FEATURE SUBSET SELECTION AND CLASSIFICATION USING HYBRID IMPROVED SVM

P.Arumugam[1],
Department of Statistics, Manonmaniam Sundaranar University,
Tirunelveli, 628012, Tamilnadu, India
sixfacemsu@gmail.com

P.Jose[2]
Research scholar, Department of Statistics, Manonmaniam Sundaranar University,
Tirunelveli, 628012, Tamilnadu, India
josekiruba@gmail.com

ABSTRACT - Many feature subset selection algorithms have been proposed, but not all of them are appropriate for a given feature selection problem. At the same time, so far there is rarely a good way to choose appropriate feature subset selection algorithms for the problem at hand. Feature selection has become an essential element in the Data Mining process. In this paper, investigate the problem of efficient feature selection for classification on High Dimensional datasets. present a feature reduction method for overcome a loss of accuracy of classification after that perform a classification process with the aid of modified fuzzy c-means clustering with rough set theory it is used to perform a feature selection process. Once the feature reduction is formed, the classification will be done based on the Hybrid kernel based Improved SVM (ISVM) classifier. In this classification, the optimal kernel is identified using Grey Wolf Optimization (GWO).

Keywords:-Feature Sub Set Selection, Data Mining, High Dimensional Data, Feature Reduction, Fuzzy C-Means, Rough Set Theory

1. INTRODUCTION

Feature selection plays an important role in classification, since it can shorten the learning time, simplify the learning classifiers, and improve the classification performance [1]. There may be complex interaction among features; it is generally difficult to find the best feature subset. Feature selection is a process of selecting a subset of optimal features according to certain criteria. It is an important and frequently used technique for dimension reduction [2]. It reduces dimensionality by removing irrelevant, redundant, or noisy features, thus bringing about significant effects for applications: reducing the cost of data acquisition and collection, speeding up a learning algorithm, improving learning accuracy, and resulting in better model comprehensibility. Feature selection techniques are generally categorized as filters, wrappers, and embedded methods [3]. A filter approach relies primarily on general characteristics of a data set to evaluate and select feature subsets without considering any machine learning approach [4]. A wrapper approach employs a classification algorithm to evaluate feature subsets, and adopts a strategy to seek for optimal subsets [6, 7]. Since the wrapper approach considers a classifier within the search process, this approach gets often better result than the filter one [5]. Hover, this approach still suffers with variety of issues such as local convergence. A meta-heuristic is a high-level problem-independent algorithmic framework that provides a set of strategies to develop heuristic algorithm [8].Feature extraction and selection are important steps in data detection and classification. An optimum feature set should have effective and discriminating features, while mostly reduce the redundancy of feature space to avoid “curse of dimensionality” problem [9]. Feature selection strategies often are applied to explore the effect of irrelevant features on the performance of classifier systems [10-11]. In this phase, an optimal subset of features which are necessary and sufficient for solving a problem is selected [12]. Feature selection improves the accuracy of algorithms by reducing the dimensionality and removing irrelevant features [13] [14].Conventional Principal Component Analysis (PCA) is one of the most frequently applied feature extraction techniques. It is based on extorting the axes on which data illustrates the maximum changeability [15].More over Cluster analysis is a commonly used data mining technique to explore the relationships among attributes, samples and the relationships beten attributes and samples. Clustering algorithms assign samples or attributes to clusters based on their similarity. Cluster analysis can be used as a preliminary method for classification or for finding new classes. Hierarchical clustering tree (HCT) [16] and k-means [17] are the two most popular clustering methods used to extract the features from the data. Hover, the rough sets offer an effective approach of managing uncertainties and can be employed for tasks such as data dependency analysis, feature identification, dimensionality reduction, and pattern classification. Rough set theory [18, 19] is a fairly recent intelligent technique for managing uncertainty that is used for the discovery of data dependencies, to evaluate the importance of attributes, to discover patterns in data, and to reduce redundancies. While wrappers and embedded methods require a frequent classifier interaction in their flow, filters do not need any classifier interaction during the construction of the feature set [20, 28].Additionally, the Adaptive Genetic Fuzzy System for medical data classification has been brought in by B. Dennis, S. Muthukrishnan*et al.* [21]. A Genetic Fuzzy System (GFS) was essentially a fuzzy system supplemented by a learning process based on a genetic algorithm (GA). In different application domains, Fuzzy systems re revealed their capability to work out diverse kinds of problems. At present, there was a rising interest to supplement fuzzy systems with learning and revision capabilities. Two of the most victorious approaches to hybridize fuzzy systems with learning and adaptation methods re prepared in the area of soft computing. The GA was combined with Fuzzy system for dissimilar purposes like rule selection, membership function optimization, rule generation, co-efficient optimization, for data categorization. For medical data classification process they proposed an Adaptive Genetic Fuzzy System (AGFS) for optimizing rules and membership functions. The main purpose of their research was 1) Generating rules from data as ll as for the optimized rules selection, acclimatizing of genetic algorithm was prepared and to describe the investigation problem in genetic algorithm, beginning of novel operator, called systematic addition was prepared, 2) Developing an uncomplicated technique for scheming of membership function and Discretization, and 3) Proposing a fitness function by permitting the frequency of incidence of the rules in the training data. At last, to launch the competence of the classifier the presentation of the expected genetic-fuzzy classifier was assessed with quantitative, qualitative and relative analysis.Yanchun Zhang, *et al.* [25] gave an overview of late advances in biomedical image examination and classification from new imaging modalities, for example, terahertz (THz) pulse imaging (TPI) and dynamic contrast-enhanced magnetic resonance images (DCE-MRIs) and recognized their underlining shared characteristics. Both time and recurrence space signal pre-processing procedures re considered: sound evacuation, spectral investigation, and principal component analysis (PCA) and wavelet changes.

