Webology (ISSN: 1735-188X)
Volume 19, Number 1, 2022

K-Nearest Neighbor And Fuzzy K-Nearest Neighbor Algorithm Performance Analysis For Heart Disease Classification

Malak Roman (Corresponding author)[1], Habib Ullah Nawab[2], Shehzad Ahmad[3], Awrang Zaib[4], Noor Wali khan[5], Mian Sahib Jan[6], Muhammad Anees Ur Rahman[7], Ishaq Ahmad Khan[8]

[1,4]Lecturer in Computer Science, University of Chitral, Khyber Pakhtunkhwa, Pakistan.

[2]Assistant Professor of Sociology, University of Chitral, Khyber Pakhtunkhwa, Pakistan.

[3]Lecturer in Computer Science, University of Engineering and Technology Mardan, Khyber Pakhtunkhwa, Pakistan.

[5]MS Computer Science, Quaid e Azam University Islamabad, Pakistan.

[6,7]MS Computer Science, Abasyn University Peshawar, Khyber Pakhtunkhwa, Pakistan.

[8]MS Computer Science, Preston University Swat Campus, Khyber Pakhtunkhwa, Pakistan.

Abstract

Computer software plays a crucial role in the health sector, facilitating effective management of medical records, enhancing the delivery of services, and improving patient outcomes through advanced diagnostic equipment. Cardiovascular disease poses a significant challenge for the medical community in the contemporary era, emerging as the leading cause of mortality. The healthcare business collects substantial quantities of health data that, however, cannot be effectively utilised for informed decision-making. Data mining techniques are employed to analyse large datasets in order to extract valuable information from vast amounts of data. This study aims to provide motivation for the development of an intelligent classification system for heart disease, utilising data mining techniques with a smaller set of characteristics or attributes. The K-nearest neighbour and FuzzyK-nearest neighbour classifier algorithms are employed in conjunction with evolutionary search and symmetric uncertainty attribute evaluator techniques to enhance the process of feature selection. The experimental findings demonstrate that each technique possesses distinct benefits in effectively meeting the objectives of cardiac disease detection with a high level of precision. The results collected indicate that the K-nearest neighbours (KNN) algorithm has demonstrated superior performance compared to the Fuzzy-KNN algorithm. The analysis further unveiled that K-nearest neighbours (KNN) regularly exhibited commendable accuracy while employing the symmetric uncertainty measure.

Keywords: Data Mining, K-Nearest Neighbor, Genetic Search, Fuzzy K-Nearest Neighbor, Symmetric Uncertainty, Cardiovascular Disease, Hospitals, Weka Software.

Introduction

Data mining is the integration of mining techniques used for scrutinizing immense datasets. That ascertains concealed and productive information from those records. Data mining tool permits to predict future line and trends, promote industries to craft cognizance and ambitious conclusion (Jabbar et al., 2013). Data mining is implemented in different areas such as, Telecommunication, Finance, Transport, Insurance, etc. Mining techniques has additional consequences in health-care facilitations. It smooths the progress of health organization to precisely as well as analytically pigeonhole the inefficiencies, also to reduce large amount expenditures. Large amount of data records turnout for heart disease foretell are extremely complex, to analyze by predictable methods. Data mining techniques know-how to replace these data records into positive sequences for assessment and future decision make purposes. Heart disease takes in all disorder that influences different regularity of the heart. Cardiovascular disease is a general term that expresses all heart diseases. Mostly causes of heart diseases are the insufficient supply of blood and oxygen from heart to other parts of body and vice versa. Also happens when the arteries are fully blocked or narrowed (Rajkumar & Reena, 2010). coronary artery disease, stroke, High blood pressure, cardio vascular heart disease, or rheumatic/rheumatic fever heart disease are the various forms of cardiovascular disease. In this research work we use KNN and Fuzzy KNN algorithms with feature subset selection, i.e. genetic search and symmetric uncertainty to prognosticate heart disease. In the afterwards section boast review to the concept. Proposed method was discussed in next segment and after that section describes investigational outcomes. Future moving measurement are discussed in section 5

Heart Disease

Heart disorders are the statement consisting of all diseases affecting various components of the heart. Cardiovascular disease is a general term that describes all heart diseases. Mostly causes of heart problems are the insufficient supply of blood and oxygen from heart to other parts of the body. It moreover transpires when arteries-normally endow with oxygen and blood back to the heart are blocked-up completely or befall narrowed. According to the 2010 world health organization report, key cardiovascular risk factors are (Sudhakar & Manimekalai, 2014).

a) Uses of tobacco
b) Make use of alcohol
c) Hypertension (high blood pressure)
d) Physical dormancy
e) Cholesterol level
f) Obesity and fatness
g) Use of unhealthy diet
h) High blood glucose

