A New Hybrid Algorithm for Detecting Autistic children learning skills

Mrs. M.S. Mythili[1], Dr. A.R.Mohamed Shanavas[2]

[1]Ph.D Research Scholar, Bharathidasan University & Assistant Professor,
Department of Computer Applications, Bishop Heber College, Tiruchirappalli 620 017, TamilNadu, India.
[2]Associate Professor, Dept. of Computer Science, Jamal Mohamed College,
Tiruchirappalli 620 020,TamilNadu, India.
[1]mythhariniee24@yahoo.com
[2]vas0699@yahoo.co.in

***Abstract*- The rate of Children influenced with extreme autism spectrum disorder (ASD) is expanding rapidly. The extremely autistic children confront the issues in comprehension aptitudes, thinking capacity, learning and related abilities. The troubles that the extremely autistic children experience are various. The autistic Children frequently encounter trouble in learning aptitudes, interaction with others and dealing with the social abilities. The most fundamental part of the paper is to examine and to propose the best optimal and ideal features to surmount the learning obstacles and to accelerate the learning capacity of autistic children. The technique proposed comprises of three stages. The First stage identifies the Feature Extraction. The Second stage indicates the Feature Selection. Third stage adopts the classifier methodology. The Principal Component Analysis act as an evaluator in the Feature Extraction. Particle Swarm Optimization is utilized as a Feature Selection. The Particle Swarm Optimization assesses the parameter optimization to retrieve the most effective and best subset of features. The successful best subsets of features are fed into the Support Vector Machine classifier. A definitive result thus emerge show that the proposed technique acquires a higher accuracy with good exactness.**

Keyword-Feature Extraction, Feature Selection, Principal Component Analysis, Particle Swarm Optimization, Support Vector Machine

I. INTRODUCTION

Data Mining is a system for separating knowledge from vast measure of database to cull out the data, designs and proficient information required. Data Mining is nearly identified with the Knowledge Discovery in Databases. The primary essential methods conceivable are Unsupervised Learning-Clustering, Association Rule Discovery and also the Supervised Learning- Classification [1]. Educational Data Mining has pleasant and awesome potential for investigating the shrouded patterns inside the datasets of the educational domain. The datasets are utilized for evaluating and diagnosing the extremely affected autistic children [9]. Dimensionality reduction is the strategy of lessening the measure of attributes. There are two sorts of dimensionality reduction techniques. The essential sort and the first sort is feature extraction. The whole introductory and the original features are joined into a new reduced set of features. There are many feature extraction reduction algorithms. The most discriminating one is Principal Component Analysis (PCA) [1].

PCA is one among the most broadly utilized dimensionality reduction techniques for data analysis and compression. It is bolstered rebuilding an outsized scope of variables into a little extent number of inconsequential variables by discovering several orthogonal direct blends of the starting variables with the most essential variance. The second technique is feature selection. Feature selection is a well known procedure used to discover the most vital and optimal subset of features for building capable learning models. An effective feature selection system dispenses with immaterial and repetitive information data and it can enhance the precision of the classification rate. There are a ton of feature selection techniques. One of the algorithmic guidelines is Particle Swarm Optimization (PSO).It is a heuristic quest or optimization technique for getting the best potential arrangement.

Particle Swarm Optimization (PSO) was first presented by Kennedy and Eberhart. PSO is easier and simpler to execute with couple of parameters. This algorithm is totally effective and generally accustomed to solve optimization issues and also feature selection problems. This paper is catalogued as follows. Section 2 lists a literature survey. Section 3 shows the thought of Autism and examines the parts of Autism Spectrum Disorder. Section 4 explains a PCA for Feature Extraction. Section 5 clarifies the PSO for Feature Selection. Section 6 portrays the SVM classifier. Section 7 gives the Data Preprocessing and the usage of model construction. Section 8 identifies the proposed methodology and algorithm. Section 9 gives the Results and Discussion. Section 10 gives the conclusion.

