Journal of Theoretical and Applied Information Technology
31st January 2015. Vol.71 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

AN EFFICIENT HEURISTIC BASED TREE CLASSIFIER MODEL FOR MEDICAL DISEASE DIAGNOSIS

K.GAYATHRI[1], DR. M.CHITRA[2].
[1]Ph.D Research Scholar,
Department of Computer Science,
Bharathiar University and
Assistant Professor, Valluvar College of Science and Management, Karur.
India
[2]Professor, Department of IT,
Sona College of Technology, Salem,
India
[1]kstgayathri@gmail.com,[2]chitra_slm@gmail.com

ABSTRACT

The explosive growth of medical data and distributed sources augments the demand for efficient disease diagnosis and identify the cause of disease. The varying filtering concept used on the existing system performs the decision making process to exploit the contextual information. But filtering was not developed on medical expert support system for predicting drug effectiveness. Data publishing using slicing partitioned high dimensional data into parallel and perpendicular vectors. Slicing technique with column generalization was not examined thoroughly with classification implications. To thoroughly predict the drug effectiveness on medical expert support system, Smoothing Heuristic Divider Method of Classification tree (SHDC) is proposed in this paper. The main objective behind SHDC is to examine and classify thoroughly drug effectiveness using classification tree on the varying set of medical data. Initially, Support Vector Machine predictive model is developed in SHDC method to predict data quality and to perform mapping operation. Data quality in SVM is measured using double exponential smoothing technique to predict the medical data changes for varied intervals. SVM in SHDC method maps the input data vectors into high dimensional space to obtain optimal hyper plane support vectors on medical data support system. Secondly, classification is performed using Heuristic Divider Method to split (i.e.,) classify the medical data hyper planes for improving the searching process. Finally, the classifier is provided to support medical expert support system using Support Vector Machine which constructs separating hyper plane. With this, the separating hyper plane maximizes classifier margin and performs experimental evaluation on factors such as true rating rate, classification accuracy and finally percentage reduction in prediction time.

Keywords: *Heuristic Divider Method, Support Vector Machine, Hyper Plane, High Dimensional Space, Smoothing Technique, Predicting, Medical Expert Support System*

1. INTRODUCTION

Data Mining is the actively researched area in real world applications that explore medical stroke disease data objects. The data mining task is defined as the process selection that processes medical data objects which act as a useful choice when tasks classify and predict the outcome of the interpreted rules.

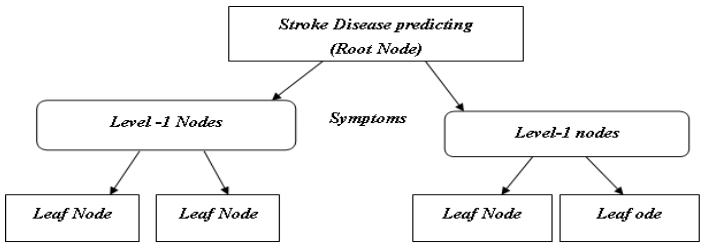

Fig 1 Structure of Classification Tree

Fig 1 shows the structure of tree based classification that classifies the stroke disease and observes the level of symptoms on selected dataset. The level of symptoms helps to identify the patient's severity of disease that classify into normal and critical position. There are three issues that have to be carefully considered during the data selection process. The first issue is to setup a

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

concise description system for handling medical data objects. The second issue is to predict the relevant data for problem definition. The third issue is to perform accurate classification on relevant data. To overcome all these above three issues, focus is made on developing the technique for predicting and classification of stroke medical data objects. The stroke medical data takes a lot of raw data and performs classification based on attribute information.

An extensive attention is provided on recommender systems to develop effective decision making process. Existing Collaborative Filtering based on Dempster-Shafer belief-theoretic framework as illustrated in [1] (CoFiDS) performed decision making process to exploit the contextual information. CoFiDS obtain prediction on smaller set of data model but did not develop a medical expert support system for predicting drug effectiveness. In [10], Group Incremental Approach was presented based on rough feature selection algorithm. Here multiple data objects were taken for construction of decision table, where new feature subset was found with much shorter time. However, incremental reduction algorithm was added up only to single data objects processing.

The survival of noisy, irrelevant, and unnecessary attributes was integrated with large set of diverse n-gram features. In [6], a rule-based multivariate text feature selection method called Feature Relation Network (FRN) controlled the survival of noise and irrelevant attributes. FRN dealt with semantic information and also the syntactic associations among n-gram features. FRN design adequately allowed addition of continuous sets of heterogeneous n-gram features for improving sentiment classification but not with additional potential medical feature relations. Ranking-based iterative classification framework, called Rank Class as described in [14] calculated the ranking allotment of objects surrounded by each feature related classes. Similarly a class membership was necessary for assessing quality ranking over a dataset.

