicative framework, different AI strategies have been applied in the past to improve the precision of heart arrhythmia grouping from ECG signals [11, 12]. The determination of a proper method for arrhythmia grouping is an intricate assignment as it relies upon the setting of the application, information examination, prior encounters, and the prerequisites of the predetermined patients.

In this paper, we have proposed an effective framework, which can characterize ECG records into typical and ailing classes, that is, separating between the presence and nonappearance of arrhythmia.

1.1 The destinations of this paper are
- To study different highlights of Arrhythmia input dataset
- To apply the SVM Recursive Feature Elimination (SVM-RFE) for diminishing the quantity of qualities
- To arrange assaults utilizing Support Vector Machine (SVM)

2. Feature Selection
Highlight determination strategies decrease the dimensionality of highlight space, eliminate excess, immaterial or uproarious information. It brings the quick impacts for application: accelerating an information mining calculation, improving the information quality and the presentation of information mining and expanding the intelligibility of the mining results [6]. Highlight choice has been a functioning exploration region in information mining networks since it permits essentially improving the understandability of the subsequent classifier models [7]. It comprises to pick a subset of information factors from a dataset with enormous of characteristics by killing highlights with practically no prescient data.

Highlight choice strategies have their most extreme importance in information mining, AI, and example acknowledgment, particularly for enormous datasets [3]. The principle point of these procedures is to eliminate immaterial or excess highlights from the dataset. Highlight determination techniques have two classifications: covering and channel [10]. The covering assesses and chooses ascribes dependent on exactness gauges by the objective learning calculation. Utilizing a specific learning calculation, covering fundamentally look through the element space by excluding a few highlights and testing the effect of highlight exclusion on the expectation measurements [4]. The element that has huge effect in learning measure suggests it does matter and ought to be considered as an excellent component.

2.1 SVM Recursive Feature Elimination (SVM-RFE)
SVM-RFE highlight determination strategy was proposed in [5] to lead quality choice for disease order. Settled subsets of highlights are chosen in a successive in reverse disposal way, what begins with all the component factors and eliminates each element variable in turn. At each progression, the coefficients of the weight vector of a direct SVM are utilized to process the component positioning score. SVM-RFE technique positions every one of the highlights as per some score work and kills at least one highlights with the most minimal scores. This cycle is rehashed until the most noteworthy arrangement precision is acquired. Because of its effectively use in choosing enlightening qualities for malignant growth arrangement, SVM-RFE acquired an extraordinary fame and is notable as perhaps the best element choice strategy [9]. In any case, the SVM-RFE is an insatiable strategy that lone desires to track down the most ideal mix for grouping. The highlights are dispensed with as per a basis identified with their help to the segregation work, and the SVM [15] is re-prepared at each progression.

2.2 The RFE-SVM calculation [5] can be broken into four stages
1. Train a SVM on the preparation set;
2. Request highlights utilizing the loads of the subsequent classifier;
3. Kill highlights with the littlest weight;
4. Rehash the interaction with the preparation set limited to the leftover highlights.

3. Proposed Methodology
In this proposed structure, include decrease technique utilizing SVM-RFE is led as an underlying advance towards lessening the quantity of traits without losing the fundamental reason and target data from the first information. The following stage is building up an indicator with an improved precision to order informational index. There are various stages in the proposed design for an efficient Arrhythmia arrangement. We are proposing another model for proficient component determination and Arrhythmia expectation. This methodology is of the accompanying strides as follows:
- Stage 1: Read the Arrhythmia dataset.
- Stage 2: Preprocess the dataset.
- Stage 3: Select the huge highlights utilizing SVM-RFE calculation.
- Stage 4: Perform Classification utilizing SVM calculation on the dataset to choose the best highlights.
- Stage 5: Evaluate execution of the classifier.