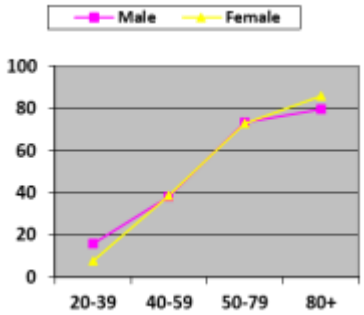

Figure 1. Prevalence of CVD in Adult by Age and Sex

Heart Diseases Classification

Heart malady infections are categorized as

i) **Coronary Disease**
Coronary problem moreover names as coronary artery syndrome. Here the arteries that's divulged oxygen and blood back to the heart are blocked-up completely or befall narrowed.

ii) **Angina Pectoris**
The Angina pectoris heart disease arises due to inadequate delivery of blood.

iii) **Congestive Heart Problem**
Congestive heart problem occurs when the cardiac muscle can't supply sufficient blood.

iv) **Cardio-Myopathy**
Cardiomyopathy problem is the weakness of heart muscles power due to insufficient heart-pumping.

v) **Congenital Disease**
This problem arises when the abnormal change occurs in the structure of the heart.

vi) **Arrhythmias**
When change occurs in the periodic movement of the heart beating.

vii) **Myocarditis**
The inflammation of heart muscle causes of fungal as well as bacterial infections that's virally affect the heart.

Literature Review

Numerous research studies have been undertaken that have concentrated on the examination of cardiovascular disease. Various data mining techniques have been employed to conduct analyses and determine probability associated with diverse practises.

Khaing (2011) presented an efficient model for the evaluation of heart attack threat from the heart disease database. For attribute evaluation K-means clustering algorithm is applied, which will extract the relevant data related to heart problems. Then determine the regular pattern effectively from the clustered data using the appropriate MAFIA (Maximum Frequent Itemset) algorithm. Finally, for the training ID3 algorithm is applied to classify the pattern that illustrates the positive risk level through decision tree method. The outcome explains that suggested model is capable to predict the heart attack effectively and more efficiently with an accuracy of 85%.

An intelligent system suggested by Dangare and Sulabha (2012), with the purpose for the assessment and analysis of heart diseases via chronological patient record files. Naive Bayes, Decision-Tree, and Neural-networks were put into practice to analyze database. WEKA tool was used to evaluate Cleveland and Stat log data sets. The UCI repository were accessed for dataset. In the projected methodology obesity and smoking were integrated as additional parameters. Later than data pre-processing process, the Neural Networks along with Decision Trees and Naive Bayes classifier algorithms were tested. In the midst of the implemented classifiers, Neural Networks presented accuracy of 0.40% , 1.30% respectively higher than the decision tree and Naive Bayes.

Jabbar et al. (2013) proposed an efficacious classification model using with a genetic approach for heart disease prediction. The data set is access from the UCI-repository databases and from an Indian hospital. The data set attributes are evaluated by Gini index and attributes having minimal Gini index were preferred for association rules. After rule evaluation, genetic functions apply to rules, and then the rules are tested over test data. The accuracy of the suggested model tested with 6 other data sets taken from an SGI-data repository and with 2 medicinal data sets obtains from the UCI-databases. The accuracy of the proposed model is 97%, much more preponderant than the other models.

Alizadehsani et al. (2013) apply multiple algorithms on the Zalizadeh Sani dataset. The algorithms used are Naïve Bayes algorithm, Artificial Neural Network, Sequential Minimal Optimization-SMO and Bagging algorithm. The rapid Miner tool was used to construct the model. Dataset having 303 patients' record and 54 features. In data preprocessing stage the effectiveness of features is evaluated through Information Gain and confidence methods. The SMO having an accuracy of 94% more preponderant than the accuracy of Bagging 93% and neural network 85%.

Beyan and Hasan (2014) suggested a FuzzyK-nearest neighbor for cancer identification amid 'Microarray Gene expression'. K-nearest neighbor and Fuzzy K-nearest neighbor classifier are functionally used for investigation by means of gene-DNA. For model testing purpose six datasets were used, that are available from geme-system.org. The Fuzzy K-NN improved with 94% than K-nearest neighbor in addition to perk up accuracy from all data sets.

Oad and Xu (2014) designed an expert system named as Fuzzy rule bases expert system. They work on Fuzzy K-nearest neighbor classifier algorithm to predict risk level of heart diseases. The dataset is obtained from the UCI machine repository database. For data preprocessing and attribute evaluation Best First search, data mining method was used. The fuzzy-KNN classifier performance is compared with J48 decision tree and Neural Network algorithms. The accuracy of fuzzy K-nearest neighbor, neural networks and J48 are 85, 78 and 73% respectively.