Opinion mining is the process of analyzing opinions and performs the classification process. In [12], Intrinsic-Domain Relevance (IDR) and Extrinsic-Domain Relevance (EDR) scores were carried out on domain dependent and domain autonomous corpora correspondingly. Intrinsic-Extrinsic Domain Relevance (IEDR) made use of the disparities in distributional characteristics of features transverse on two corpora. IDR achieved greater than an additional threshold value on candidate features and EDR attained less than threshold value. IDR and EDR failed on employing fine-grained topic modeling approach to mutually classify opinion features including non-noun features, infrequent features, as well as inherent features.

Micro data contain records each of which contains information about an individual entity, such as a person, symptoms, personal medical report. There are several types of recoding for generalization. The recoding conserves the information with high privacy level. An existing privacy preserving data publishing called slicing approach as described in [2] partitions high dimensional
data into parallel and perpendicular points. Slicing technique with column generalization was not examined thoroughly with privacy implications. Encrypted data still provide the privacy for the patterns in [4] using spam filtering. The filtering of information formalized and generalized main ideas but failed to develop techniques for simulating non-linear medical data objects. However, encrypted index technique failed to perform classification with insertion and deletion medical information through selected data object points.

The proposed work is to perform effective classification process on the stroke medical data objects using Smoothing Heuristic Divider Method of Classification tree method. The support vector machine works on SHDC method to predict patient's diseases level by using on ego cardiogram medical data objects. The predictive model is effectively mapped in SHDC method and quality of mapping is attained using double exponential smoothing technique. The double exponential smoothing technique maps the input data vectors with a higher dimensional space. Heuristic Divider Method is used on constructing the classification tree by splitting medical data hyper planes. Finally work attained higher classifier rate margins by separating hyper plane in that space using SVM.

This section1 describes about basic concept of classification and also about the drawbacks on each types of existing classification process on data objects. The rest of this paper is structured as follows: Section 2 discusses the techniques of Smoothing Heuristic Divider Method of Classification tree using detailed architecture

Journal of Theoretical and Applied Information Technology
31st January 2015. Vol.71 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

model. Section 3 presents experimental setting and section 4 represents the results for evaluating the proposed SHDC method. Section 5 provides with related works. Finally, Section 5 gives concluding remarks.

2. Smoothing Heuristic Divider of Classification Tree Method on Medical Expert Support Data System

Data mining with medical data points is the most actively researched work for providing higher classification result on stroke medical data objects. Data mining with medical data points is the most actively researched work for providing higher using individual report of each user. Each user extracts the information and performs support vector machine based prediction. Support vector machine format in SHDC method is shown in Fig 2.

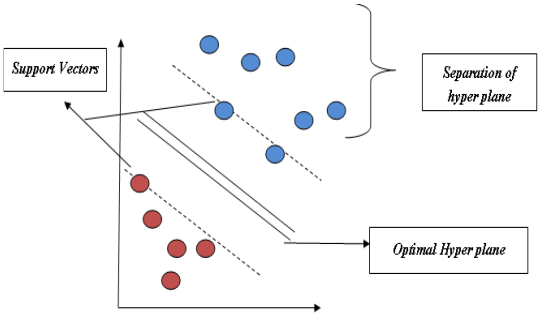

Figure 2 Support vector of SHDC method

Fig 2 describes the uses of prediction operation in SHDC method where different predictable result is generated for different medical data model. SVM represent yet an effective system to predict the data quality using mapping operation.