2. PROPOSED METHODOLOGY

The primary intension of this research is to achieve promising results in data classification. have planned to utilize novel feature selection method and Support Vector Machine classifier. Initially, the pre-processing will be applied to extract useful data and to convert suitable sample from raw datasets. Here, input dataset will be as high dimensional or high features which are a great barrier for classification. Therefore, feature dimension reduction method will be applied to reduce the features' space without losing the accuracy of classification. Here, clustering with rough set theory based feature selection will be developed and used to reduce the feature dimension. In MFCM- RST (Modified Fuzzy c-means (MFCM) clustering algorithm with Rough Set Theory) will be combined and used to feature selection process. Once the feature reduction is formed, the classification will be done based on the Hybrid kernel based Improved SVM (ISVM) classifier. In this classification, the optimal kernel is identified using Grey Wolf Optimization (GWO). The overall process of the proposed framework is divided into three steps, such as 1) Pre-processing, 2) Feature selection and 3) Classification using Hybrid kernel based ISVM in fig.1. Here, for experimentation, the dataset given in the UCI machine learning repository will be subjected to analyze the performance of the proposed technique utilizing accuracy, sensitivity and specificity and other performance metrics. The implementation will be done by MATLAB.

2.1. Pre-processing

The pre-processing will be applied to extract useful data and to convert suitable sample from raw datasets

2.2. Feature Selection Using FCM-RS Algorithm

Here, input dataset will be as high dimensional or high features which are a great barrier for classification. Therefore, feature dimension reduction method will be applied to reduce the features' space without losing the accuracy of classification. Here, clustering with rough set theory based feature selection will be developed and used to reduce the feature dimension.

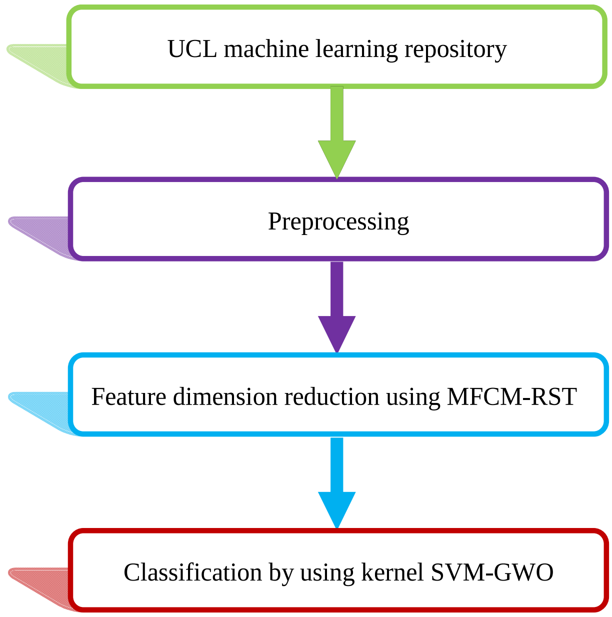

Fig 1: Proposed Feature Subset Selection and Classification

In RS-MFCM (Modified Fuzzy C-Means (MFCM) clustering algorithm with rough set theory) will be combined and used to feature selection process.The basic idea of this section is to reduce the dimension of the features using fuzzy c-means clustering with roughest theory (FCM-RS) algorithm**.** The high number of features is a great obstacle for the prediction; so have to reduce the dimension of the features. As a result, feature dimension reduction method is required to decrease the features' space without losing the precision of prediction. In addition, reduce the number of features and remove the not related, unnecessary or noisy information. Besides, this improves the presentation of information prediction with speeding up the processing algorithm. To improve the prediction accuracy, use the FCM-RS algorithm for feature reduction. In FCM-RS algorithm, Fuzzy c-means clustering (FCM) is combined with Rough Set Theory (RST) to use the feature selection process. The detailed explanation of dimension reduction is explained in the following steps.