3.1 Support Vector Machine (SVM)
The SVM is another kind of AI techniques dependent on factual learning hypothesis. In view of good advancement and a higher precision, SVM has become the exploration focal point of the AI people group. SVMs are set of related managed learning techniques utilized for grouping and relapse [15]. A few late investigations have detailed that the SVM for the most part are equipped for conveying better as far as order exactness than the other information arrangement calculations. SVM is based on factual learning hypothesis by Vapnik *et al.* [15] proposed another learning strategy, which is based on a set number of tests in the data contained in the current preparing text to get the best grouping results.

An uncommon property of SVM will be, SVM at the same time limit the experimental grouping mistake and augment the mathematical edge. So SVM called Maximum Margin Classifiers. SVM depends on the Structural danger Minimization. SVM map input vector to a higher dimensional space where a maximal isolating hyperplane is built. Two equal hyperplanes are developed on each side of the hyperplane that different the information. The isolating hyperplane is the hyperplane that boost the distance between the two equal hyperplanes. A supposition that is made that the bigger the edge or distance between these equal hyperplanes the better the speculation mistake of the classifier [12].

4. Experimental Details and Results
In this paper, another model is proposed for arranging arrhythmia patients utilizing the ECG dataset taken from UCI AI store [14]. The proposed model initially chooses the most distinctive highlights utilizing an improved component choice method, a covering calculation worked around SVM-RFE. Subsequent to choosing the huge highlights, SVM put together strategy are applied with respect to the chose include set to order the patients into sixteen subclasses of arrhythmia.

The analysis was led with double center 2.20 GHz with 4.00 Go of memory on windows stage, and we carried out the calculation utilizing WEKA [8]. WEKA represents Waikato Environment for Knowledge Analysis. WEKA is made by analysts at the University of Waikato in New Zealand. The product is written in the Java language and contains a GUI for communicating with information documents. WEKA additionally gives the graphical UI of the client and gives numerous offices. WEKA is a cutting-edge office for creating AI (ML) methods and their application to true information mining issues. The information record typically utilized by WEKA is in ARFF document design. ARFF represents Attribute Relation File Format, which comprises of extraordinary labels to demonstrate separating in the information document. WEKA implements algorithms for data pre-processing, classification, regression and clustering and association rules. It also includes visualization tools.

4.1 Dataset

This exploration is directed on the Cardiac arrhythmia dataset taken from the UCI AI store [14]. The dataset is made out of 452 examples ordered into 16 distinct classes and absolute number of traits are 279. The top of the line addresses the ordinary cases, while the other 15 classes address various kinds of arrhythmias including ischemic changes (coronary course infection), old front myocardial localized necrosis, old sub-par myocardial dead tissue, sinus tachycardia, sinus bradycardia, ventricular untimely compression, supraventricular untimely withdrawal, left group branch block, right pack branch block, first-degree atrioventricular (AV) block, second-degree AV block, third-degree AV block, left ventricular hypertrophy, and atrial fibrillation. The dataset has a sum of 279 characteristics for each given example where the initial four ascribes contain general data like age, tallness, sex, and weight, while the remainder of the properties are extricated from the ECG signals recorded by a standard 12-lead recorder including the P, Q, R, S, and T waves data. As the dataset has a huge arrangement of highlights, include determination is applied to choose the most important and huge highlights containing helpful data needed for information order. The point is to recognize the presence and nonappearance of heart arrhythmia and to arrange it in one of the 16 gatherings as demonstrated in the table-1. Class 1 alludes to 'typical' ECG classes 2 to 15 alludes to various classes of arrhythmia and class 16 alludes to the remainder of unclassified ones. We utilize 70% of records as the preparation information and the other 30% as the testing information.