II. LITERATURE SURVEY

Nasser H. Sweilam *et al*. [2] presented Particle swarm optimization, Quantum-behave Particle Swarm for training SVM. F Heppner *et al.* [3] suggested that PSO is an amazingly straight forward algorithm that is by all accounts effecting for enhancing an extensive variety of functions. Chung-Jui Tu *et al.* [4] described that PSO is utilized to actualize a feature selection, and SVMs with the one-versus-rest system serve as a fitness function of PSO for the classification problem. Bo-Tsuen Chen *et al.* [5] found that the normal arrangement precision rate of the methodology is 100% in the training subset, and be 88.98% in the test subset, and it is clear that the PSO-SVM methodological approach is as good as the grid search for SVM and original SVM. Dimitrios Bouzas *et al*. [6] demonstrated that by applying PSO on the sub-class Linear Discriminant Error Correcting Output Codes structure get a huge change in the classification performance. Enrique Alba *et al*. [7] centered that PSOSVM is able to find interesting genes and to give classification competitive execution. Smruti Rekha Das *et al.* [8] concentrated focused on SVM prepared utilizing linear, polynomial and RBF kernels.

III. AUTISM SPECTRUM DISORDERS

Autism Spectrum Disorders influence a child's mind health. Children with Autism Spectrum Disorders have challenges with correspondence and social collaboration abilities and show tedious practices. They have learning and formative difficulties. Autism Spectrum Disorders cause issues in significant regions of improvement such as Verbal and nonverbal correspondence, Social collaboration, Imaginative play, Sensory handling. Children on the autism spectrum may experience difficulty in understanding or imparting their needs to instructors and individual understudies [12]. They can experience issues seeing some classroom headings direction, and instruction, along with unpretentious vocal and facial signs of educators. Unseemly social connection can prompt testing practices, tormenting, each child with autism is distinctive. By and large, taking into consideration, the most widely recognized issue that youngsters with extreme introvertedness have is troubles with dialect. Both tuning in (responsive dialect) and talking (expressive dialect) are influenced in somebody with a mental imbalance. Due to challenges in correspondence, every branch of knowledge in school can be testing and tormenting. They are fit for adapting, yet they require an organized domain to do as such. Dialect abilities must be taught unequivocally. Connected Behavior Analysis is a treatment that helps set up nature to empower kids with extreme introvertedness to learn [13].

1V. PRINCIPAL COMPONENT ANALYSIS FOR FEATURE EXTRACTION

PCA is a factual system utilized for removing data from multi-assortment of dataset. It is an unsupervised dimension reduction approach. It aims at representing data utilizing orthogonal vectors so that the first voluminous data can be anticipated on a smaller space. It employs normalization to yield out the redundant features. The components are in view of the Eigen vectors of the covariance matrix acquired for the normalized dataset. It serves as an effective tool for analyzing data and offers the advantages of data compression by lessening the dimensions without any loss of information [1].

The PCA can be summarized by the following steps

1. The mean M of the set is defined by

$$M= \frac{1}{N}\sum_{i=1}^{N} X_i$$

2. The covariance Matrix C

$$C= \frac{1}{N}\sum_{i=1}^{N}(X_i - M)(X_i - M)T$$

3. The Eigen Vector and the Eigen value is determined by the equation

$CV=\lambda v$

4. Sorting the Eigen vector in the descending order from the Eigen value

5. The Eigen face is determined

6. The new set is retrieved.

V. PARTICLE SWARM OPTIMIZATION FOR FEATURE SELECTION

PSO is a powerful stochastic improvement system in light of the development and brainpower of swarms.PSO applies the idea of social collaboration to issue solving. It was created in 1995 by James Kennedy and Russell Eberhart. It utilizes various operators called particles. It constitutes a swarm moving around in the quest space searching for the best solution [3]. Each particle is dealt with as a point in an N-dimensional space which modifies its "flying" as indicated by its own particular flying knowledge and additionally the flying background of other particles. Each particle stays informed concerning its position in the arrangement space which is connected with the best arrangement that is wellness and that has accomplished so far by that particle. This quality is called individual best, pbest.Another best esteem that is followed by the PSO is the best esteem got so far by any particle in the area of that particle. This quality is called gbest [5].The essential idea of PSO lies in quickening every particle toward its pbest and the gbest areas, with an irregular weighted acceleration at every time