Step1: Subset formation

The FCM algorithm assigns cluster centre to each category by using fuzzy memberships

$$J_m = \sum_{i=1}^{I}\sum_{j=1}^{J}(\mu_{ij})^m \left\| x_i - z_j \right\|^2 \qquad (1)$$

In Eqn. (1), x_i represents the features $t(w(s_n^i))$, $c(s_n^i)$ extracted from the input database, z_j is the j thcluster centre and m is the constant value. The membership function represents the probability that a cluster center belongs to a specific cluster. In the FCM algorithm, the probability is dependent on the distance beten the pixel and each individual cluster center in the feature domain. The membership functions and cluster centers are updated by the equations (2) and (4).

$$u_{ij} = \frac{1}{\sum_{k=1}^{J}\left(\frac{\left\| x_i - z_j \right\|}{\left\| x_i - z_k \right\|}\right)^{\frac{2}{m-1}}} \qquad (2)$$

Repeat the algorithm until the coefficients' change beten two iterations is no more than ξ, for the given sensitivity threshold.

$$\max_{ij}\left\| U_{ij}^{(k)} - U_{ij}^{(k+1)} \right\| < \xi \qquad (3)$$

In equation (3), ξ is a termination criterion beten 0 and 1, whereas k are the iteration steps. The clusters centroid values are computed by using the equation (4).

$$z_j = \frac{\sum_{i=1}^{I} u_{ij}^m . x_i}{\sum_{i=1}^{I} u_{ij}^m} \quad (4)$$

To enhance the performance of the fuzzy-C-means clustering method, adaptiveness is invoked by measuring the Clustering effectiveness (α) and Absolute density (β). On the basis of these two, set two thresholds to ensure the clustering being good. After the FCM process, obtain the number of cluster set such as $I_1, I_2, I_3, ..., I_n$. This clustered dataset is used for the further processing.

Step 2: Attribute selection

After the clustering process using FCM, have to do the attribute selection process. Here, calculate the minimum $A_{\min}$ and maximum $A_{\max}$ value of each column (or each attribute) in the dataset D_1 and D_2 . If the $A_{\min}$ and $A_{\max}$ value is similar to corresponding dataset I_1 and I_2 , have to neglect that column.

Step 3: Discretization

Discretization is a significant step in data processing to convert the data into specific interval, means that the range of values is confined into a specific interval. Here, have used one discretization function based on the predictable way. perform the discretization process at first, identify the maximum and minimum values of every attribute, and the K interval is tracked by taking the ratio beten the deviated value and the K value.

$$D^L = \min(A_j) \leq \left[\min(A_j) + Dev^j \right] \quad (6)$$

$$D^M = \left[\min(A_j) + Dev^j \right] \leq \left[\min(A_j) + 2 \bullet Dev^j \right]$$

$$D^H = \left[\min(A_j) + 2 \bullet Dev^j \right] \leq \max(A_j) j$$

Using equation (6) can adjust all the feature values in the specific interval. Now obtain the new feature values, which feature values varies from specific interval .Then; every value that comes under within the range is replaced with the interval value so that the input data is transformed to the discretized data. Consequently, the training dataset D^{TR} is concerted to the discretized format D^D where, the entire data element D^D contains only the L, M, and H $if\ k = 3$.

Step 4: Reduct and Core analysis

Reduct and core study is a significant step of rough set theory categorize the significance of the attributes so that the relevant attributes can be able to recognize without compromising the accuracy of classification. Now the dimensionality reduction can simply be considered as removal of some attributes from the decision table (actually some features from the feature vector) preserving its basic classification capability. In this case would face two conditions;

i) Whether a decision table contains some redundant or superfluous data, if yes then
ii) Collect those redundant data and remove them.

To perform such reduction of attributes in rough set theory select indispensable attributes. In roughest theory indispensable attributes are selected using two fundamental concepts: reducts and core: they are defined as follows:

Let Q be a subset of P and let a belong to Q

- A is dispensable in Q if I(Q-{a}); otherwise a is indispensable in Q
- Set Q is independent if all its attributes are indispensible
- Subset Q' of Q is a reduct of Q if Q' is independent and $I(Q') = I(Q)$.

Thus a reduct is the minimal subset of attributes that enables the same classification of elements of the universe as the whole set of attributes i.e., attributes that do not belong to a reduct are redundant or superfluous.

$$Core = \bigcap reduct \quad (7)$$

Reduct is computed by taking the relative discernability function and eliminating unnecessary attributes. The equivalence class is signifying in table 5, here the equivalence class is denoted as 'E'. Once the feature reduction is formed, the classification will be done based on the Hybrid kernel based Improved SVM (ISVM) classifier.

3.Classification Using Hybrid Kernel based-Improved Support Vector Machine

Once the feature reduction is formed, the classification will be done based on the Hybrid kernel based Improved SVM (ISVM) classifier. In this classification, the optimal kernel is identified using Grey Wolf Optimization (GWO).

