IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727
PP 51-57
www.iosrjournals.org

Onto Fuzzy Based Support Vector Machine

[1]J.Karunya, [2]N.vijayalakshmi
[1]*Research scholor, Computer science, Kovai kalaimagal college of arts and science, Narasipuram, Coimbatore.*
[2]*Associate professor, head of the department, Computer science, Kovai kalaimagal college of arts and science, Narasipuram, Coimbatore*.

Abstract: *Classification consists of predicting a certain outcome based on a given input. This research work proposes fuzzy ontology based support vector machine classifier. In order to predict the outcome, the algorithm processes a training set containing a set of attributes and the respective outcome, usually called goal or prediction attribute. The algorithm tries to discover relationships between the attributes that would make it possible to predict the outcome. The proposed research work focuses on classification problem among the medical dataset to predict heart diseases.*

I. Data Mining

Data mining is the act of consequently looking for huge stores of data to find examples and patterns that go past basic investigation. Data mining utilizes complex scientific calculations to section the data and assess the likelihood of future occasions. Data mining is otherwise called Knowledge Discovery in Data (KDD). The key properties of data mining are:

- Automatic disclosure of examples
- Prediction of likely results
- Creation of significant data
- Focus on substantial data sets and databases

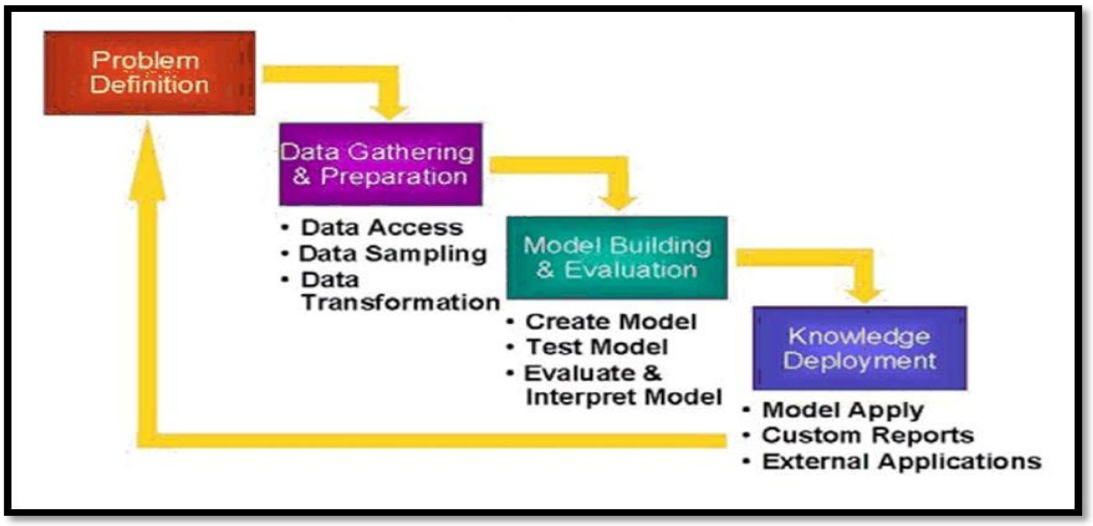

Fig.1.1 Data Mining Process Model

Fig.1.1 outlines the stages, and the iterative nature, of a data mining project. The procedure stream demonstrates the data mining project does not stop when a specific arrangement is conveyed. The after effects of data mining trigger new business questions, which thus can be utilized to grow more engaged models. This underlying period of a data mining project concentrates on understanding the project targets and requirements. A data mining project begins with the comprehension of the business issue. Data mining specialists, business specialists, and space specialists work confidently together to characterize the project targets and the fundamentals from a business point of view. The project objective is then converted into a data mining issue definition.