Steps:
Step 1: In every iteration, each particle is updated by following two "best" values, Personal best and Global best.
Step 2: After finding the two best values, the particle updates its velocity and positions with following equation
*v[m]=v[m]+c1*rand(m)*(pbest[m] - present[m]) +c2*rand(m)*(gbest[m]–present[m])(1)*
present[m]= present [m]+v[m] (2)
v[m] is the particle velocity, *present[m]* is the current particle (solution), *pbest[m]* and *gbest[m]* are defined as stated before, *rand(m)* is a random number between *(0,1)*
and *c1*, *c2* are learning factors. Usually *c1=c2=2*.

VI. SUPPORT VECTOR MACHINE CLASSIFIER

Support vector machine (SVM) is a prominent example for pattern classification method with numerous various applications. Support vector machine (SVM) is a new technique for data classification was initially recommended by Vapnik in 1995. SVM is utilizing Separating Hyper plane to recognize the data of two or several different Class that deal with the data mining issue of classification [2]. Sequential minimal optimization (SMO) is a standout among the most prevalent algorithms for large-margin classification by SVM. Kernel parameter setting in the SVM preparing methodology, along with the feature selection, fundamentally influences the classification accuracy. Several kernel functions help the SVM in obtaining the optimal solution. It contains diverse sorts of kernel such as Poly kernel, RBFKernel, Normalized Kernel, String Kernel. The RBF is generally applied most often, that it can characterize multi-dimensional data, dissimilar to a linear kernel function. The RBF has fewer parameters to set than other kernel. RBF is an effective option for kernel function. Proper parameters setting can enhance the classification accuracy of SVM. The parameters that ought to be improved and optimized incorporate the penalty parameter *C* and the parameters with different kernel function that is the *gamma value of* RBFkernel [1].

VII. DATA PREPROCESSING

Data preprocessing is an imperative one to eliminate the uproarious data, missing data and repetitive characteristics over the Datasets. The Datasets used in the classification algorithmic principle applied the data preprocessing to evacuate and purge the undesirable data's. The data must be clear, right and it is preprocessed for taking care of missing or repetitive attributes. The data ought to be a complete and it ought to be reliable data to deliver the best result from the Data Mining methodology.