3.1. Kernel Based Support Vector Machine:

In these mobile commerce system have predicted a user next purchase behaviour with the aid of these kernel based SVM. Afterward, the finest attributes are delivered to fusion kernel support vector machine for the principle of categorization. Now, the chosen attribute from the previously progression is efficiently engaged for the isolation of the two module. For the principle of processing the non-linear procedure, the kernel functions are started in the SVM categorization. There are two very important phases in the SVM procedure such as the preparation phase and the effortless stage.

*3.2 Training phase***:** Currently, the output of attribute choice is provided as the input of the preparation stage. The input utility supplies the group of values which cannot be alienated. Approximately each and every one of the probable isolation of the position places are comprehend by a hectic plane. In the Lagrange pattern, it is probable to put the partition of the hectic plane standard vector during the divergent kernel task. In this association, a kernel symbolizes a few tasks, which communicate to a dot product for definite kind of attribute recording. Yet, recording a position into a better quality dimensional gap is probable to direct to unnecessary assessment period and enormous storage requirements. By the outcome, in concrete perform, an original kernel task is initiated which h is competent of openly estimating the dot product in the better-quality dimensional gap. The frequent edition of the kernel task is provided as follows.

$$K(U,V) = \varphi(U)^T \varphi(V) \qquad (8)$$

In this view, the majority broadly engaged kernel tasks contain the linear kernel, Polynomial kernel, Quadratic kernel, Sigmoid and the Radial Basis task. Specified beneath are the terms for the different kernel task.

For Linear Kernel: $linear_k(U,V) = u^T v + c$ (9)

Where u, v represents the inner products in linear kernel and c is a constant.

For Quadratic Kernel: $quad_k(U,V) = 1 - \frac{\|u-v\|^2}{\|u-v\|^2 + c}$ (10)

Where, u, v - are the vectors of the polynomial kernel function in the input space

For Polynomial Kernel: $poly_k(U,V) = (\lambda u^T v + c)^e, \lambda > 0$ (11)

For Sigmoid Kernel: $sig_k(U,V) = \tanh(\lambda u^T v + c), \lambda > 0$ (12)

The effectiveness of the SVM consistently oriented on the variety of the kernel. In the occurrence of the attribute gap being linearly indivisible, it has to be recorded into a better-quality dimensional gap by the Radial basis task kernel, in order that the concern appears as linearly detachable. Additionally, the amalgamation of any two kern el task is proficient to defer outstanding accuracy than that acquired by utilizing some single kernel task. In the original procedure, an original KSVM is predicted, dedicated for the noteworthy development in the categorization system. At this point, two kernel tasks such as the linear and the quadratic kernel task are mutual to defer outstanding presentation ratios. The uniting Equations 6 and 7 the standard is predictable as recommended in the original technique. The mutual kernel task is successfully engaged in the KSVM and the standard of the kernel task, $avg_k(U,V)$ is delivered beneath.

$$avg_k(U,V) = \frac{1}{2}\left(lin_k(U,V) + quad_k(U,V)\right) \qquad (13)$$

$$avg_k(U,V) = \frac{1}{2}\left(\left(u^T v + c\right) + \left(1 - \frac{\|u-v\|^2}{\|u-v\|^2 + c}\right)\right) \qquad (14)$$

In the kernel Support Vector Machine, two kernels such as the linear and quadratic are used into description for the principle of classify the search links. The merging of two outcomes, the standard of the outcome is accomplished and developed to classification.

3.3 Testing phase<u>**:**</u> In the training phases productivity from the classification choice is provided as to the experiment stage and the productivity specifies the subsistence or else the absence.

4.Gray Wolf Optimization

In our past work have used a GSO-AGA calculation for streamlining. In this dim wolf streamlining calculation is utilized for prioritization it will be the last procedure there are all the more no of clients in distributed computing is expanding in an exponential rate at consistently. Grey wolf optimizer agent (GWO) is another method, which can be connected effectively to solve advanced issues. The GWO in fact mimics the initiative progression and chasing component of dim wolves.

The dim wolves adequately encase a Canidae's segment predecessors and are esteemed as the head predators seeing their course of action at the sustenance's nourishment grouping. They typically delineate a prejudice to detail appropriate as a get together The pioneer speak to a male and a female, set apart as alpha, which are for the larger part division in allegation of captivating legitimate assortment screening differing highlights, for instance, the chasing, resting area, time to wake, et cetera. The determinations arranged by the alpha are acknowledged on to the gathering. The Beta locations to the second grade in the pecking exhibit of the dark wolves. They are, in a general sense, optional wolves which successfully proposed a couple of sponsorship to the alpha in the assortment developing or proportional gathering utility.