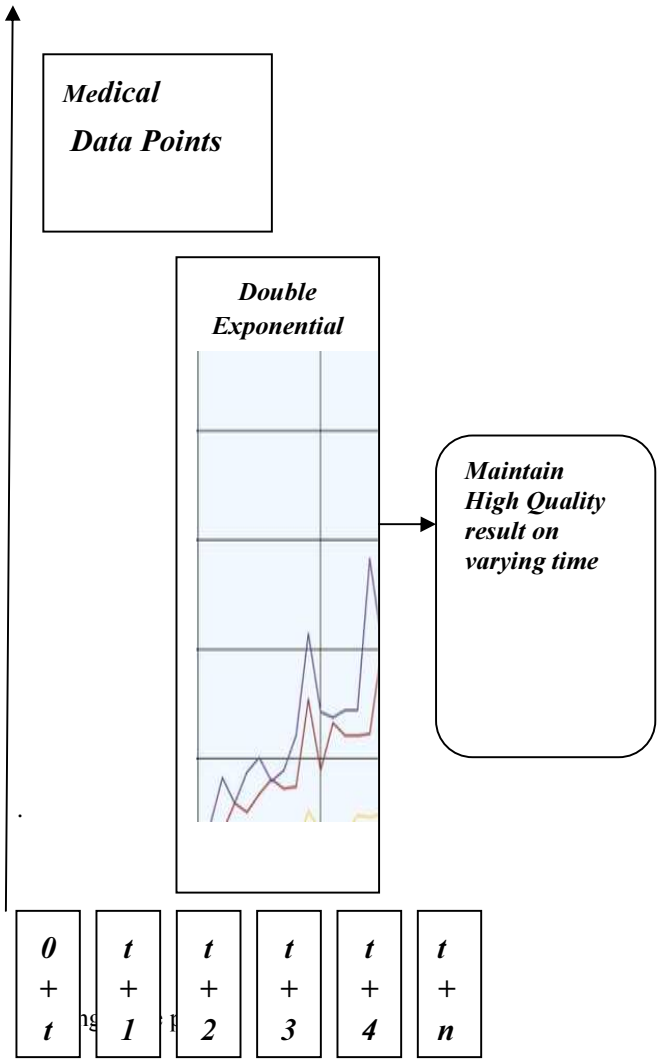

Figure 3.Double Exponential Smoothing On Varying Period Time

As illustrated in fig 3, the classification in SHDC method uses tree structure for easier search process. The classification is carried out with Heuristic Divider Method. The Heuristic Divider method classifies medical data points in the form of tree for easier classification. Support Vector Machine predictive model in SHDC method now uses the predictable medical data points for improving classification process. Predictable data points use the separating hyper plane points to improve classifier rate. All these steps are clearly illustrated with the aid of architecture diagram in Fig 4. Smoothing Heuristic Divider of Classification tree (SHDC) Method clearly elaborate the overall system process through medical data report.

Journal of Theoretical and Applied Information Technology
31st January 2015. Vol.71 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

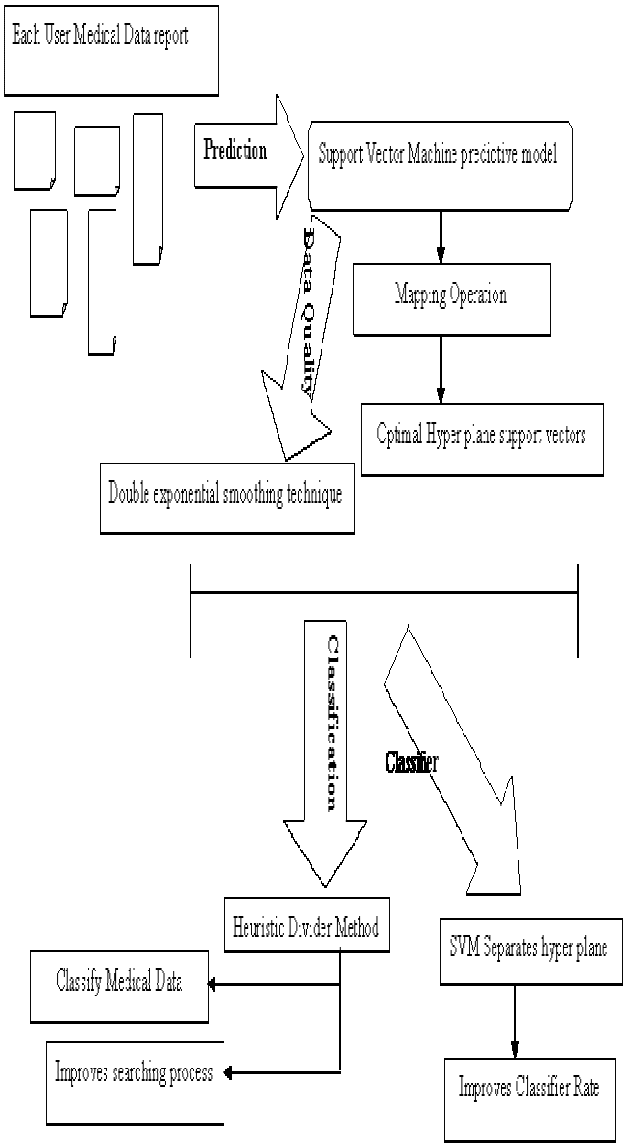

Figure 4 Architecture Diagram Of Shdc Method

As demonstrated in Fig 4, each user's medical report is taken into consideration for providing higher classifier rate after the disease model prediction. The medical disease is predicted using Support Vector Machine predictive model. The support vector model performs mapping operation to obtain optimal hyper plane support vectors to predict medical diseases accurately. The disease prediction obtains optimal hyper planes and data quality is maintained using double exponential smoothing technique. Double exponential smoothing technique in SHDC method computes equation using weighting function to measure data quality level.

The high quality medical data performs tree classification using Heuristic Divider Method. The Heuristic Divider Method classifies medical data to achieve effective searching process. SHDC method provides effective classification using SVM where proposed work also separates hyper plane effectively and maximizes classifier margin results. The sections given below briefly explain each processing step. Section 2.1 describes about support vector machine prediction steps, whereas section 2.2 and 2.3 describes the Heuristic Divider Method and SVM based tree classifier.