Sasha et al. (2015) Determines the gestures that causes muscle and other joints pain problems and disorders. The data are accumulated by using Kinect sensor. For best and optimal feature selection ReliefF algorithm is used. Then for further prediction and classification Fuzzy KNN classifier is weigh against Support Vector Machine, K-nearest neighbor, Ensemble Decision tree (EDT) and Neural Networks. Among the classifiers Fuzzy-KNN prove better performance with accuracy of 90.63%.

Souza (2015) proposed a prototype model for predicting cardiac diseases using K-Means Clustering, Neural Network and Frequent Item Set generation techniques. The Cleveland data set is taken from Cleve-land heart disease database which having 14 traits and 303 records. After removing missing value attributes 297 were left. The algorithms effectiveness is analysed through sensitivity, accuracy and specificity. The comparison shows that the ANN-artificial neural networks surpass K Means clustering in all respects, having an accuracy of 79.36 and 63.29% respectively.

Research Methodology

In this research methodology, a machine learning programming tools named WEKA-waikato environment for knowledge analysis is used for heart disease dataset evaluations. In the first movement the cardiovascular heart disease data were accumulated from the Lady Reading Hospital and Hayatabad Medical Complex hospital in Peshawar. In favour of data pre-processing, symmetric uncertainty and genetic search process were used discretely used in search of preeminent features. On the preeminent features K-nearest neighbor (KNN) classifier with Fuzzy k-nearest neighbor (F-KNN) were applied individually to test out accuracies of the model. Firstly, the models are trained and after that test data is used to calculate accuracies as shown in fig. 2. The next steps were adopted to elongate out the inquiry.

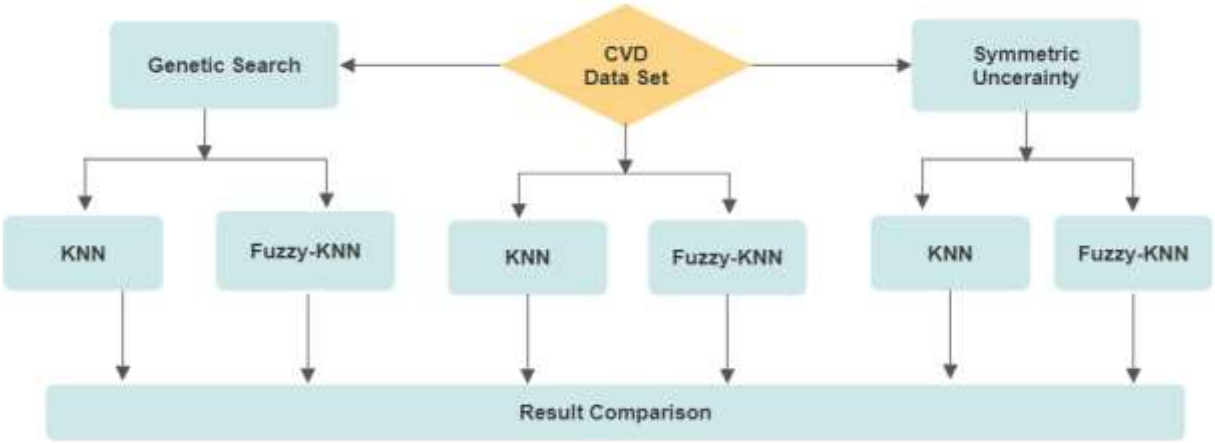

Figure 2. Flow Diagram of Research Methodology

The below fig.3 diagrammatically represents our research methodology that how data is acquired from patients and what transpired after applying feature selection methods. How the re-ports/results are evaluated after applying classifiers.