II. Fuzzy Logic

In the writing sources, we can discover various types of legitimization for fuzzy frameworks hypothesis. Human learning these days turns out to be progressively critical – we pick up it from encountering the world inside which we live and utilize our capacity to motivation to make arrangements in the mass of data (i.e., to figure human information in a precise way). Since we are altogether constrained in our capacity to see the world and to significant thinking, we get ourselves wherever faced by vulnerability which is a consequence of the absence of data (lexical impression, deficiency), specifically, error of estimations. The other restricting component in our want for exactness is a characteristic dialect utilized for portraying/sharing information, correspondence, and so on. We comprehend center implications of the word and can impart precisely to a worthy degree, however, for the most part, we can't accurately concur among ourselves on the single word or terms of presence of mind meaning. To put it plainly, common dialects are unclear.

III. Ontology

Ontologism assumes a key part in the coming of the Semantic Web. An imperative issue when managing ontologism is the alteration of a current philosophy in light of a specific requirement for change. This issue is a confusing and complex one since it can take a few unique structures and incorporates a few related sub issues, similar to heterogeneity determination or monitoring metaphysics variants. Accordingly, it is being tended to by a few unique, however firmly related and regularly covering research disciplines. Shockingly, the limits of each such teacher are not clear, as a similar term is regularly utilized with various implications in the important writing, making a specific measure of perplexity.

IV. Objectives Of This Research Work

As talked about in the past segment the proposed look into work concentrates on grouping issue among the restorative data set. The proposed investigate work has the accompanying goals concerning the picked look into the issue.

- ✓ To propose a novel fuzzy cosmology procedure that fulfills the info parameters to be picked so as to perform characterization errand.
- ✓ To fuse bolster vector machine with the novel fuzzy philosophy procedure.
- ✓ To perform data mining arrangement undertaking utilizing the proposed fuzzy cosmology based help vector machine calculation with the Z-Alizadeh Sani dataset.
- ✓ To lead broad re-enactments utilizing MATLAB apparatus for assessing the execution of the proposed calculation as far as exactness, affect ability, and specificity.

V. Literature Review

Shivnarayan Patidar et al.,2015 presented a new method for the finding of computer aided design using tunable-Q wavelet transform (TQWT) based highlights extricated from heart rate signals. The heart rate signals are decomposed into different sub-bands using TQWT for better diagnostic component extraction. The nonlinear component called centered correntropy (CC) was computed on decomposed detail sub-band. Mohamad M. Al Rahhal et al.,2015 proposed a novel interactive ensemble learning approach in light of the extreme learning machine (ELM) classifier and the induced ordered weighted averaging (IOWA) administrators. Nikola Kasabov et al.,2014 presented a novel method and system for personalized (individualized) modeling of spatio/spectro-temporal data (SSTD) and prediction of events. A novel evolving spiking neural network reservoir system (eSNNr) was proposed for the reason. Fernando Jimenez et al.,2014 presents a novel govern based fuzzy classification methodology for survival/mortality calculation in seriously well-done patients.

M.H. Vafaie et al.,2014 presented classification method to characterize ECG signals more accurately in view of the dynamical model of the ECG signal. Jennifer N. Cooper et al.,2015 aimed to compare the performance of LR to several data mining algorithms for predicting 30-day surgical morbidity in children. The prediction of hospitalization was formulated as a directed classification problem in Wuyang Dai et al.,2015. M. Akhil Jabbar et al.,2013 proposed an algorithm which joins KNN with genetic algorithm for powerful characterization.

VI. Onto Fuzzy Based Support Vector Machine (Ofsvm)

6.1 Fuzzy SVM

Fuzzy Support Vector Machine was presented by Lin and Wang. A participation s_j is assigned for each info sample (x_j, y_j), where $0< s_j <1$. Since the participation sj s_j is the disposition of the comparing point xj toward one class, and the parameter ξ_j is a measure of a blunder in the SVM, the term $s_j\xi_j$ is a measure of mistake with various weighting. The ideal hyperplane issue is then viewed as the answer for:

$$\text{minimize } \frac{1}{2} w.w + C \sum_{i=1}^{k} S_j \xi_j$$

subject to

$$y_j(w.z_j + b) \geq 1 - \xi_j = j \quad 1, K, n$$

To utilize FSVM, a participation work should be characterized for each info sample. Here, it utilized the participation work definition. From Eq. (3.23) one can see that if the ξ_i of a misclassified information x_i is expanded, the recently learned hyperplane will tend to accurately classify x_i keeping in mind the end goal to dispose of the bigger blunder that x_i acquainted with the classifier. Correspondingly assigning a bigger enrollment s_i for an information expands the likelihood of accurately classifying that sample while a little participation diminishes the likelihood of effectively classifying the sample. In view of this perception, the enrollment work is characterized as takes after.