Table 1: Different classes of Cardiac Arrhythmia Dataset

S. No	Class	No. of instances
1	Normal	245
2	Ischemic changes (coronary artery disease)	44
3	Old Anterior Myocardial Infarction	15
4	Old Inferior Myocardial Infarction	15
5	Sinus tachycardy	13
6	Sinus bradycardy	25
7	Ventricular Premature Contraction (PVC)	3
8	Supraventricular Premature Contraction	2
9	Left bundle branch block	9
10	Right bundle branch block	50
11	degree Atrio Ventricular block	0
12	degree AV block	0
13	degree AV block	0
14	Left ventricule hypertrophy	4
15	Atrial Fibrillation or Flutter	5
16	Others	22

i. Data Visualization

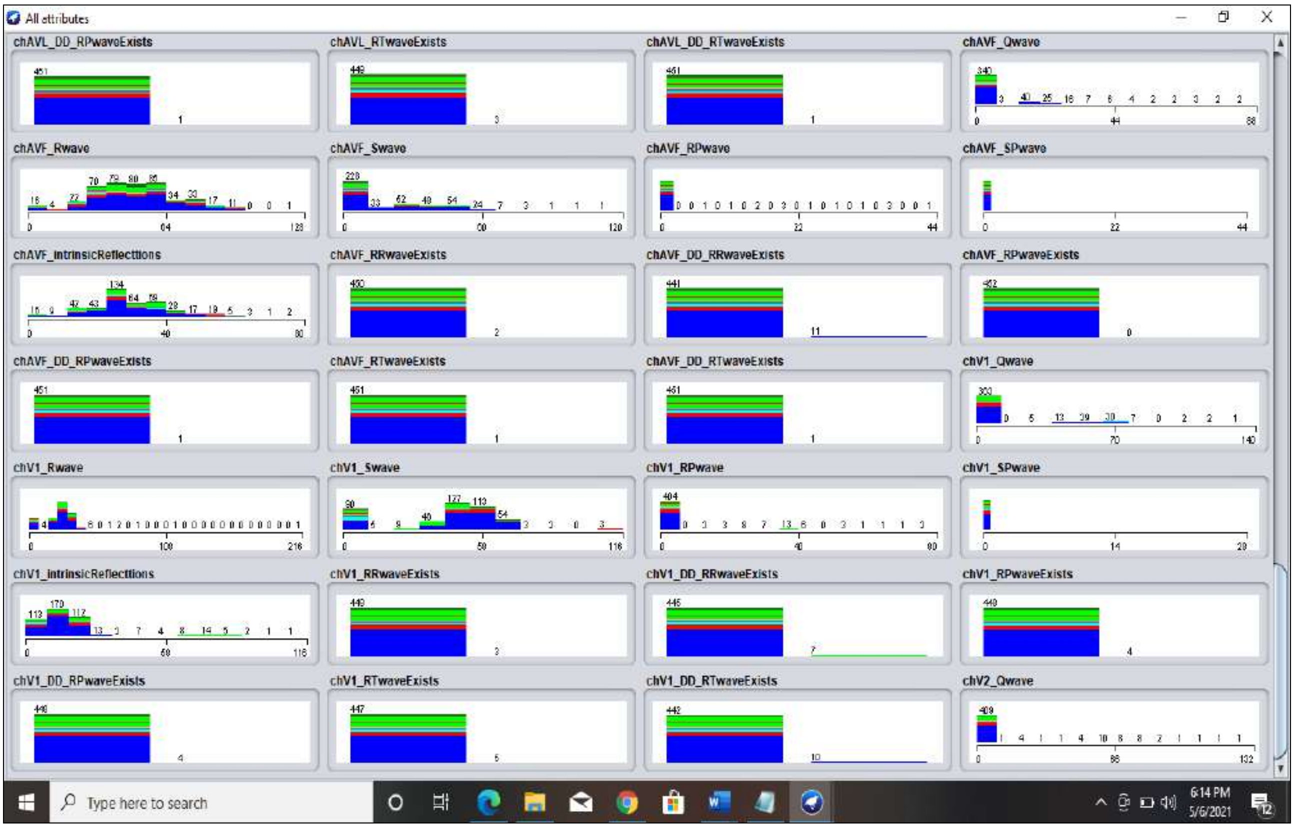


Fig 1: Statistical Summary of Dataset

ii. SVM-RFE Features

Fig 2: Statistical Summary of reduced Dataset

4.2 Results

In the first stage SVM algorithm is trained on the original set of features was used in the experiment. In the second stage we implement a SVM-RFE algorithm for obtaining the adequate number of features to identify the features selected. The results that we got for SVM without feature selection and with feature selection are shown below in figure-3 with their corresponding values.