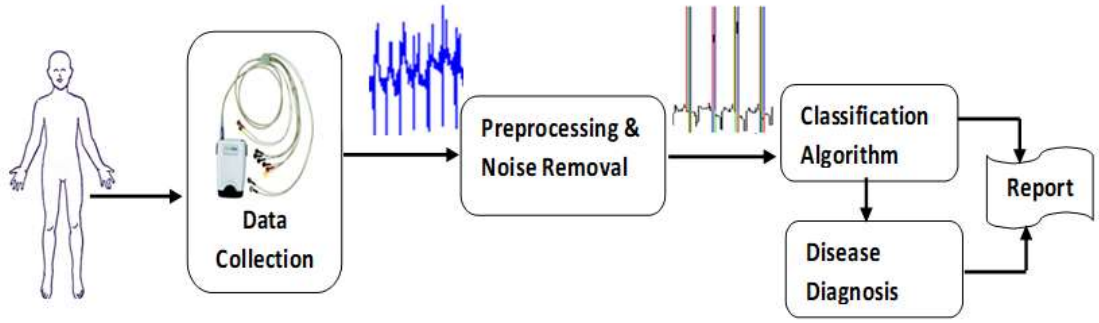

Figure 3. Architecture view of Research Methodology

Data Set

The cardiovascular heart disease dataset consists of 16 attributes. The attributes are age, gender, pain location, chest pain, BP-resting, resting heart rate, serum-cholesterol, fasting-blood sugar, diabetes, resting electrocardiography, slope, number of vessels, exercise reduce angina, family history, smoking, number of angiographic vessels. The attribute values are explicated in table 1.

Table 1.

Attribute Description					
S.No	**Attribute**	**Value**	**S.No**	**Attribute**	**Value**
1	Age	Numerical	9	Diabetes	0=No, 1=Yes
2	Gender	0=Female, 1=Male	10	Resting Electrocardiography	0=Normal, 1=ST-TWave, 2=Left Ventricular
3	Pain Location	0=Otherwise, 1= Substernal	11	Slope	0=Horizontal, 1=Upslope, 2=Downslope
4	Chest Pain Type	0=Atypical, 1=Typical	12	Exercise Reduce Angina	0=No, 1=Yes
5	Resting Blood Pressure	Numerical	13	Family History	0=No, 1=Yes
6	Resting Heart rate	Numerical	14	Smoking	0=No, 1=Yes
7	Serum Cholesterol level	Numerical	15	Angiographic Status	0=<50, 1=>50
8	Fasting Blood Sugar	0=No, 1=Yes (if FBS>120 then Yes	16	Class label	0=No, 1=Yes

Feature Selection

Feature subset evaluation is used to find out the least number of parameters. The selected attributes presented results like attained with all of attributes. It is extraordinarily helpful to eradicate unrelated attributes and increases effectiveness. By means of paramount feature the model turns into simpler, transparent and facile to interpret. Here two feature selection techniques are applied

- Genetic Search
- Symmetric Uncertainty

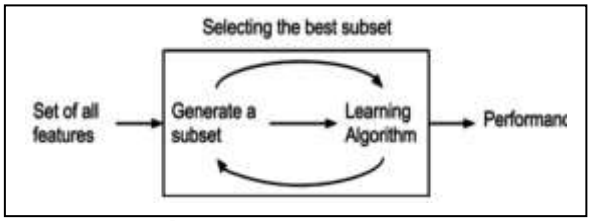

Figure 4. Feature Selection Process

Genetic Search

A genetic algorithm is a meta heuristic searching technique that is motivated by natural selection, genetics and evolution developed. An individual selection is based on genetic operators. The operator functions are Selection, Crossover, Mutation, and fitness. Individual with higher the fitness function value will have more the possibility of selections. As a result

the general purpose of Genetic algorithm is to minimize the cost of searching more favourable individual (Nopiah et al., 2010).

To create the next generation the following stages are uses.

a) Selection stage: In this stage an individual is selected for further process called parent, to incipient a new generation.
b) Crossover stage: the individual as parents are intersect to reproduce next offspring's.
c) Mutation stage: pick up the superlative offspring, that offspring is completely transmuted from the prior population.

This algorithm repetitively transmutes a population of individuals. At every one stage, the genetic algorithm arbitrarily picks an individual from the present population and presented as parents to breed after that generation. Over prospering generations, the population steps forward toward superior group of individuals.

Figure 5. Working Flowchart of Genetic Algorithm

Symmetric Uncertainty

Symmetric uncertainty knows how to analyze the potency of features. Potency is calculated between the feature and the target class. The attribute having high value of SU obtain more significance (Singh et al., 2014).

In symmetric uncertainty those attributes are removed, which having fewer SU value than defined threshold λ value and the attributes with high SU values are having higher weight. The attribute having the diminutive SU value are pruned.

Symmetric Uncertainty has a numerous advantages, i.e. it is symmetric in nature as a result SU(i,j) is equal to SU(j,i) thus it decreases the number of comparisons where i and j are two separate features. SU is not affected by multivalve attributes as is the situation of information gain their values are normalized. The equation used for symmetric uncertainty value (Ali & Shahzad, 2012).

'SU (X, Y) =2*(IG (X;Y) / (H(X) + H(Y)))'

IG(X;Y) = Mutual Information

Information gain a correlation measurement, is based on entropy with information theoretical concept.