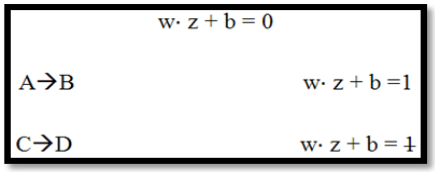

Fig. 3.3: Different regions in high dimension space.

- Initial, a customary SVM is prepared using the first preparing set.
- After stride 1, the hyperplane w. z + b = 0 is found. Expecting that if w. z + b > 0, the information is assigned to the positive class; generally, the information is assigned to the negative class. There additionally are two different hyperplanes w· z + b =1and w· z + b = −1. As showed in Figure, the high dimensional space is isolated into four areas by these three hyperplanes. For the positive samples, area A speaks to the information focuses that are accurately classified and the related ξ s is 0. District B speaks to the information focuses that are likewise accurately classified yet the related ξ s are non-zero. Locale C and D speaks to the information focuses that are mistakenly classified. For the negative samples, similar locales with the same properties concerning the positive samples can be gotten by simply swapping the district positions. In the accompanying discussion, just consider positive cases are considered. Applying these same standards to negative cases is direct.
- The focuses in District A have no commitment to the improvement since their ξ s is 0. Along these lines, regardless of what participation is assigned to them, it won't influence the resultant hyperplane. Here for simplicity, a consistent esteem sA = s1 is assigned to them where 0 < s1 < 1.
- The focuses in district B are accurately classified, yet they have non-zero ξ s. In this way, they add to the enhancement condition, however, ought to be dealt with as less vital than the focuses in areas C and D, since they are effectively classified. The more close to the hyperplane w. z + b > 0, the more critical in the following preparing strategy to accomplish a superior classification result. Given d=w· z + b, where z=Φ(x) for input point x, the participation for local B is characterized as:
  $S_B = S_1 + (1 - d) \times S_2$
  where $S_2 > 0$, $0 < s1 + s2 < 1$ and $0 \leq d \leq 1$ in locale B.

- The focuses in local C are inaccurately classified. It can be anticipated that in the following preparing method, the hyperplane can move towards these focuses, accordingly permitting a greater amount of them can be classified effectively. The closer the focuses to the hyperplane w. z + b > 0, the less essential they are in the following preparing system. As clarified in step 4, in any case, they are more vital than the focuses in area B. Using the same documentation as step 4, the fluffy enrollment for area C is characterized as
  $S_c = (S_1 + S_2) + |d| \times S_3$
  where $s_3 > 0$, $0 < s_1 + s_2 + s_3 < 1$, and $-1 \leq d \leq 0$ in area C.

- The focuses in area D are inaccurately classified. The further away the focuses are from the hyperplane w. z + b > 0, the all the more most likely an exception exists; in this manner, the littler enrollment ought to be assigned. The enrollment for locale D is characterized as:
  $S_D = (S_1 + S_2 + S_3)/|d|^k$

✓ where k>0 and d ≤ −1 in locale D. Here, k is a positive whole number, and the bigger k is, the speedier the participation diminishes with the expansion of separation d. The estimation of k is picked as 9.