2.1 Support Vector Machine Predictive Model

Support Vector Machine prediction model learn the process of mapping. The mapping is a medical class data points a→b where object 'a' is in class 'b. The prediction on medical data points in SHDC method is formulized as,

$$a \in R^n, b \in \{0\} \quad \text{..... Eqn (1)}$$

R^n helps to predict relationship between medical data points for producing higher prediction result. The prediction using SVM in SHDC method is described as,

$$R(\alpha) = \int l(f(a,\alpha)dD(a,b) \quad \text{..... Eqn (2)}$$

Where, D (a, b) is divider data points on 'a' and 'b'. 'l' is the optimal super plane result attained on SVM. The relational function points 'α' produces effective prediction result after smoothing operation.

2.1.1 Double Exponential Smoothing Technique

Double exponential smoothing technique in SHDC method uses weighting function to reduce outliers and formularized as,

$$Weighting\ Function\ f(a(w,b) = sign(w.x + b)$$

..... Eqn (3)

Eqn (3) computes weighted function of object 'a' on class 'b'. The weighted points are computed as predicting equation based weight measurement changes from period to period time. The predicting algorithm in SHDC method is described as,

Journal of Theoretical and Applied Information Technology
31st January 2015. Vol.71 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

// Double Exponential Smoothing technique
Begin
Step 1: Predicting result

$$P_t = Single_t + (Single + 1_t)$$

Step 2: Performs the Single exponential prediction as,

$$Single_t = a_t + (1-\alpha)^2 e_t \quad \text{..... Eqn (4)}$$

Step 3: Performs the Double exponential prediction as,

$$Double\ Exponential = (Single + 1_t) = b_{t-1} + (\alpha)^2 e_t$$

..... Eqn (5)

Step 4: Smoothing is carried out as,

$$e_t = P_t - a_t$$

End

The prediction from varying period time 't' removes the outliers from medical data points. The smoothing constant 'α' indicate that the smoothing take place effectively in SHDC method. The single exponential prediction predicts medical stroke data objects with mild symptoms. The double exponential prediction predicts medical data objects of complicated attribute information. The double exponential effectively handles larger symptom set of patients medical data.

2.2 Heuristic Divider Method

The classification is carried out using Heuristic Divider Method. Heuristic Divider Method with 'N' positive integers objects A= $\{a_1, a_2, a_n\}$ of medical disease attain divider D= {1, 2…..N}.
The minimization discrepancy on classification tree is represented as,

$$Divider\ Discrepancy = |\Sigma_{i \in D}\, a_i - \Sigma_{i \notin D}\, a_i|$$

…… Eqn (6)

Divider in SHDC method attains perfect partitions, where it determines perfect classification with medical data information. The divider performs without overlapping of super planes in smoothing heuristic divider classification tree. The tree based structure helps to perform easier searching process with minimal search time in SHDC method.

2.3 SVM based Tree Classifier

SHDC method maintains classifier where 'p' entities privately own the blocks of classified medical data. The classifier objects are private from $a_1, a_2, \ldots a_n$, so that p-1entities are maintained in private classifier set. Entities all agree the data points and construct separate hyper plane space set to improve classifier margin rate. The tree classifier algorithmic procedure is described as,

//Tree Classifier Procedure
Step 1: SVM constructs the private classifier blocks with rows and column
Step 2: Predicted data points taken into consideration
Step 3: For predicted data
Step 4: Classify the data points through hyper plane divider tree
Step 5: Separating hyper plane carried out to improve classifier rate
Step 6: Class of classifier objects is maintained as,

$$b \in R_n$$

Step 7: Classifier margin attained on SVM

The classifier rate is attained on higher level using tree structure. The algorithmic step is clearly described to support the machine by attaining higher classifier margin on using stroke medical data objects. The predicted data on stroke medical data objects is then used to perform effective classification using Heuristic Divider method. The separating hyper plane maximizes classifier margin and attains higher performance rate.

3. EXPERIMENTAL EVALUATION

Smoothing Heuristic Divider of Classification tree (SHDC) Method is experimented on medical dataset. The stroke type medical dataset named Echocardiogram Data Set from UCI repository is used for experimenting different parameters. SHDC Method is compared against the existing Collaborative Filtering based on Dempster-Shafer belief-theoretic framework (CoFiDS) and Data Publishing through Slicing (DPS) approach. The dataset taken for experiment contain all information about patients who suffer from stroke disease on heart, while some of stoke diseased patients are alive and some are dead. The survival and still-alive variables are taken jointly and point out whether a patient endures for at least one year from the date heart attack.

Journal of Theoretical and Applied Information Technology
31st January 2015. Vol.71 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The main work of the SHDC Method is to classify the stroke disease and predict whether the patient will survive or not. There are almost 13 attributes with 132 instances. The attribute characteristics are categorical, integer, real form. The results are denoted in the binary form after the prediction. The '0' denotes the dead at end of the survival period. The '1' means the still in alive condition. SHDC Method performs the experimental evaluation on the factors such as true rating rate, classification accuracy, percentage reduction in prediction time, searching efficiency, computational time and false positive rate.