$H(x) = - \sum \{ p(x)*\log2 (P(x)) \}$

The decrease in uncertainty of X to Y, namely IG-information gain, is define as

$IG(X/Y) = H(X) - H(X|Y)$ (4.5)

Where gain = H (Y) – H (Y | X)

= H(X) – H (X|Y)

= H(Y) + H (X) – H (X, Y)

Information gain is symmetrically measures the total of information gained for X after evaluating with Y, is resembles to the total of information gained for Y to X.

Classifiers

The two classifiers k-nearest neighbors and fuzzy k-nearest neighbor are used for heart disease classification. The classifiers KNN and FKNN were applied separately on with and without best features selected by genetic search and symmetric uncertainty techniques.

K-Nearest Neighbor

The k-nearest neighbor algorithm is proven as lazy learning classifier algorithm. The KNN classifier evaluation is on an association among the test and training data sets. The evaluation is carrying on similarity relationship basis. The sample called test data is evaluated via measure up within trained data set. Once when an unfamiliar data record is provided, the KNN classifier looks for the similar data record in pre-trained data set. K-nearest classifier selects the closest similar record and assigns that record class label to the tested record (Han & Micheline, 2005). For an unknown test record a majority vote of its nearby neighbours are fined. The class-label of that class is assigned from the trained data which is closest to the test record. The closeness is evaluated by various distance formula functions. If K having value=1, then case is undoubtedly assigned to its nearest similar neighbor.

The familiarity is determined in terms of various distances metric Cartesian, such as Euclidean, Manhattan and Murkowski distance formulas. The trendiest formula used to find out the similarities between two points (sample and training) is Euclidean distance formula. Assumes the first second instances points are (a_1, a_2, - - - a_n) and (b_1, b_2, - - - b_n). Then distance between both of the instances with Euclidean distance formula is (Han et al., 2011).

Euclidean distance : $\sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \ldots\ldots (a_n - b_n)^2}$

Knn Algorithm Pseudocode:

1. Compute "d(x, x_i)" i =1, 2, ….., **n**; where **d** stand for the Euclidean distance between the points.
2. Sort n Euclidean distances, calculated in non-descending order.
3. Let **m** be a +ive integer, obtain the first **m** distances from the sort out list.
4. Locate those **m**-points consequent to these m-distances.
5. Let m_i symbolize the number of dots belongs near the i^{th} class amid **m** points i.e. $m \geq 0$
6. If $m_i > m_j \; \forall \; i \neq j$ then locate x in class i.

Fuzzy K-Nearest Neighbor

In Fuzzy K-Nearest Neighbor classification techniques, multiple class membership's value is assigned to test sample instead of placing it in any single class / group. In fuzzy-KNN algorithm an element can fit into more than a single class. Fuzzy KNN classification method is an alternative of other two value traditional methods as KNN.

The Fuzzy K-nearest neighbor classifier does not classify a value into completely true and false groups. The Fuzzy K-nearest neighbor classifier is same to the simple KNN classifier. In KNN classifier, a sample belongs to only one of the nearest neighbor class. Where in fuzzy-KNN classification techniques, a sample belongs to various classes having distinct membership value linked to these classes. The membership value represents that an element is in which category class.

For example, if a data member is given 0.9 class memberships and memberships value of 0.05 in second class, it is quite safe to say that the vector fit in 0.9 class membership. Same as, if a value is determined with 0.55 membership to 1st class, with 0.44 membership to 2nd class and 0.01 membership to 3rd class then might we should be confused to relate and assign vector class. Though, we are sure that class three is not where it belongs. In such condition, The vector displays a high degree of membership in both classes one and two, suggesting further analysis to determine its categorization. Hence, The membership allocations resulting from the approach may have significance during the categorization process.

The membership value is found out by using the following function (Han et al., 2011).

$$u_i(x) = \frac{\sum_{j=1}^{K} u_{ij}\left(1/\|x - x_j\|^{2/(m-1)}\right)}{\sum_{j=1}^{K}\left(1/\|x - x_j\|^{2/(m-1)}\right)}$$

i=1, 2... c

Where Ui(x) is the membership function value of the test sample instance x to class i. || x-x j|| is the Euclidean distance between sample x and x j whereas Ui(Xj) is the fuzzy membership value of the j-th neighbor to the i-th class. m is a fuzzy strength membership parameter value between 1 and 2 determines that how the distance is measured and variable k shows the number of nearest neighbors.