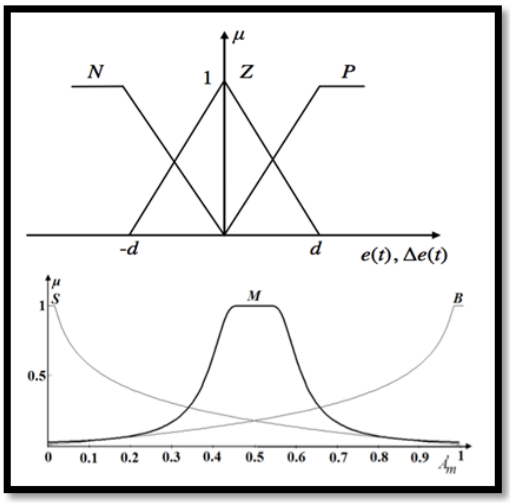

Fig. 4.4: Characterized Enrollment Capacity

The parameters s1, s2, s3 are customizable. They have introduced under the limitation that every one of them is positive and the total of them is littler than 1. An FSVM is prepared using the resultant participation work from those underlying parameters. At that point, the same preparing information is utilized as the test information on the prepared FSVM and those three parameters are changed in accordance with accomplishing the most elevated classification comes about. At long last, the prepared FSVM is utilized to classify the genuine test information.

6.2 Onto Fuzzy based Support Vector Machine (OFSVM)

The final stage toward our fuzzy ontology based support vector machine method is fuzzy taxonomy generation based on the subsumption relations among the extracted concepts. Let $Spec(c_x, c_y)$ denote that concept c_x is a specialization (subclass) of another concept c_y. The degree of such a specialization relation can be estimated from

$$\mu R_{CC} \approx Spec(c_x, c_y) = \frac{\sum_{t \in c_x \cap c_y} \mu_{c_x}(t) \otimes \mu_{c_y}(t)}{\sum_{t_x \in c_x} \mu_{c_x}(t_x)}$$

where $\otimes$ is a fuzzy conjunction operator which is equivalent to the main function. The above formula states that the degree of subsumption (specificity) of c_x to c_y is based on the ratio of the sum of the minimal membership values of the common terms belonging to both concepts to the sum of the membership values of terms in the concept c_x. For instance, if every attribute of $c_y$is also an attribute of c_x, a strong specificity relation exists and the value of $Spec(c_x, c_y)$is high. The range of the $Spec(c_x, c_y)$ relation falls in the unit interval [0, 1] and the subsumption relation is asymmetric.

One problem of the standard fuzzy conjunction operation is that the specificity value is highly influenced by the weakest terms (attributes) of the concepts. Therefore, we explore another alternative of estimating the degree of subsumption between two concepts based on the method successfully applied to image analysis. In particular, any two concepts c_x and $c_y$could be said to be similar if their structural similarity is high and the corresponding structural similarity value $SSIM(c_x, c_y)$ approaches. On the other hand, two concepts are dissimilar if their structural similarity value $SSIM(c_x, c_y)$ is low (e.g., close to zero).

The terms $Q_1 = 0.0255, Q_2 = 0.2295, and\ Q_3 = 0.1148$ are constants, and they are applied to image analysis work before [Wang et al.,2004]. It adopt $Q_1 = [0{,}5 \times 0.0255]$, $Q_2 = [0{,}5 \times 0.2295]$, $Q_3 = [0{,}5 \times 0.1148]$. When it apply the structural similarity measure to estimate the degree of subsumption between two concepts. For instance, if most attributes t_i belonging to the concept c_y are also belonging to the concept c_x, the concept c_x is a subconcept of c_y to a high degree. To formulate our $Spec(c_x, c_y)$ function

based on the structural similarity, it first compute the common concept $c_g = c_x \cap c_y$. Then, it is necessary to examine if this common subconcept is more subsumed by which concept to determine the direction of the specialization relation.

VII. About Matrix Laboratory

Dr. Cleve Moler, the Boss researcher at MathWorks, Inc., initially composed Matlab, to give simple access to framework programming created in the LINPACK and EISPACK ventures. The principal version was composed in the late 1970s for use in courses in grid hypothesis, direct polynomial math, and numerical investigation. Matlab is in this manner based on an establishment of refined lattice programming, in which the fundamental information component is a grid that does not require pre-dimensioning.