A. Attribute Identification

Dataset collected from autism children database consists of

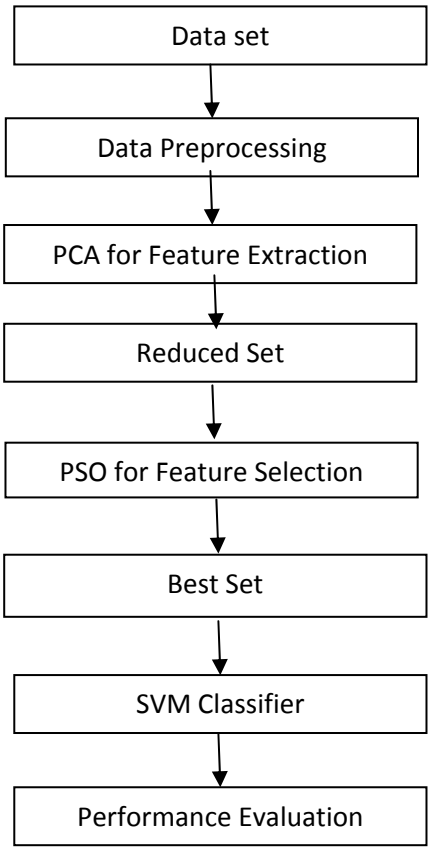
TABLE I
AUTISTIC DATASET

Attributes	Description
Attention level	Attention Skills of autism child
Hand writing level	Hand Writing Skills of autism child
Spell level	Spelling Skills of autism child
Lang level	Language Skills of autism child
Reading level	Reading Skills of autism child
Writing level	Writing Skills of autism child
Interaction level	Interaction Skills of autism child
Memory of child level	Memory Skills of autism child
Listen level	Listening Skills of autism child
Speaking level	Speaking Skill of autism child s
Maths level	Mathematical Skills of autism child
Science level	Science Skills of autism child
Art level	Art Skills of autism child
Music level	Music Skills of autism child
Physical level	Physical Skills of autism child
Child class label	Levels of autism child(High,Medium,Low)

B. IMPLEMENTATION OF MODEL CONSTRUCTION

The Weka is the open programming framework software system tools and it is the mix of machine learning calculations algorithms and the data preprocessing. It is broadly used in data mining applications which contain the association rule mining, classification and clustering. It acknowledges only the ARFF (Attribute Relational File Format), CSV (Comma Separated Value) file. The recognition of autism children learning skills has distinctive properties of attributes such as attention skills, Handwriting skills, spelling skills, language skills, Reading skills, writing skills, memory skills, listening skills,maths skills, science skills, music skills, physical skills, speaking skills, art skills, and the interaction skills, 200 instances were taken for the analysis and detection. The Feature Extraction and the Feature Selection approaches have been implemented in Weka tool.

VIII. PROPOSED METHODOLOGY



A. Proposed Algorithm

The proposed Algorithm is as follows:
Input: D (F_1, F_2,…..F_n)// Training data with n features
Output: Best Subset with maximum Accuracy and minimum error rate

//Initilization

Var=0.25, center data=false, max.attributes=-1, Best data=max_value

//Feature Extraction

1. For each unsupervised Attribute evaluator S

2. Input data=Filter PCA(S,var) //Get Input data.

3. Subtract the mean value

4. Compute the covariance matrix and set the var to be 1

5. While [(Input data > Best data) and var! =1] do

6. Best data=Input data

7. var=var+0.05

8. Calculation of eigen vector with eigen values and Ranked attributes in descending order.

9. The Best data is obtained

// Feature Selection

10. within each Best data

11. Initialize the parameters of PSO search ie Population size=100, Number of generations=50, c1=c2=2.0, Report frequency=50, Random number seed=1

12. Repeat

13. For i= 1 to n do

// $f(x_{i)=}$fitness function measures the error rate of x_i

14. If $f(x_i)<f(pb)$ // calculation of pbest

15. $pb=x_i$

16. End

17.End

18. If f(pb)< f(gb) //calculation of gbest

19. end

20.end

21.for i=1to n do

22.update the velocity according to (1)

23.update the position according to (2)

22.end

23.until it returns the gbest

//**Classification Process**

24. The gbest particle in the swarm is fed to the SVM classifier with 10-fold cross validation.
25. Make the build logistic models to be true and the c parameter=2.0, with the Normalized PolyKernel
26. Performance Evaluations gives the maximum Accuracy and minimum error rate

IX. RESULTS AND DISCUSSION

The analysis and interpretation of classification is lengthy procedure that needs a profound comprehension of insights. The process needs a lot of time to evaluate and come to a definite conclusion to arrive at the classification and relationships within the data.

TABLE II

PERFORMANCE RESULT OF CLASSIFIERS

Evaluation Criteria	PCA+SVM	PCA+PSO+SVM
Correctly classified instances	95	98.5
Incorrectly classified instances	5	1.5
Time taken	0.39	0.30

In Table II the percentage of correctly classified instances is generally eluded as accuracy of the model. Henceforth the Principle Component Analysis and Particle Swarm Optimization with the Support Vector Machine with Normalized Polykernel can be termed as more precise than other classifiers.