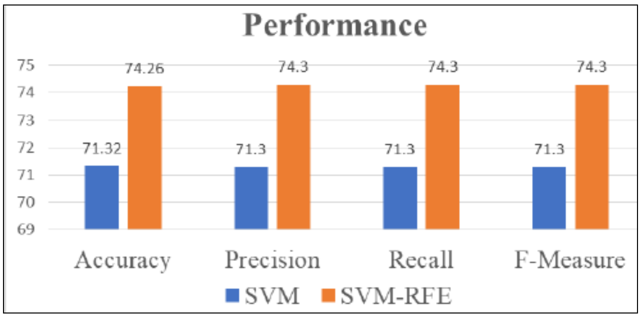


Fig 3: Performance of SVM and SVM-RFE

The Cardiac Arrhythmia dataset is either labeled as normal or one of 16 different types of arrhythmias. From 279 attribute we have filtered to 28 feature vectors by using SVM-RFE technique to get an optimum selection from complete dataset for training as well as for testing experiments. Figure-3 shows the performance of classifying Cardiac Arrhythmia by using the SVM algorithm for the full dimension data and also after the feature reduction with SVM-RFE technique.

From the figure-4, we observe the performance of SVM without SVM-RFE based on accuracy has got 71.32%, whereas the performance of SVM with SVM-RFE feature selection based on accuracy has achieved 74.26%. However, there is an improvement in the accuracy with feature selection. The accuracy rate is increased 2.94% with feature selection.