4.3 The Medical Dataset

The Z-Alizadeh Sani informational collection (Roohallah Alizadehsania et al.,2013) contains the records of 303 patients, each of which has 54 highlights. All highlights can be considered as markers of computer aided design for a patient, as per medical writing [Bonow et al.,2012]. Be that as it may, some of them have never been utilized as a part of information mining based methodologies for computer aided design determination. The highlights are orchestrated in four gatherings: demographic, symptom and examination, ECG, and research center and resound highlights. Table 4.1 presents the highlights of Z-Alizadeh Sani informational index alongside their legitimate reaches, separately. Every patient could be in two conceivable classes computer aided design or Normal. A patient is sorted as computer aided design, if his/her diameter narrowing is more noteworthy than or equivalent to half, and generally as Normal [Bonow et al.,2012]. Some of the highlights in the displayed tables ought to be additionally clarified: HTN distinguishes the historical backdrop of hypertension, DM is a background marked by Diabetes Mellitus, Ebb and flow Smoker is momentum consumption of cigarettes, Ex-Smoker is a past filled with past consumption of cigarettes, and FH is a past filled with coronary illness in first-degree relatives.

VIII. Results And Discussions

The following are the performance metrics used to evaluate the performance proposed classification method namely Fuzzy Ontology based Support Vector Machine (FOSVM). The classification accuracy, sensitivity and specificity can be calculated using the following metrics.

Classification Accuracy	It measures proportion of true results (both true positives and true negatives) among the total number of records examined. Thus, Classification Accuracy = (TN+TP)/(TN+TP+FN+FP)
Sensitivity	It measures the ability of a test to detect the condition when the condition is present. Thus, Sensitivity = TP/(TP+FN).
Specificity	It measures the ability of a test to correctly exclude the condition (not detect the condition) when the condition is absent. Thus, Specificity = TN/(TN+FP).

8.1 Classification Accuracy Analysis

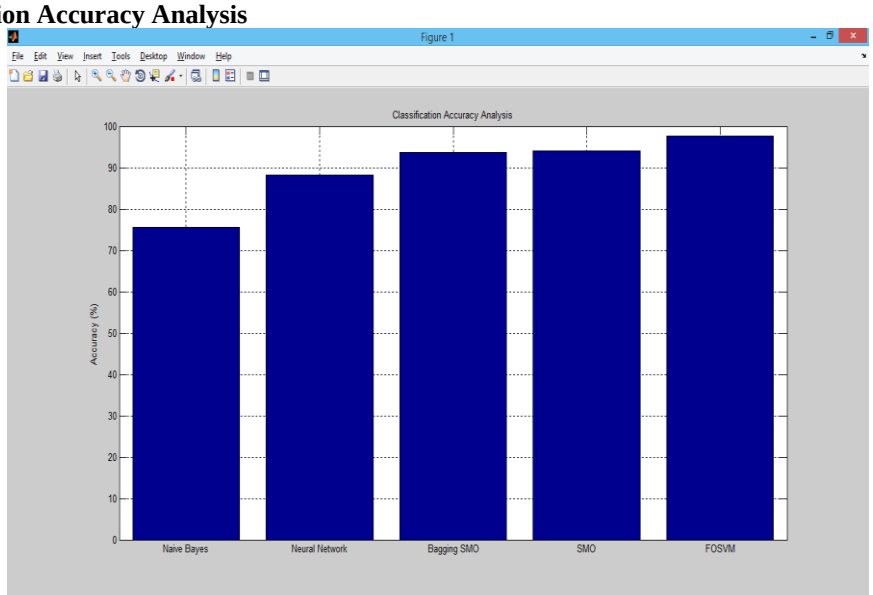

Fig 8.1 Classification Accuracy Analysis

The Fig.8.1. depicts the classification accuracy performance for Z-Alizadeh Sani dataset of the existing algorithms (Roohallah Alizadehsania et al.,2013) such as Naive Bayes, Neural Network, Bagging SMO, SMO and the proposed algorithm in chapter 3, i.e., Fuzzy Ontology based Support Vector Machine (FOSVM). It can be clearly understood that the proposed work FOSVM provides better classification accuracy 97.69% respectively.