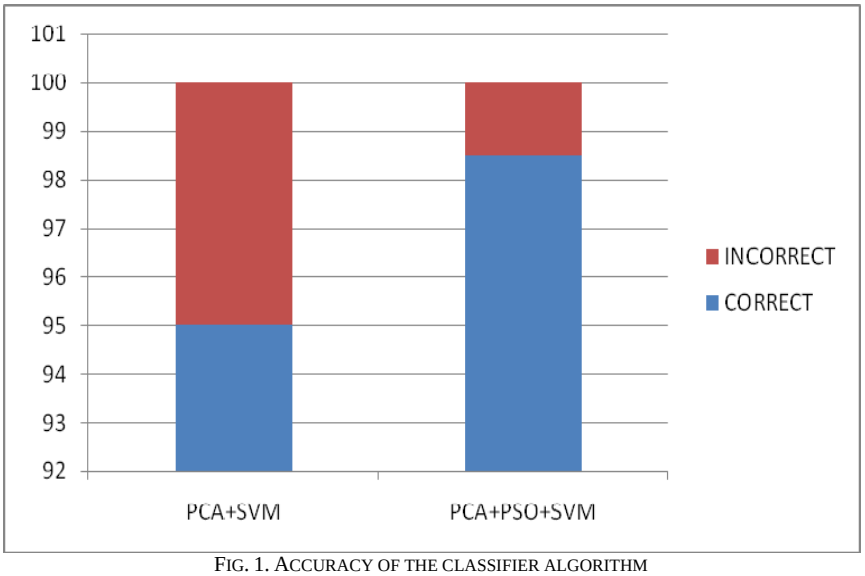


FIG. 1. ACCURACY OF THE CLASSIFIER ALGORITHM

The Figure 1 clarifies the graphical representation of correctly classified instances of results of autism children learning skills construct basically with respect to extreme introvertedness on autism children dataset. The highest percentage of correctly classified instances is the Principle Component Analysis and Particle Swarm Optimization with the Support Vector Machine with Normalized Polykernel is an absolute best classifier for analyzing the autism children performance result.

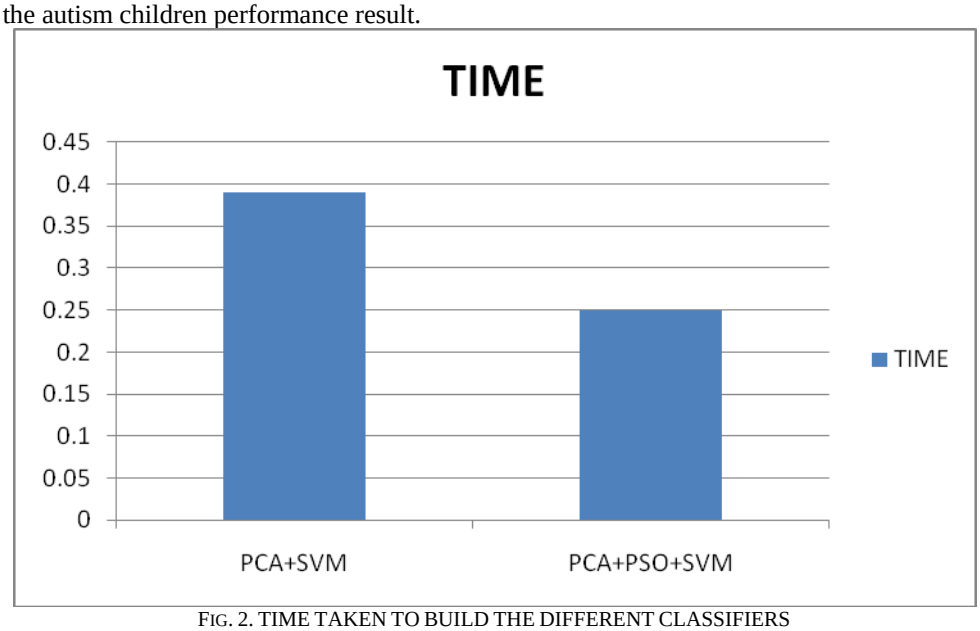


FIG. 2. TIME TAKEN TO BUILD THE DIFFERENT CLASSIFIERS

The Figure 2 demonstrates that the graphical representation of time taken to fabricate the consequences of extreme introvertedness autism children learning skills based on dataset. It plainly uncovers that the Principle Component Analysis and Particle Swarm Optimization with the Support Vector Machine with Normalized Polykernel is an absolute best classifier for investigating the autism children performance result consuming less time coupled with great precision.