The omega is the littlest division of the dark wolf pack and immense undertaking as a substitution present into the further premier wolves nearly on each event and is permitted to incorporate only the small remains charming after a great banquet by the pioneer wolves. A wolf is demonstrated as optional or as delta every so as often as possible in the event that it doesn't fit in among the gathering of an alpha, beta, or omega. Because of reality that these delta wolves require to reverence the alphas and betas, they have a prospering high extent over the omegas. In our technique, the alpha (α) is esteemed as the most suitable accumulation by a standpoint to recreating judiciously the group pecking arrange of wolves though imagining the GWO. In this way, the second and the third most great arrangement are beta (β) and delta (δ) freely. The left over cheerful arrangement are seen to be the omega (ω). In the GWO strategy the chasing (improvement) is directed by then α , β , δ and ω .

The step by step procedure of gray wolf optimization algorithm is provided as follow,

Initialization process

Here start the pre-processed output data as ll $a, A, and\ C$ as coefficient vectors

Fitness evaluation

Assess the fitness utility based on the evaluation valuen and furthermore choose the best result

Separate the solution based on the fitness

Presently, get the divided result based on the fitness value. Let the initial finest fitness results be d_α , the second finest fitness results d_β and the third finest fitness results d_δ .

4.1 Encircling prey

The tracking is aimed at by α , β , δ and ω tag the length of these three contenders. In order for the compilation to track an injured party is leading nearby it.

$$d(t+1) = d(t) + \vec{A}.\vec{K} \qquad (15)$$

$$\vec{K} = |\vec{C}.d(t+1) - d(t)| \qquad (16)$$

$$\vec{A} = 2\vec{a}r_1 - \vec{a} \text{ And } \vec{C} = 2r_2 \quad (17)$$

Where, t depicts the iteration number.

d (t) corresponds to the prey position.

A and C depicts the coefficient vector.

$\vec{a}$ is linearly lessen from 2 to 0.

r_1 *and* r_2 corresponds to the random vector [0, 1].

4.2 Hunting

assume that the alpha (best competitor arrangement), beta and delta incorporate the upgraded data about the plausible position of the casualty with a specific end goal to mirror precisely the following exercises of the dim wolves. Since a result, amass the soonest three finest result fulfilled up to now and require the further investigate middle person (counting the omegas) to adjust their posture in view of the course of action of the finest investigate arbiter. For replication, the novel result is $d(t+1)$ unsurprising by the formulae expressed underneath.

$$\vec{K}^{\alpha} = |\vec{C}_1.d_{\alpha} - d|, \quad \vec{K}^{\beta} = |\vec{C}_2.d_{\beta} - d|, \quad \vec{K}^{\delta} = |\vec{C}_3.d_{\delta} - d| \tag{18}$$

$$d_1 = d_{\alpha} - \vec{A}_1.(\vec{K}^{\alpha}),\ d_2 = d_{\beta} - \vec{A}_2.(\vec{K}^{\beta}),\ d_3 = d_{\delta} - \vec{A}_3.(\vec{K}^{\delta}) \tag{19}$$

$$d(t+1) = \frac{d_1 + d_2 + d_3}{3} \tag{20}$$

It can be trial that the closing area would be in an easygoing position encompassed by a circle which is particular by the area of alpha, beta, and delta in the investigate hole. By means alpha, beta, and delta figure the area of the casualty, and further wolves' modernize their area unpredictably in the district of the casualty.

4.3 Attacking prey (exploitation) and Search for prey (exploration)

Examination and usage are guaranteed by the versatile estimations of a and A. The versatile estimations of confinement n and A grant GWO to proficiently transformation among examination and use. By declining A, half of the cycles are consistent to examination $|A| \geq 1$ and the further half are focused on use (|A| < 1). The GWO contains just two premier restrictions to be acclimated (a and C). Hover, include saved the GWO calculation as easy as achievable through the littlest number of agent to be acclimated the technique will be tenacious anticipating the best precision is procured. Finally the finest characteristic is chosen and supply to the extra system.

5. RESULTS AND DISCUSSION

The proposed fuzzy based classifier is experimented with the three dataset namely Cleveland, Hungarian, Switzerland. These datasets are taken from the UCI machine learning repository. This data base encloses 76 characteristics; on the other hand all allocated tests refer to exploiting a subset of 14 of them. Particularly, ML researchers employ only the Cleveland database still today. The "goal" field refers to the presence of heart disease in the patient. It is integer valued from 0 (no presence) to 4. Experiments with the Cleveland database have contemplated on simply attempting to differentiate presence (values 1, 2, 3, 4) from absence (value 0).

5.1. Evaluation Metrics

Here will explain some evaluation metrics used for our work such as sensitivity, specificity, accuracy, FPR, FNR.The performance of the proposed feature subset selection and classification method was evaluated by the three metrics Sensitivity, Specificity and Accuracy, FPR, FNR. The results of proposed work help to analyze the efficiency of the process. The subsequent table I tabulates the results.