True rating rate is defined as the effective measure of proposition on identifying stroke disease. The stoke disease which correctly identifies rate of actual positives which are correctly identified (i.e., predicted).

$$True\ Rating\ Rate = \frac{True\ Positive\ Value}{True\ Positive + False\ Negative}$$

The correct classification of stoke disease using echo cardiogram information is termed as the classification accuracy. Classification accuracy is measured in terms of accuracy percentage. The prediction time is defined as the amount of time taken to predict the disease rate using SHDC Method. Prediction time is measured in terms of milliseconds.

$$Prediction\ Time = Start\ Time\ to\ Predict - End\ Time$$

Searching efficiency is defined as the effective rate of searching result to perform classification process. Searching is measured in terms of percentage. The searching efficiency value is improved to show higher margin rate on classification. The computational time is the time taken to compute overall process of classification.

$$Computational\ Time = Start\ Time\ of\ Classification\ process - End\ Time\ of\ particular\ process$$

False Positive rate is defined as the probability of identifying faulty identification of stroke diseases from particular medical objects.

$$False\ Positive\ Rate = \frac{False\ Result\ Value}{False\ rate + True\ rate}$$

The stroke disease set is reduced to fractional percent in SHDC Method when compared with existing system. The false positive rate is the expectancy of the false positive ratio.

4. RESULTS AND DISCUSSION

SHDC Method performs experimental result in section 4, where analyzed using the existing Collaborative Filtering based on Dempster-Shafer Belief-theoretic framework (CoFiDS) and Data Publishing through Slicing (DPS) approach. SHDC Method performs the parametric evaluation and Produce the result through table and graph.

Table 1 Tabulation Of True Rating Rate

Tree Level of Classification 'T'	True Rating Rate (%)		
	CoFiDS Frame work	DPS approach	SHDC Method
2	85.4	90.2	92.5
4	86.2	92.4	95.4
6	85.7	90.1	94.4
8	89.2	95.6	99.2
10	85.6	90.4	94.1
12	86.9	92.7	98.4
14	87.2	933	96.7
16	86.6	92.4	97.3

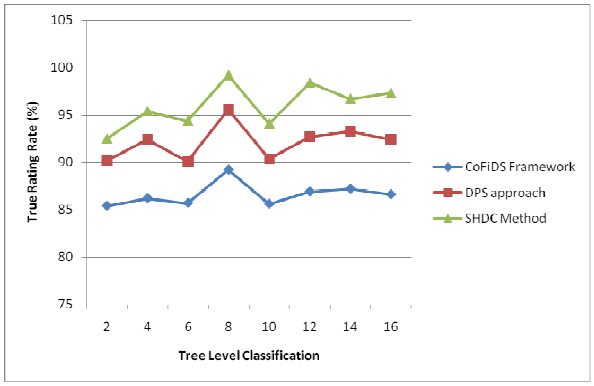

Figure 5 Measure of True Rating Rate

Fig 5 shows the true rating rate result using echo cardiogram data objects on CoFiDS Framework [1], DPS approach [2], and SHDC Method. The tree level of 'T'= 2 to 16 are taken for

Journal of Theoretical and Applied Information Technology
31st January 2015. Vol.71 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

experimenting the true rating rate. The SHDC method uses heuristic divider on each level of the tree structure and improves the true positive rate by 8 – 13 % when compared with CoFiDS Framework [1].The divider performs without any overlapping of super planes in smoothing system, so that the true rating rate is improved by 2 – 6 % in SHDC method when compared with existing DPS approach [2]. The tree based structure helps to perform higher true positive rate result on the SHDC method.

Table 2 Tabulation for Classification Accuracy

Medical Data Objects 'D'	Classification Accuracy (Accuracy %)		
	CoFiDS Framework	DPS approach	SHDC Method
50	79	82	85
100	75	79	82
150	72	77	80
200	73	79	83
250	74	78	82
300	75	80	85
350	77	82	88

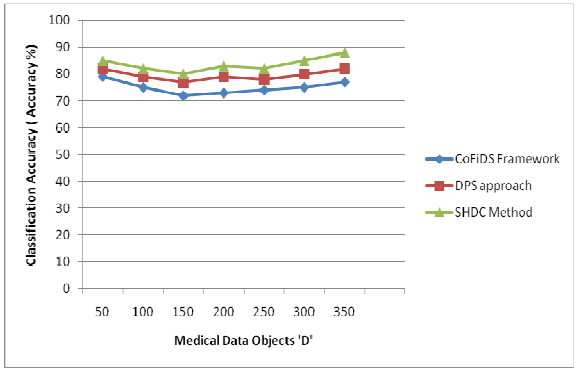

Figure. 6 Classification Accuracy Measure

Fig 6 describes classification accuracy level based on medical data objects 'D'. The higher accuracy on classification is attained using Heuristic Divider Method in the proposed system. The medical data objects count varies from the 50 to 350 to perform experimental evaluation. The higher count of data objects produces higher percent of classification result. Heuristic Divider method classifies medical data points in the form of tree for easier classification and produce 7 – 14 % higher accuracy percent in SHDC Method when compared with existing CoFiDS Framework [1]. Smoothing operation clearly improves the accuracy rate of classification by 3 – 7 % when compared with the DPSapproach[2]