Result and Evaluations

This part provides a comprehensive analysis of the experimental results obtained in the course of this research, together with a detailed discussion of these findings in relation to the research objectives. The empirical results illustrate that individual classifiers provide distinct merits for effectively detecting cardiovascular ailments. The performance of two classification algorithms is evaluated using two distinct feature selection strategies. The effectiveness of a classification algorithm can be determined by evaluating its accuracy, recall, f-measure, and precision values, particularly when using the best attribute assessment technique. This evaluation is conducted on both the training and test data sets. This part provides a comprehensive analysis of the experimental results obtained in the course of this research, together with a detailed discussion of these findings in relation to the research objectives. The empirical results illustrate that individual classifiers provide distinct merits for effectively detecting cardiovascular ailments. The performance of two classification algorithms is evaluated using two distinct feature selection strategies. The effectiveness of a classification algorithm can be determined by evaluating its accuracy, recall, f-measure, and precision values, particularly when using the best attribute assessment technique. This evaluation is conducted on both the training and test data sets.

Table 2.

Comparison of Fuzzy K-Nearest Neighbor and K-Nearest Neighbor						
Algorithms	**Preprocessing**	**Accuracy**	**Precision**	**Recall**	**F-Measure**	**ROC Area**
	Simple	65%	0.65	0.667	0.657	0.571

Fuzzy K-Nearest Neighbor	Genetic Search	74%	0.793	0.742	0.74	0.65
	Symmetric Uncertainty	75%	0.768	0.75	0.757	0.7
K-Nearest Neighbor	Simple	90 %	0.933	0.933	0.933	0.911
	Genetic Search	93%	0.97	0.967	0.967	0.942
	Symmetric Uncertainty	95%	0.992	0.992	0.992	0.968

Table 2 Shows the summarize results of fuzzy-k nearest neighbor and k-nearest neighbor algorithm classifiers with and without attribute evaluation techniques. The classifiers are compared on the basis of accuracy, precision, recall and ROC curves.

- The accuracy of KNN model is analyzed by varying the value of k from 1 to 5.
- During testing the KNN with SU, GS and simply gives best results with k value 5.
- The accuracy of FKNN model is analyzed by varying the value of k from 1 to 10 and fuzzifier value from 3.0 to 1.5.
- During testing the FKNN with SU gives best result with k value 1 and fuzzifier 3.0, the FKNN with GS shows best result with k value 5 and fuzzifier 2.0 and FKNN (simply) gives best results with k value 9 and fuzzifier 3.0.

Through-out the precision, Recall and F-Measure, accuracy is calculated with True (T) and False (F) parameter values. Precision, Recall and F-measure, accuracy are measures by use of the following equations

- Precision-P = TP / (TP+FP)
- Recall-R = TP / (TP+FN)
- F-measure = 2 * (P*R) / (P+R)

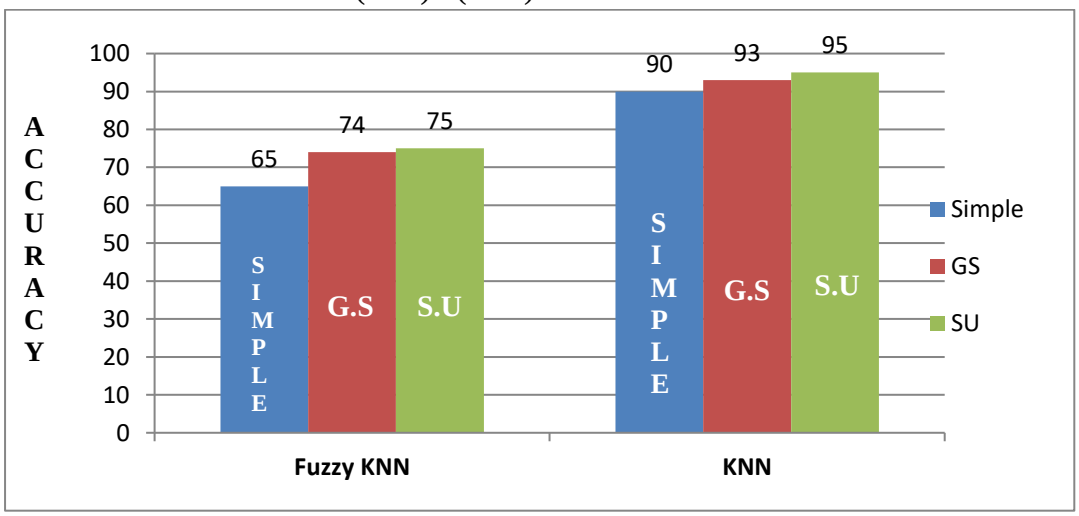

Figure 6. Comparing Results of KNN and F-KNN in terms of Accuracy.