Table I: Evaluation measures for Cleveland dataset

K-cross validation for Cleveland	Accuracy	Sensitivity	Specificity	FPR	FNR
1	0.8	0.789	0.818	0.307	0.181
2	0.9	0.9375	0.857	0.076	0.142
3	0.9	0.882	0.923	0.142	0.076
4	0.833	0.928	0.75	0.076	0.25
5	0.966	1	0.916	0	0.083
6	0.866	0.933	0.8	0.076	0.2
7	0.933	0.937	0.928	0.071	0.071
8	0.933	0.947	0.909	0.090	0.090
9	0.8	0.937	0.642	0.1	0.357
10	0.806	1	0.666	0	0.333

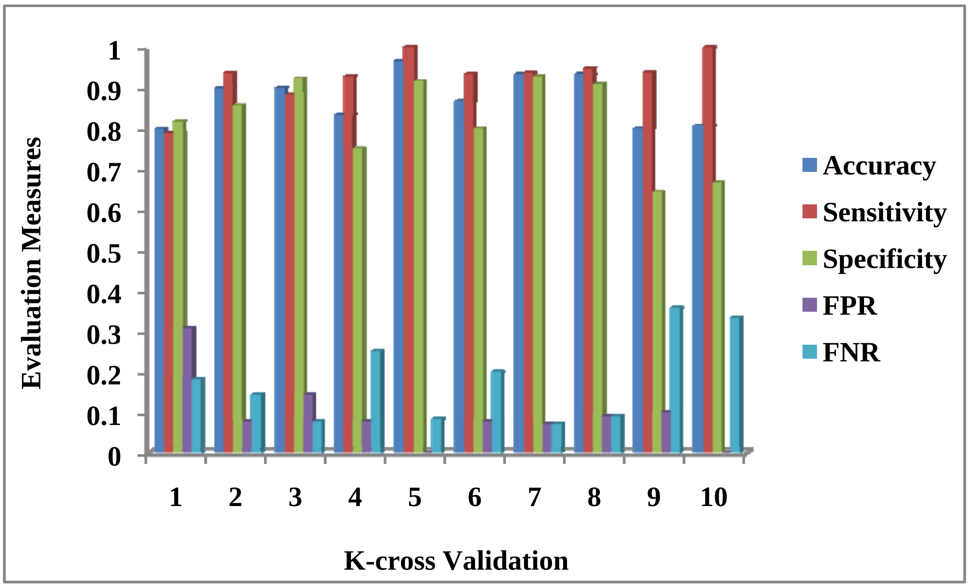

Fig 2: Graph for evaluation measures in Cleveland dataset

From table I, the evaluation metrics are analyzed for the Cleveland dataset, by which can observe the efficiency of proposed feature subset selection system. The results of the measures Sensitivity, Specificity, Accuracy, FPR, and FNR for Cleveland dataset are graphically represented in fig. 2. The sensitivity for a given data set is explained. The FPR and FNR values are very low for our proposed work. These low values offer a way to raise the classification accuracy.

5.2. Comparative Analysis

In our proposed feature subset selection and classification research makes use of SVM-GWO classifier. can establish that our proposed work helps to attain very good accuracy for the categorization of images utilizing SVM-GWO classifier from the above sections. And also can establish this categorization accuracy outcome by comparing other classifiers. have utilized SVM and Kernel SVM for our comparison in our work.