8.2 Sensitivity Analysis

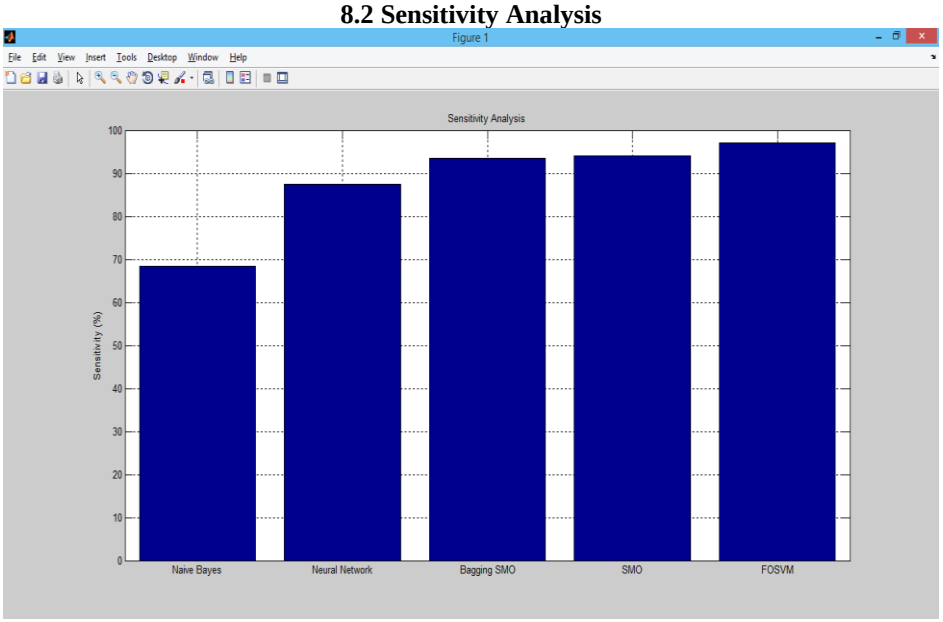

Fig 8.2 Sensitivity Analysis

The Fig.8.2. depicts the classification accuracy performance for Z-Alizadeh Sani dataset of the existing algorithms (Roohallah Alizadehsania et al.,2013) such as Naive Bayes, Neural Network, Bagging SMO, SMO and the proposed algorithm in chapter 3, i.e., Fuzzy Ontology based Support Vector Machine (FOSVM). It can be clearly understood that the proposed work FOSVM provides better sensitivity 97.11% respectively.

8.3 Specificity Analysis

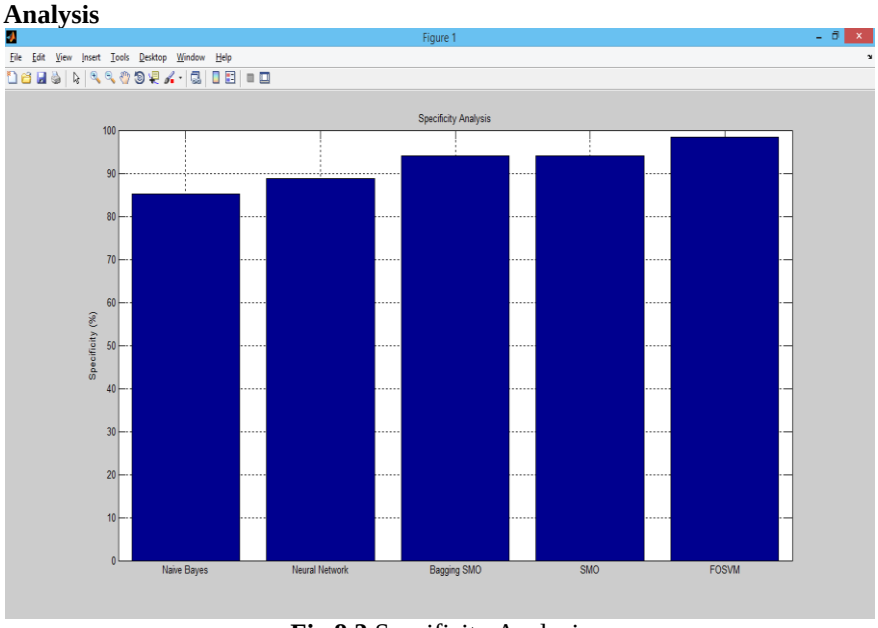

Fig 8.3 Specificity Analysis

The Fig.8.3. depicts the classification accuracy performance for Z-Alizadeh Sani dataset of the existing algorithms (Roohallah Alizadehsania et al.,2013) such as Naive Bayes, Neural Network, Bagging SMO, SMO