Fig. 6 graphically represents the result of k-nearest neighbor and fuzzy k-nearest neighbor algorithms in terms of accuracy, which shows that k-nearest neighbor out performs then fuzzy k-nearest neighbor algorithm with and without symmetric uncertainty and genetic search attribute evaluators. The accuracy of k-nearest neighbor with symmetric uncertainty, genetic

search and without preprocessing methods is 95, 93 and 90% respectively. The fuzzy k-nearest neighbor shows the accuracy of 75, 74 and 63% respectively.

From the experiment and results, it was found that the k-nearest neighbor classifier algorithm is much better than fuzzy k-nearest neighbor algorithm in terms of accuracy. The accuracy of k-nearest neighbor with symmetric uncertainty is 95%. The accuracy of k-nearest neighbor with genetic search is 93% and the accuracy of k-nearest neighbor (simple) is 90%. The accuracy of fuzzy k-nearest neighbor with symmetric uncertainty is 75%. The accuracy of fuzzy k-nearest neighbor with genetic search is 74% and the accuracy of fuzzy k-nearest neighbor (simple) is 65%.

To the extent that the calculation relates, k-nearest neighbor is more effective for learning classification then the fuzzy k-nearest neighbor.

The fig.7 graphically represents the result of k-nearest neighbor and fuzzy k-nearest neighbor algorithms in terms of accuracy, Precision, recall and f-measure. Which shows that k-nearest neighbor performed better then fuzzy k-nearest neighbor algorithm with and without symmetric uncertainty and genetic search attribute evaluators.

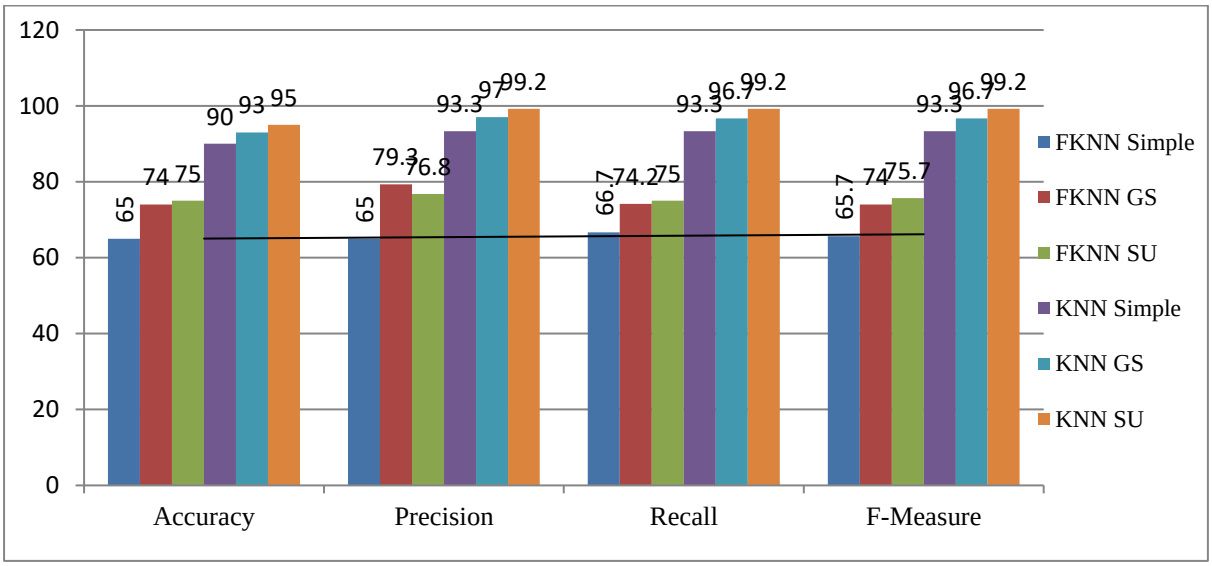

Figure 7. Graphically Representation of Accuracy, Precision, Recall and F-Measure of KNN and F-KNN with GS and SU

Discussion

Based on the obtained results, it was determined that the k-nearest neighbor classification algorithm outperforms the fuzzy k-nearest neighbor approach. The utilisation of the k-nearest neighbor algorithm with symmetric uncertainty attribute assessment demonstrates superior performance in terms of accuracy, precision, and recall. The k-nearest neighbour classifier algorithm is characterised by its simplicity; however, it exhibits superior performance compared to other sophisticated classification algorithms. The findings shown in Table 1 indicate that the k-nearest neighbor classifier exhibits superior speed and accuracy compared to the fuzzy k-nearest neighbor approach. In terms of the calculation's relevance, it can be argued that k-nearest neighbor (k-NN) has greater efficacy in classification learning compared to fuzzy k-nearest neighbor (fuzzy k-NN).