Cleveland Dataset

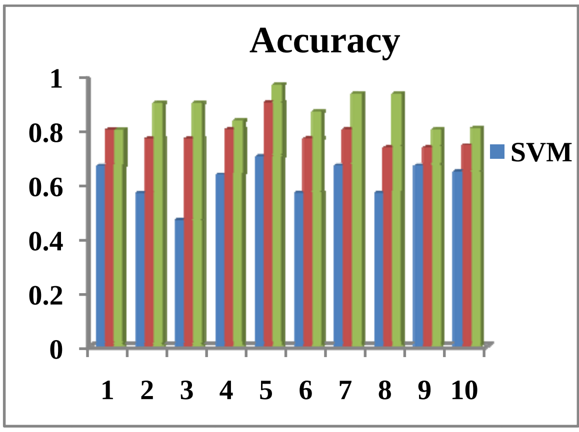

Fig 5: Accuracy comparison for Cleveland

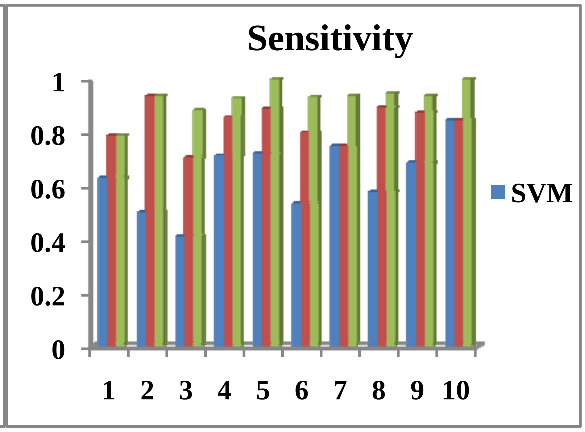

Fig 6: Sensitivity Comparison for Cleveland

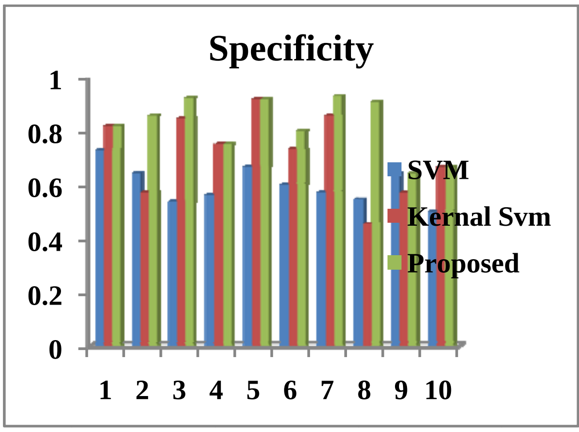

Fig 7: Specificity comparison for Cleveland

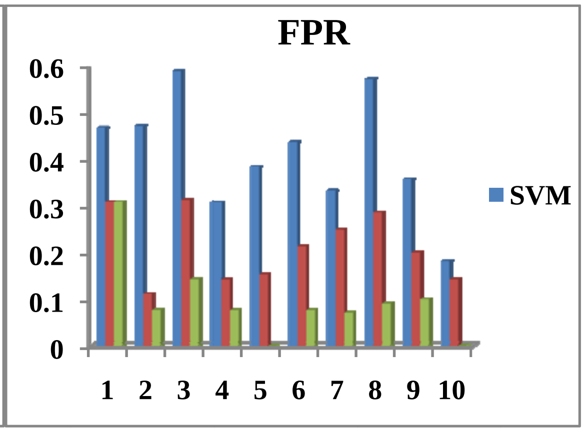

Fig 8: FPR Comparison for Cleveland

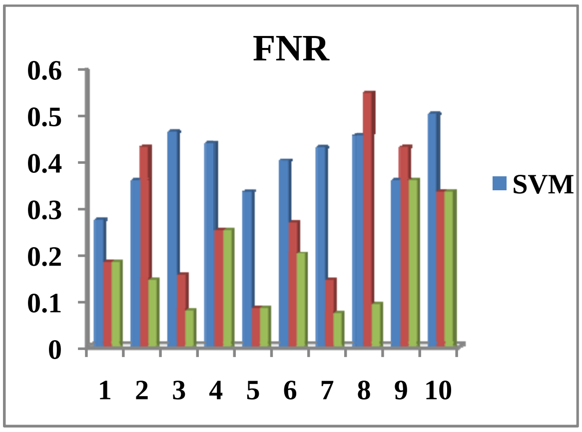

Fig 9: FNR Comparison for Cleveland

The overall accuracy evaluation measures as shown in fig.5 obtained for the existing SVM and Kernel SVM is 0.61 and 0.78 and sensitivity values fig.6. for existing method obtained value is 0.63 and 0.83 also the specificity value fig.7 forthe existing method is 0.59 and 0.71 which is low when compare these results to our proposed work it will given a better results for our proposed research obtained accuracy, sensitivity and specificity is 0.87, 0.92 and 0.82. the specificity value for the existing method is 0.57 and 0.68 which is low when compare these results to our proposed work it will given a better results for our proposed research obtained accuracy, sensitivity and specificity is 0.91, 0.87 and 0.82.and FPR fig.8,FNR fig.9.

CONCLUSION

In this paper propose a feature subset selection process can be performed with the aid of SVM-GWO classification technique. Here, input dataset will be as high dimensional or high features which are a great barrier for classification. Therefore, feature dimension reduction method will be applied to reduce the features' space without losing the accuracy of classification. Here, clustering with rough set theory based feature selection will be developed and used to reduce the feature dimension. In MFCM- RST (Modified Fuzzy c-means (MFCM) clustering algorithm with Rough Set Theory) will be combined and used to feature selection process. Once the feature reduction is formed, the classification will be done based on the Hybrid kernel based Improved SVM (ISVM) classifier. In this classification, the optimal kernel is identified using Grey Wolf Optimization (GWO). For experimentation, the dataset given in the UCI machine learning repository such as Cleveland, Hungarian and Switzerland etc., will be subjected to analyze the performance of the proposed technique in class imbalance problem utilizing accuracy, sensitivity and specificity. The results of our proposed method have shown that, our optimal classifier achieves better result when compared to other method.