Table 3 Tabulation of Prediction Time

Patients Count	Prediction Time (ms)		
	CoFiDS Framework	DPS approach	SHDC Method
5	225	200	180
10	415	375	345
15	625	595	565
20	850	810	780
25	1000	955	925
30	1125	1080	1050
35	1245	1165	1125

The table 3 clearly elaborates the system of percentage reduction on prediction factor in SHADC method. The prediction time is reduced in SHDC method when compared with the existing system [1, 2].

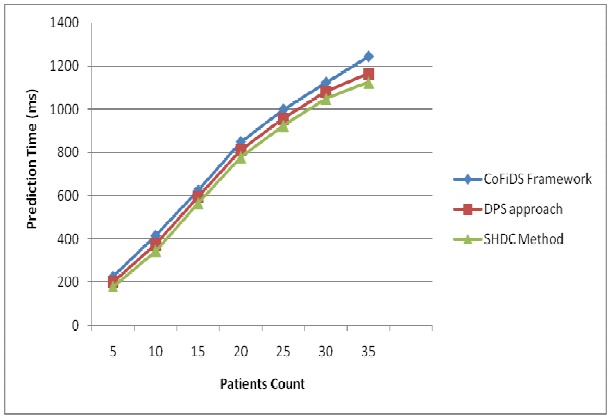

Figure 7 Prediction Time Measure

Fig 7 predicts the stroke disease based on prediction time. The prediction time of different set of patients count is taken for experimental work. The patient's prediction time is reduced in SHDC method using Support Vector Machine predictive operations. The SVM reduces the prediction time by 6 – 20 % in SHDC method when compared with the CoFiDS Framework [1]. The higher the patient count attains the lesser predictive time taken in SHDC method. SHDC method is also improved by 2 – 10 % when compared with the DPS approach [2]. Predictable data points separate the hyper plane points, thereby reducing the prediction time rate. For instance, on taking the 10 patients record for predicting the stoke disease, SHDC method consumes only 345 ms, whereas CoFiDS Framework consumed 415 ms and DPS approach takes 375 ms.

Journal of Theoretical and Applied Information Technology
31st January 2015. Vol.71 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 4 Searching Efficiency Tabulation

Data Object Size (KB)	Searching Efficiency (%)		
	CoFiDS Framework	DPS approach	SHDC Method
100	72	83	87
200	74	84	88
300	75	87	91
400	78	89	93
500	79	88	93
600	82	90	95
700	83	91	96
800	83	92	97

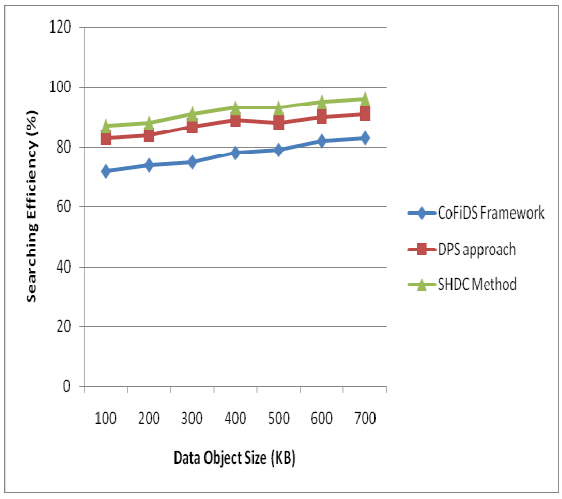

Figure 8 Searching Efficiency Measure

Fig 8 illustrates the efficiency of searching process based on data object size. The data object size (record size) is measured in terms of Kilo Bytes (KB). The searching efficiency is gained in SHDC method using double exponential smoothing technique The double exponential smoothing technique identifies all type of critical stroke disease in SHDC method, thereby improving search percent by 15 – 21 % when compared with CoFiDS Framework [1]. . The double exponential predict medical data changes from period to period time, so that proposed improved the searching efficiency by 4 – 5 % when compared with DPS approach [2]. The search efficiency rate of different data object size is taken for result evaluation work.