Conclusion and Future Work

Conclusion

This research aims to analyse various classifiers in order to determine the optimal classifier for the categorization of cardiovascular heart disease. According to existing literature, the classification data mining methodology has been widely regarded as the most effective method for classifying cardiovascular diseases. This study presents a methodology for the categorization of cardiovascular disease (CVD) data through the use of k-nearest neighbour and fuzzy k-nearest neighbour classifier algorithms. The collected results demonstrate that the proposed model, which incorporates the k-nearest neighbour algorithm with a symmetric uncertainty pretreatment technique, outperformed the existing models. The results obtained from the proposed model are compared to those of existing models in terms of accuracy. The present study employed a highly efficient methodology for extracting valuable insights from cardiovascular disease (CVD) data.

Future Work

Based on the observed outcomes of our study and the proposed model, it is recommended that this model be extended for the purpose of analysing and classifying many other forms of data, including stroke data, wind data, and text mining, among others. The recommended model can be further improved by the development of an automated hybrid software solution, which would facilitate user-friendly utilisation. While the suggested model demonstrates efficient performance and achieves improved accuracy compared to previously implemented models, it still requires further modifications. In subsequent iterations, the integration of clustering techniques with the suggested model could potentially enhance processing efficiency and bolster accuracy.

Acknowledgments

We express our gratitude to Almighty Allah for providing us with assistance and guidance throughout our lives, enabling us to possess the necessary capabilities and perseverance to undertake this scientific endeavour. Furthermore, we express our gratitude to our friends and families for their invaluable assistance.

References

Ali, S. I., & Shahzad, W. (2012, October). A feature subset selection method based on symmetric uncertainty and ant colony optimization. In Emerging Technologies (ICET), 2012 International Conference on (pp. 1-6). IEEE.

Alizadehsani, R., Habibi, J., Hosseini, M.J., Mashayekhi, H., Boghrati, R., Ghandeharioun, A., Bahadorian, B. & Sani, Z.A., (2013). A data mining approach for diagnosis of coronary artery disease. Computer methods and programs in biomedicine, 111(1), pp.52-61.

Beyan, C., & Ogul, H. (2014). A fuzzy K-NN Approach for Cancer Diagnosis with Mocroarray Gene Expression Data. Departement of Computer Engineering, Baskent University.

D'Souza, A. (2015). Heart Disease Prediction Using Data Mining Techniques. International Journal of Research in Engineering and Science (IJRES) ISSN (Online), 2320-9364.

Dangare, C.S. & Apte, S.S., (2012). Improved study of heart disease prediction system using data mining classification techniques. International Journal of Computer Applications, 47(10), pp.44-48.

Deekshatulu, B.L. & Chandra, P., (2013). Classification of heart disease using k-nearest neighbor and genetic algorithm. Procedia Technology, 10, pp.85-94.

Han, J., Pei, J., & Kamber, M. (2011). Data mining: concepts and techniques. Elsevier.

Han, Jiawei & Micheline K. (2005). Data mining concepts and techniques. Morgan Kaufmann.

Jabbar, M.A., Deekshatulu, B.L. & Chandra, P., (2013). Heart disease prediction system using associative classification and genetic algorithm. arXiv preprint arXiv:1303.5919.

Khaing, H.W. (2011, March). Data mining based fragmentation and prediction of medical data. In Computer Research and Development (ICCRD), 2011 3rd International Conference on (Vol. 2, pp. 480-485). IEEE.

Nopiah, Z. M., Khairir, M. I., Abdullah, S., Baharin, M. N., & Arifin, A. (2010, February). Time complexity analysis of the genetic algorithm clustering method. In Proceedings of the 9th WSEAS International Conference on Signal Processing, Robotics and Automation, ISPRA (pp. 171-176).

Oad, K. K., DeZhi, X., & Butt, P. K. (2014). A fuzzy rule based approach to predict risk level of heart disease. Global Journal of Computer Science and Technology, 14(3-C), 17.

Rajkumar, A. & Reena, G.S., (2010). Diagnosis of heart disease using datamining algorithm. Global journal of computer science and technology, 10(10), pp.38-43.

Saha, S., Pal, M., Konar, A., & Bhattacharya, D. (2015). Automatic Gesture Recognition for Health Care Using ReliefF and Fuzzy kNN. In Information Systems Design and Intelligent Applications (pp. 709-717). Springer India.

Singh, B., Kushwaha, N., & Vyas, O. P. (2014). A Feature Subset Selection Technique for High Dimensional Data Using Symmetric Uncertainty. Journal of Data Analysis and Information Processing, 2(04), 95.

Sudhakar, K. & Manimekalai, D.M., (2014). Study of heart disease prediction using data mining. International Journal of Advanced Research in Computer Science and Software Engineering, 4(1).