REFERENCES

[1] S. D. Pachgade, S. S. Dhande, "Outlier Detection over Data Set Using Cluster-Based and Distance-Based Approach", International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 2, no. 6, 2012.
[2] Kononenko, "Inductive and Bayesian learning in medical diagnosis," Application of Artificial Intelligence, vol.7, no.3, pp. 317–37, 1993.
[3] Wolberg W-H, Street W-N and Mangasarian., "Machine learning techniques to diagnose breast cancer from fine-needle aspirates," Cancer Lett., vol.77, 1994.
[4] Kurgan, Cios, Tadeusiewicz, Ogiela and Goodenday, "Knowledge discovery approach to automated cardiac SPECT diagnosis", Artificial Intelligence in Medicine, vol.23pp.49–69, 2001.
[5] Jose Hernandez Orallo, M Jose Ramírez Quintana, Cesar Ferri Ramirez ,"Introducción a la minería de datos", Prentice Hall,2004.

[6] L. Goodwin, M. VanDyne, S. Lin and S. Talbert ,"Data mining issues and opportunities for building nursing knowledge" Journal of Biomedical Informatics, vol:36,pp:379-388,2003.

[7] Miller, A., B. Blott and T. Hames, "Review of neural network applications in medical imaging and signal processing. Med. Biol. Engg. Comp., 30: 449-464, 1992.

[8] Shortliffe, E. H., "Computer Programs to Support Clinical Decision Making", Journal of the American Medical Association, Vol. 258, No. 1, 1987.

[9] AlirezaOsareh and BitaShadgar, "A Computer Aided Diagnosis System for Breast Cancer" IJCSI International Journal of Computer Science Issues, Vol. 8, no.2, 2011

[10] NurettinAcır, OzcanOzdamar and CuneytGuzelis, "Automatic classification of auditory brainstem responses using SVM-based feature selection algorithm for threshold detection," Engineering Applications of Artificial Intelligence, Vol.19, 209-218, 2006

[11] Valentini, G., Muselli, M., and Ruffino, F., "Cancer recognition with bagged ensembles of support vector machines", Neuro computing, Vol. 56, 461-466, 2004.

[12] Y.L.Zhang, N.Guo, H.Du and W.H Li, "Automated defect recognition of C- SAM images in IC packaging using Support Vector Machines," The International Journal of Advanced Manufacturing Technology 25, 1191 -1196,2005.

[13] Karabatak, M., Ince, M.C., "An expert system for detection of breast cancer based on association rules and neural network," Expert Systems with Applications, Vol.36, pp.3465-3469, 2009.

[14] A.MehmetFatih, "Support vector machines combined with feature selection for breast cancer diagnosis," Expert Systems with Applications, Vol. 36, pp.3240-3247,2009.

[15] Jollife, I. T. Principal Component Analysis. Springer-Verlag, New York. 1986

[16] Eisen MB, Spellman PT, Brown PO and Botstein D: "Cluster analysis and display of genome-wide expression patterns", Proceedings of the National Academy of Sciences of the United States of America, vol.95, no.25, pp.14863-14868, 1998.

[17] Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM: "Systematic determination of genetic network architecture", Nature genetics, vol.22, no.3, pp. 281-285, 1999.

[18] Z. Pawlak, "Rough Sets- Theoretical Aspects of Reasoning about Data". The Netherlands, 1991.

[19] L. Polkowski, "Rough Sets- Mathematical Foundations", Physica-Verlag Heidelberg, 2003

[20] S. Dudoit, J. Fridlyand and T.P. Speed, " Comparison of discrimination methods for the classification of tumors using gene expression data" Journal of the American Statistical Association,vol.97,no.457,2002.

[21] B. Dennis, S. Muthukrishnan, "AGFS: Adaptive Genetic Fuzzy System for medical data classification", Applied Soft Computing, Vol. 25, pp. 242-252, 2014.

[22] Napolitano, Giulio, Adele Marshall, Peter Hamilton, and Anna T. Gavin, "Machine learning classification of surgical pathology reports and chunk recognition for information extraction noise reduction", Elsevier on Artificial Intelligence in Medicine, pp.77-83, 2016.

[23] Molina, Marco E., Aurora Perez, and Juan P. Valente, "Classification of auditory brainstem responses through symbolic pattern discovery", Elsevier on Artificial Intelligence in Medicine, Vol.70, pp.12-30, 2016.

[24] Last, Mark, Olga Tosas, Tiziano Gallo Cassarino, ZisisKozlakidis, and Jonathan Edgeworth, "Evolving classification of intensive care patients from event data", Elsevier on Artificial intelligence in medicine, Vol.69, pp.22-32, 2016.

[25] Yin, Xiao-Xia, SillasHadjiloucas, Yanchun Zhang, Min-Ying Su, Yuan Miao, and Derek Abbott, "Pattern identification of biomedical images with time series: Contrasting THz pulse imaging with DCE-MRIs", Elsevier on Artificial intelligence in medicine, Vol.67, pp. 1-23, 2016.