Screen shots

i. SVM Results

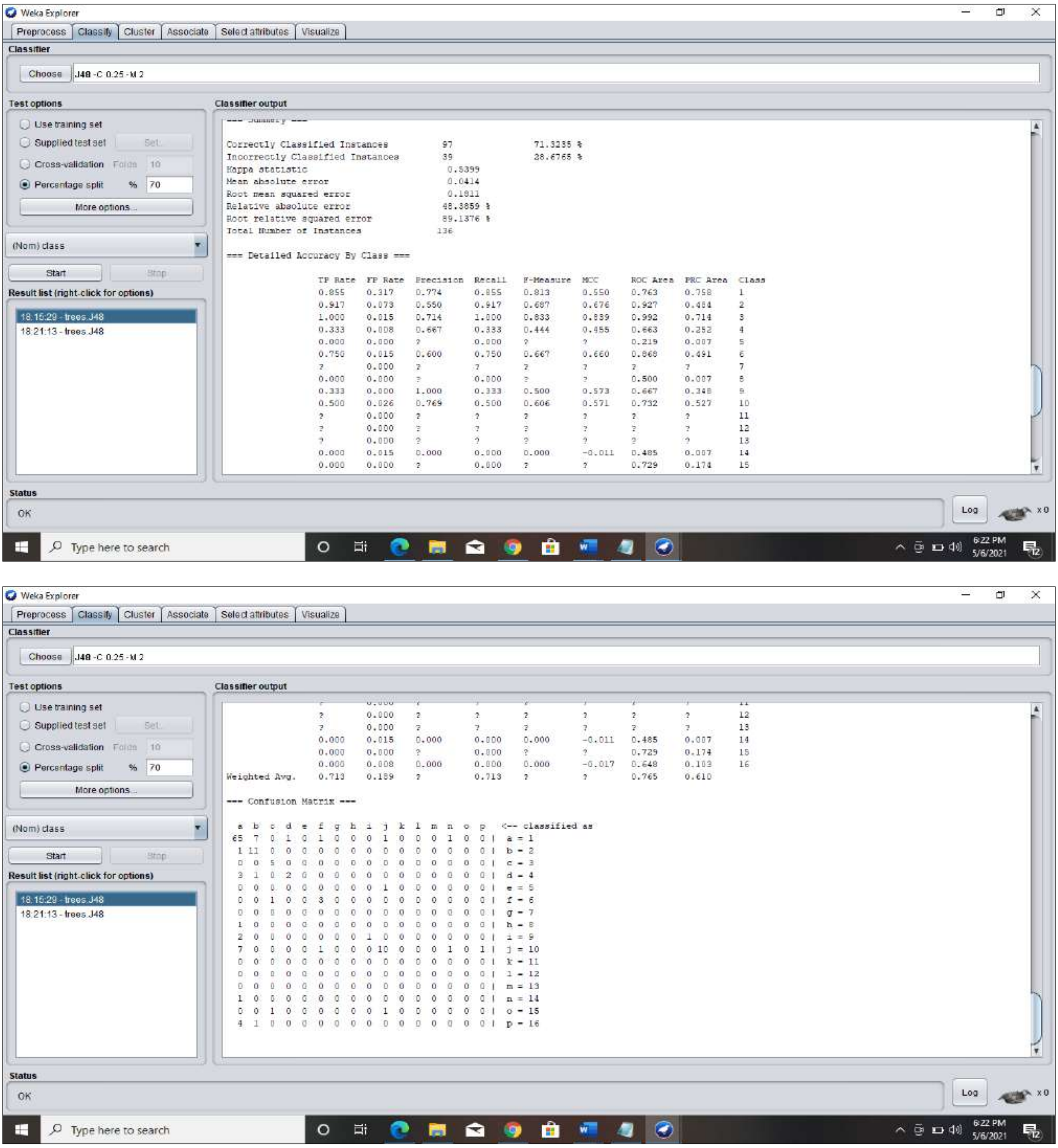