TABLE III
ERROR MEASUREMENT FOR CLASSIFIERS

Evaluation Criteria	PCA+SVM	PCA+PSO+SVM
Kappa statistic	0.9175	0.9595
Mean absolute error	0.0272	0.0179
Root mean squared error(RMSE)	0.1436	0.1288

In Table III Kappa statistics is a measure of the degree of non random agreement between observers and measurement of a particular categorical variable. The root mean square error and Mean absolute error of

Principle Component Analysis and Particle Swarm Optimization with the Support Vector Machine with Normalized Polykernel is a very best classifier compared to other classifiers.

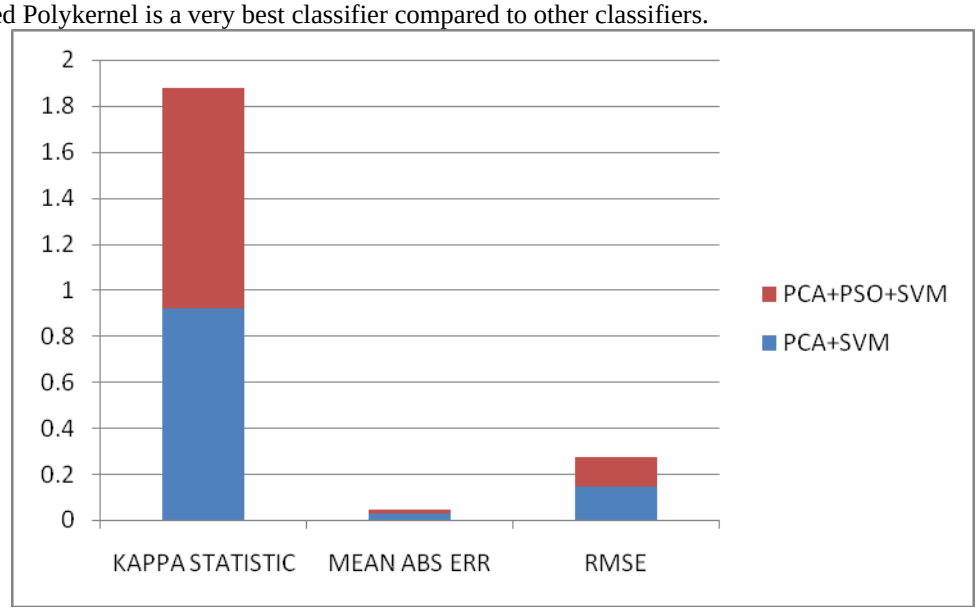


FIG. 3. ERROR RATE OF DIFFERENT CLASSIFIERS

The Figure 3 deals with the errors among totally with the diverse classifiers. Principle Component Analysis and Particle Swarm Optimization with the Support Vector Machine with Normalized Polykernel is an absolute best classifier has lower error rate contrasted with distinctive classifiers. In this way the Principle Component Analysis and Particle Swarm Optimization with the Support Vector Machine with Normalized Polykernel is the effective classification technique among remaining classifiers.

TABLE IV
CLASS LABEL ACCURACY FOR CLASSIFIERS

classifier	TP	FP	Precision	Recall	Class
PCA+SVM	0.714	0	1	0.714	Low
	1	0.095	1	0.95	High
	1	0	1	1	Medium
PCA+PSO+SVM	0.857	0	1	0.857	Low
	1	0	1	1	High
	0.857	0	1	0.857	Medium

The Table 4 unmistakably demonstrates the execution of each classifier in view of the true positive rate and false positive rate, precision, recall and distinctive measures. These measures are extremely useful for looking at the classifiers taking into account the exactness. The Principle Component Analysis and Particle Swarm Optimization with the Support Vector Machine with Normalized Polykernel beat every distinctive classifier inside the children dataset.