Table 5 Tabulation for Computational Time Measure

No of User(Patients)	Computational Time (sec)		
	CoFiDS Framework	DPS approach	SHDC Method
10	321	245	237
20	468	373	324
30	592	512	491
40	667	583	540
50	822	751	714
60	1153	1098	971
70	1247	1178	1038
80	1378	1298	1187

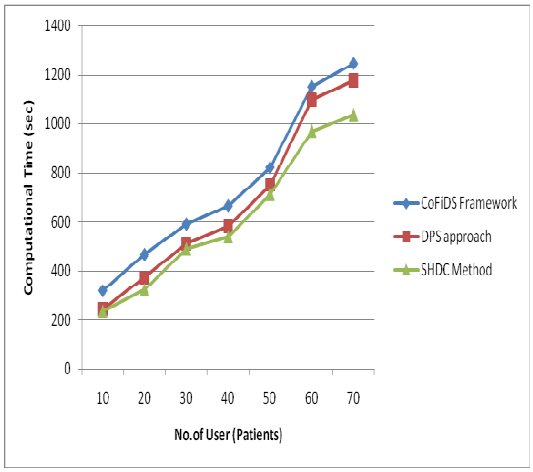

Figure 9 Measure of Computational Time

The attained computational time factor result is shown in Fig 9 briefly. The computational time of the SHDC method are compared against the existing CoFiDS Framework and DPS approach. Smoothing constant in SHDC method reduces the computational time factor by 13 – 30 % averagely in SHDC Method when compared with CoFiDS Framework [1]. The single exponential prediction predicts the medical stroke data objects with mild when compared with DPS approach [2]. The double exponential prediction predicts the medical data objects of complicated attribute information and also reduces the computational time. The computational time is completely reduced maximum rate f about 30 % in SHDC method when compared with existing systems

Journal of Theoretical and Applied Information Technology
31st January 2015. Vol.71 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 6 Tabulation of False Positive Rate

Tree Level of Classification	False Positive Rate (%)		
	CoFiDS Framework	DPS approach	SHDC Method
2	9.5	8.8	7.5
4	5.9	5.2	4.6
6	7.1	6.5	5.6
8	1.1	1.0	0.8
10	6.9	6.5	5.9
12	2.2	1.9	1.6
14	4.3	3.9	3.3
16	4.0	3.2	2.7

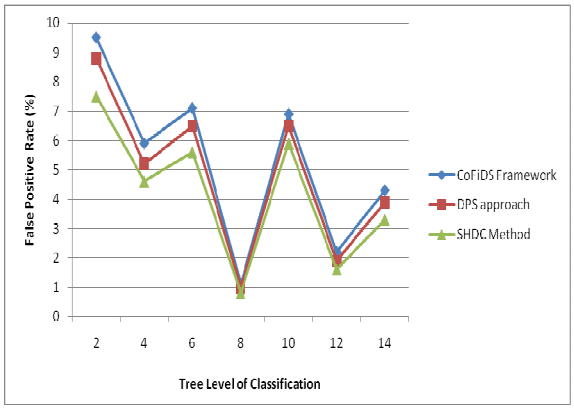

Figure 10 Measure of False Positive Rate

Fig 10 illustrates the false positive rate result based on the tree level of classification. The true rating rate result is used to easily identify the false positive rate result. The Double exponential smoothing technique in SHDC method uses the weighting function to reduce the outlier and remove the false positive rate ratio. The weighting function reduces the false positive rate by 10 – 25 % in SHDC method when compared with CoFiDS Framework [1]. The weighting function is analyzed on each tree level to identify the false positive result ratio at the initial stage. The predicting result is taken for performing the weighting function and reduces the false positive ratio by 14 – 27 % when compared with existing DPS approach [2].

Finally, Smoothing Heuristic Divider Method of Classification tree method performs the effective classification on the stroke medical data objects. The predictive model is effectively mapped in SHDC method and quality of mapping is attained using double exponential smoothing technique. The larger amount of patient's record contributes the higher accuracy of classification result in SHDC method when compared with existing systems.

5. RELATED WORK

Several feature based on class measures are used to guess statistical characteristics of training partitions. To assess the effectiveness of dissimilar training partitions, a huge number of disjoint training partitions with distinctive distributions is generated. In [3], Clustering, De-clustering, and Selection (CDS) partitioning method are used to remove the collision occurrences. The mounting events for training data subsets with overlaps fail to unite the filter based data partitioning approach with a wrapper based method. In [13], Probabilistic Sequence Translation Alignment Model (PSTAM) plan to confine both feature alignment and mapping among series. The sequences are analogous to translating one language into another in the field of machine translation but warping band constraints is occurred.

The users' activities essentially reproduce their habits, preferences, and the ability to understand users' activity information. In [7], an iterative learning procedure is formulated as gradient descent in data learning function space. A user's individuality is identified and correct activity classification is carried out with filtering procedure. A new collaborative boosting framework includes a text-to-activity classifier for each user and also involves a method for collaboration among classifiers of users having social connections. An online interpretation called Learning with Local and Global Consistency method (OLLGC) in [14] is fundamentally a second-order online learning algorithm. The second-order online learning algorithm with structural property of a graph classification is proved with selective sampling.