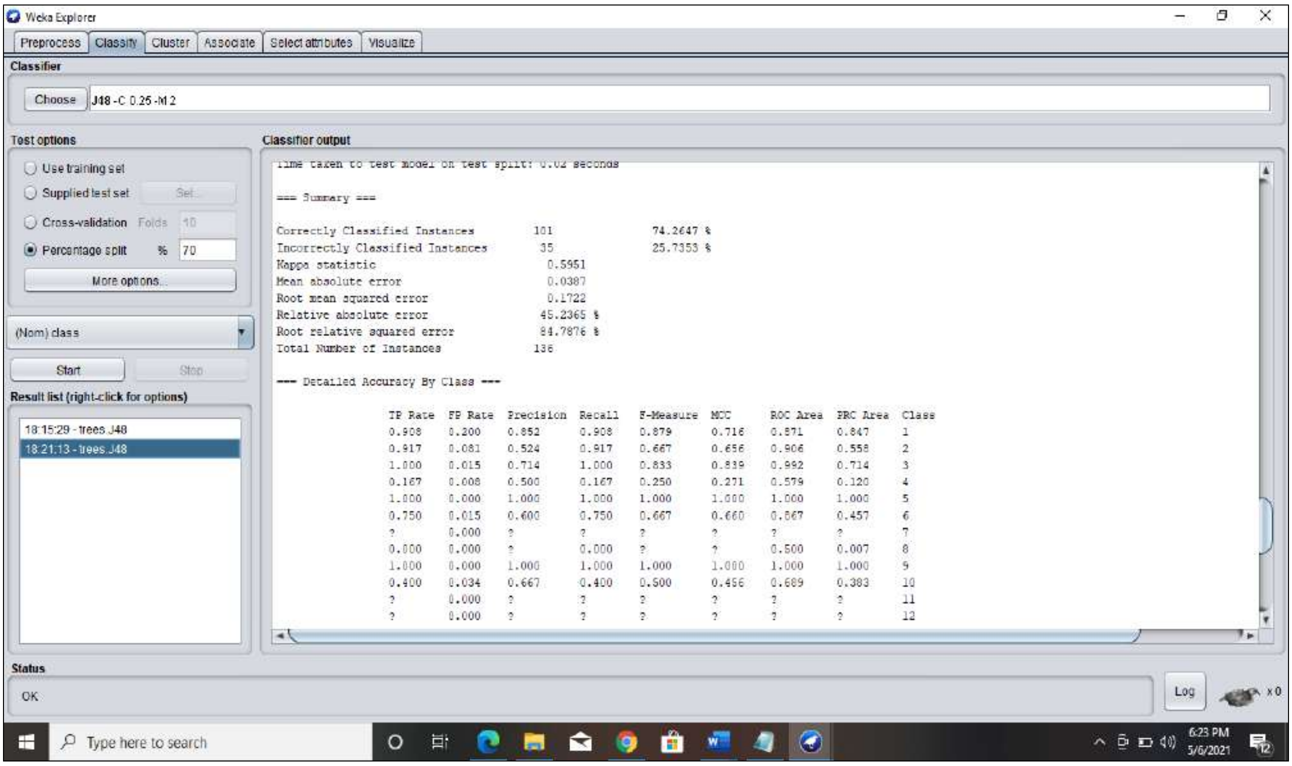