and the proposed algorithm in chapter 3, i.e., Fuzzy Ontology based Support Vector Machine (FOSVM). It can be clearly understood that the proposed work FOSVM provides better specificity 98.46% respectively.

IX. Conclusion

Data mining offers promising approaches to reveal concealed examples inside a lot of data. These shrouded examples can conceivably be utilized to anticipate future conduct. The accessibility of new data mining calculations, be that as it may, ought to be met with the alert. Most importantly, these strategies are just comparable to the data that has been gathered. Great data is the primary prerequisite for good data investigation. Accepting great data is accessible, the subsequent stage is to pick the most proper procedure to mine the data. Nonetheless, there are tradeoffs to consider while picking the suitable data mining method to be utilized as a part of a specific application. There are distinct contrasts in the sorts of issues that are conducive to every procedure. The "best" model is regularly found by experimentation: attempting distinctive advances and calculations. Intermittently, the data examiner should think about or even join accessible systems keeping in mind the end goal to acquire ideal outcomes. This exploration work proposed a novel fluffy philosophy based help vector machine. Reproductions are done utilizing MATLAB apparatus and the proposed FOSVM calculation accomplishes better execution regarding arrangement precision, affect ability, and specificity.

This exploration work can additionally be stretched out by consolidating the accompanying strategies and procedures.

- ✓ Optimization strategies like a hereditary calculation, molecule swarm improvement, ant colony optimization and so on can be additionally included.
- ✓ By fusing type – 2 fuzzy sets and frameworks the proposed classifier can be supplanted with customary fluffy rationale framework.

References

[1]. Fernando Jimenez, Gracia Sanchez, Jose M. Juarez, Multi-objective evolutionary algorithms for fuzzy classification in survival prediction, Artificial Intelligence in Medicine, Volume 60, Issue 3, March 2014, Pages 197-219.

[2]. Jennifer N. Cooper, Lai Wei, Soledad A. Fernandez, Peter C. Minneci, Katherine J. Deans, Pre-operative prediction of surgical morbidity in children: Comparison of five statistical models, Computers in Biology and Medicine, Volume 57, 1 February 2015, Pages 54-65

[3]. M. Akhil jabbar, B.L. Deekshatulu, Priti Chandra, Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm, Procedia Technology, Volume 10, 2013, Pages 85-94

[4]. M.H. Vafaie, M. Ataei, H.R. Koofigar, Heart diseases prediction based on ECG signals' classification using a genetic-fuzzy system and dynamical model of ECG signals, Biomedical Signal Processing and Control, Volume 14, November 2014, Pages 291-296.

[5]. Mohamad M. Al Rahhal, Yakoub Bazi, Naif Alajlan, Salim Malek, Haikel Al-Hichri, Farid Melgani, Mansour A. Al Zuair, Classification of AAMI heartbeat classes with an interactive ELM ensemble learning approach, Biomedical Signal Processing and Control, Volume 19, May 2015, Pages 56-67

[6]. Shivnarayan Patidar, Ram Bilas Pachori, U. Rajendra Acharya, Automated diagnosis of coronary artery disease using tunable-Q wavelet transform applied on heart rate signals, Knowledge-Based Systems, Volume 82, July 2015, Pages 1-10,

[7]. Wuyang Dai, Theodora S. Brisimi, William G. Adams, Theofanie Mela, Venkatesh Saligrama, Ioannis Ch. Paschalidis, Prediction of hospitalization due to heart diseases by supervised learning methods, International Journal of Medical Informatics, Volume 84, Issue 3, March 2015, Pages 189-197