Classification is a most accepted machine learning task. In [9], Lazy Associative Classifiers (LAC) based on entropy minimization method which provides accurate evaluations of class membership probabilities. The class membership probabilities are employed on well calibrated classifiers, but the information retrieval (i.e.,) search process is not relayed accurately. Classification model as described in [11] assume that the true label of a data point perform the accurate information retrieval. Conversely, more than one novel class appears simultaneously, and

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

then all the classes of classification are regarded as a single class regardless of the instances.

The classifier building either a Rough Sets (RS) or Fuzzy-Rough Set (FRS) is generally divided into different categories. In [5], Fuzzy-Rough Set (FRS) Approach generalizes the system with respect to consistence degree. The consistence degree includes one prohibited threshold to attain the non-linear data information. FRS is not effective in specifying the reasonable high dimensional classification system with the development of Generalized Fuzzy-Rough Sets. Fuzzy unsupervised clustering method as established in [8] incorporates a mechanism for data classification. The data classification is based on segment result definition but fails to employ Gaussian Mixture model on medical non-linear objects. Gaussian Mixture model is not extended to alter the classification process with larger datasets.

6. CONCLUSION

In this paper, we proposed a Smoothing Heuristic Divider Method of Classification tree (SHDC) method to design an effective classification tree on stroke medical data objects. SHDC method utilizes Support Vector Machine predictive model system to efficiently predict the patient's medical data objects. With the predicted medical data objects obtained, Heuristic Divider Method is employed to construct the classification tree. The constructed classification tree finally constructs the SVM for separating the hyper plane and improves the classifier margin rate. Experimental results demonstrate that the proposed SHDC method not only leads to noticeable improvement on predicting stokes diseases on medical data objects, but also outperforms the classification margin accuracy by 5.015 percent. The computational time on classification is also reduced to 8.036 percent averagely while comparing with existing methods.

REFRENCES:

[1] Thanuka L. Wickramarathne., Kamal Premaratne., Miroslav Kubat, and Dushyantha T. Jayaweera., "CoFiDS: A Belief-Theoretic Approach for Automated Collaborative Filtering," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 23, NO. 2, FEBRUARY 2011

[2] Tiancheng Li., Ninghui Li., Jian Zhang., Ian Molloy., "Slicing: A New Approach to Privacy Preserving Data Publishing," IEEE Transactions on Knowledge and Data Engineering, volume:24, Issue:3, 2012

[3] Rozita A. Dara., Masoud Makrehchi., and Mohamed S. Kamel., "Filter-Based Data Partitioning for Training Multiple Classifier Systems," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 22, NO. 4, APRIL 2010

[4] Battista Biggio., Giorgio Fumera., Fabio Roli., "Security evaluation of pattern classifiers under attack," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING., 2013

[5] Suyun Zhao., Eric C.C. Tsang, Degang Chen, and XiZhao Wang., "Building a Rule-Based Classifier-A Fuzzy-Rough Set Approach," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 22, NO. 5, MAY 2010 Ahmed Abbasi, Stephen France, Zhu Zhang, and Hsinchun Chen, "Selecting Attributes for Sentiment Classification Using Feature Relation Networks", IEEE Transactions on Knowledge and Data Engineering, March 2011

[6] Yangqiu Songx, Zhengdong Luz, Cane Wing-ki Leungz and Qiang Yang, "Collaborative Boosting for Activity Classification in Micro blogs", ACM 2013

[7] Wenjing Zhang., and Xin Feng., "Event Characterization and Prediction Based on Temporal Patterns in Dynamic Data System," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, 2013

[8] Adriano Veloso, Wagner Meira Jr, Marcos Gonçalves, Humberto M. Almeida, Mohammed Zaki, "Calibrated lazy associative classification", ELSEVIER Science Direct 2011

[9] Jiye Liang, Feng Wang., Chuangyin Dang., and Yuhua Qia., "A Group Incremental Approach to Feature Selection Applying Rough Set Technique," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 26, NO. 2, FEBRUARY 2014

[10] Mohammad M. Masud., Jing Gao., Latifur Khan., Jiawei Han., and Bhavani Thuraisingham., "Classification and Novel Class Detection in Concept-Drifting Data Streams under Time Constraints," IEEE TRANSACTIONS ON KNOWLEDGE AND

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

DATA ENGINEERING, VOL. 23, NO. 6, JUNE 2011

[11] Zhen Hai., Kuiyu Chang., Jung-Jae Kim., and Christopher C. Yang., "Identifying Features in Opinion Mining via Intrinsic and Extrinsic Domain Relevance," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 26, NO. 3, MARCH 2014

[12] Minyoung Kim., "Probabilistic Sequence Translation-Alignment Model for Time-Series Classification," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 26, NO. 2, FEBRUARY 2014

[13] Ming Ji, Jiawei Han and Marina Danilevsky, "Ranking-Based Classification of Heterogeneous Information Networks", ACM 2011

[14] Quanquan Gu, Charu Aggarwal, Jialu Liu and Jiawei Han, "Selective Sampling on Graphs for Classification", ACM 2013