Fig 3: Results of SVM

ii. SVM-RFE Results



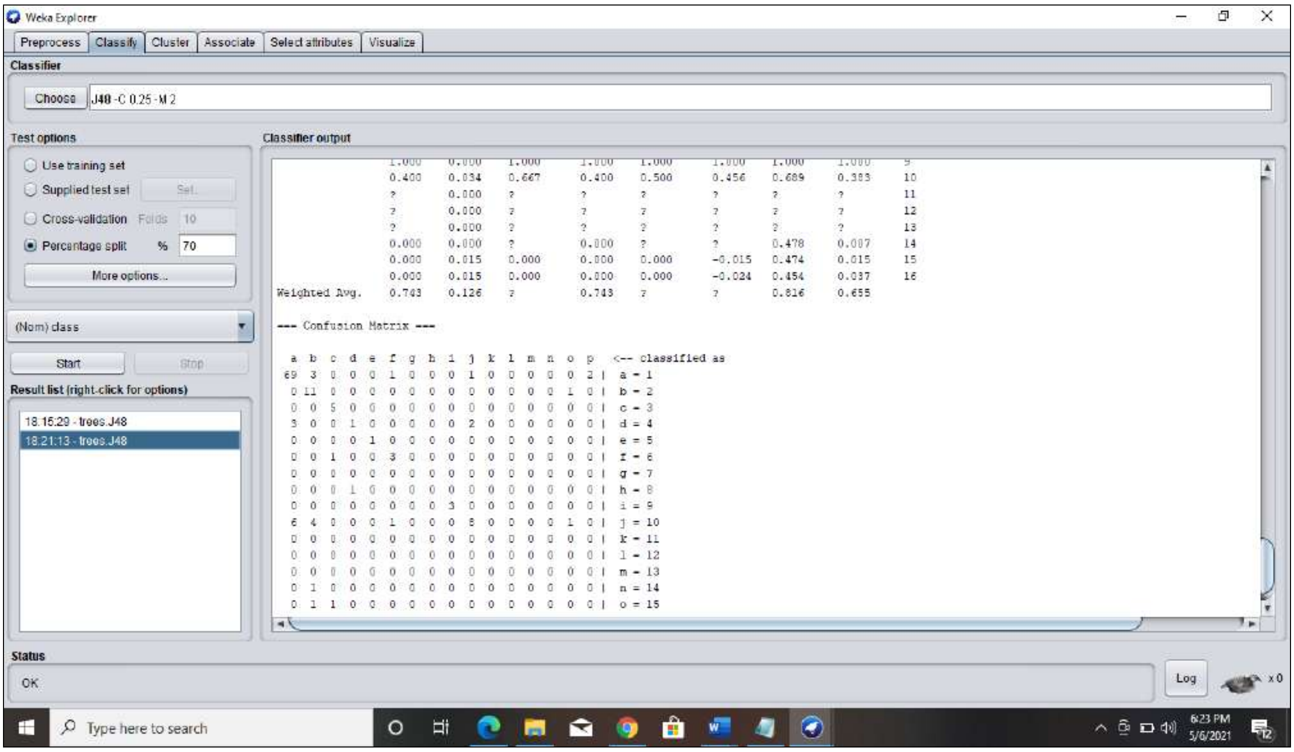


Fig 4: Results of SVM-RFE

In our experimental result the SVM with SVM-RFE algorithm shows the highest accuracy compared with SVM without SVM-RFE. With the improvement the accuracy, the proposed model demonstrated that it performs well after selecting relevant features. This result provided new insight using a classification learning algorithm and reduction technique to selection relevant and important feature in order to improve the accuracy of the system and to identify possible features which may contribute to this improvement. Most of the proposed research system could effectively utilize feature selection process to improve detection rate of their system and minimize considerably the false alarm rate.

5. Conclusion

In this paper, the significance of using a set of relevant features with the SVM classification learning algorithm for Cardiac arrhythmia disease prediction has been demonstrated. This paper proposes a method for multiclass classification of Cardiac arrhythmia using ECG records with SVM based approach. A presentation and proposition of a feature selection method which consist of a recursive feature elimination using a SVM classifier to identify important features have been done. The feature selection, preprocessing, and classification techniques have produced a combination which provides promising results for disease classification. The evaluation the effectiveness of the method using different classification metric measurement has been made and it has been proved that by reducing the number of features, the accuracy of the model was improved. In order to detect Cardiac arrhythmia disease from large dataset, detection algorithm, and feature selection method have too more efficient.

6. References

1. Elsayad M. Classification of ECG arrhythmia using learning vector quantization neural networks, in Proceedings of the 2009 International Conference on Computer Engineering and Systems, ICCES'09 2009;139-144.
2. Alickovic E, Subasi A. Medical decision support system for diagnosis of heart arrhythmia using DWT and random forests classifier, Journal of Medical Systems 2016;40(108):1-12.
3. Ravi Kumar G, Nagamani K, Anjan Babu G. A Framework of Dimensionality Reduction Utilizing PCA for Neural Network Prediction", Lecture Notes on Data Engineering and Communications Technologies, ISBN 978-981-15-0977-3. Springer Nature Singapore Pte Ltd 2020;37:173-180.
4. Ravi Kumar G, Venkata Sheshanna Kongara, Dr. Ramachandra GA. An Efficient Ensemble Based Classification Techniques for Medical Diagnosis, International Journal of Latest Technology in Engineering, Management and Applied Sciences 2013;2(8):5-9. ISSN-2278-2540.
5. Guyon Weston, Barnhill, Vapnik. Gene selection for cancer classification using support vector machines," MACHLEARN: Machine Learning 2002, 46.
6. Liu H, Yu L. Toward integrating feature selection algorithms for classification and clustering, IEEE Trans. Knowl. Data Eng 2005;17(4):491-502.
7. Liu H, Sun J, Liu L, Zhang H. Feature selection with dynamic mutual information, Pattern Recognition 2009;42(7):1330-1339.
8. Witten H, Frank E. Data mining: practical machine learning tools and techniques with Java implementations, San Francisco, CA, USA: Morgan Kaufmann Publishers Inc 2000.
9. Guyon, Elisseeff A. An introduction to variable and feature selection, J Mach. Learn. Res 2003;3:1157-1182.
10. Han J, Kamber M. Data Mining concepts and Techniques, the Morgan Kaufmann series in Data Management Systems, 2nd ed. San Mateo, CA; Morgan Kaufmann 2006.
11. Mitra M, Samanta R. Cardiac Arrhythmia Classification Using Neural Networks with Selected Features, Procedia Technology 2013;10:76-84.
12. Kohli N, Verma N. Arrhythmia classification using SVM with selected features, International Journal of Engineering, Science and Technology 2012;3(8):22-31.
13. Kumar SU, Inbarani HH. Neighborhood rough set-based ECG signal classification for diagnosis of cardiac diseases," Soft Computing 2017;21(16):4721-4733.
14. UCI Machine Learning Repository. https://archive.ics.uci.edu/ml/.
15. Vapnik VN. The nature of statistical learning theory, New York, NY, USA: Springer-Verlag New York, Inc 1995.