TABLE V
CONFUSION MATRIX

Classifier	High	Low	Medium	Class
PCA+SVM	6	1	0	Low
	0	18	1	High
	0	0	14	Medium
PCA+PSO+SVM	6	0	1	Low
	0	18	1	High
	0	0	14	Medium

The Table 5 describes that the confusion matrices are very useful for examining the classifiers.

X. CONCLUSION

The work investigates the power of machine learning algorithms in choosing the impact of result, different components like attention ,Hand writing, spelling, language, Read, write, memory skills and result from the autism school student learning abilities and analyze of autism children performance. It is found that the Principle Component Analysis and Particle Swarm Optimization with the Support Vector Machine with Normalized Polykernel is best than that of distinctive algorithms utilized in the study. This study is appallingly valuable for the instructive foundations. In future, it is feasible to expand the investigation by utilizing diverse classification procedures and association rule mining for the autism children dataset.

REFERENCES

[1] Jiawei Han, Michelin Kamber, Data Mining: concepts and techniques, 2nd edition, Morgan Kaufmann publishers, 2008.

[2] N.H. Sweilam, A. Tharwat, A. Abdel, N. K. Moniem, “Support vector machine for diagnosis cancer disease”, Egyptian Informatics Journal , Volume (11), pp. 81–92, 2010.

[3] F. Heppner, U. Grenander, “A stochastic nonlinear model for coordinated bird flocks”, AAAS publications, Washington, 1990.

[4] Tu, Chung-Jui, Li-Yeh Chuang, Jun-Yang Chang, and Cheng-Hong Yang. "Feature selection using PSO-SVM." IAENG International journal of computer science 33, no. 1 (2007): 111-116.

[5] B.T. Chen, M.Y. Chen, “Applying particles swarm optimization for support vector machines on predicting company financial crisis”, International Conference on Business and Economics Research, 2010.

[6] N. Arvanitopoulos, D. Bouzas, A. Tefas, “Subclass Error Correcting Output Codes Using Fisher’s Linear Discriminant Ratio, Pattern Recognition”, pp. 2953-2956, 2010.

[7] E.Alba, J. Garc´ıa-Nieto, L. Jourdan, E. Talbi, “Gene selection in Cancer Classification using PSO/SVM and GA/SVM Hybrid Algorithms”, Congress on Evolutionary Computation, 2007.

[8] Smruti Rekha Das, Pradeepta Kumar Panigrahi, Kaberi Das, Debahuti Mishra, “Improving RBF Kernel Function of Support Vector Machine using Particle Swarm Optimizaton”, International Journal of Advanced Computer Research ,pp 2277-7970,Volume -2,Number -4,Issue-7,December-2012.

[9] M.S. Mythili, A.R.Mohamed Shanavas,”A Study on Autism Spectrum Disorders using classification Techniques”,IJSCE,ISSN:2231-2307,Volume 4,Issue 5,Nov 2014.

[10] M.S. Mythili, A.R.Mohamed Shanavas,”A Novel Approach to predict the learning skills of Autistic children using SVM and Decision Tree, IJCSIT, ISSN: 0975-9646, vol5 (6), 2014, 7288-7291

[11] M.S. Mythili, A.R.Mohamed Shanavas,” An Improved Feature Selection (IFS) Algorithm for Detecting Autistic Children learning Skills”, BBRA, April 2015, Vol.12 [1], 499-505.

[12] “Guidelines for Educating Students with Autism Spectrum Disorders”, October 2010 Virginia Department of Education, Office of Special Education and Student Services.

[13] “Guidelines for Identification and Education of Children and Youth with Autism “, July 2005, DEPARTMENT OF EDUCATION Division of Teaching and Learning Programs and Services Bureau of Special Education