A Proficient Optimized Feature Selection Method Germane For Autism Spectrum Disorder Classification

R.Abitha, Dr S. Mary Vennila

Abstract: In this modern scenario, Autism is a growth state in connection with healthcare expenditures, consequently, initial examining of autism indications can be censored down on these expenses. The autism examining method comprises of bestowing a sequence of queries for caretakers, paternities, and household members to respond in support of the child to regulate the perspective of autistic behaviours. Every so often, present autism screening devices, like the Autism Quotient (AQ), contain numerous queries, along with the cautious scheme of the inquiries, which sorts the autism screening method to prolonged duration. The main objective is to detect and proliferation the appropriate actions which can support to detect the features that disturb erudition. As the technology, Data mining grips such health grounds to forecast by evaluating designs in colossal data sets. The framework of the proposed research is to regulate the high pertinent and low redundant features from the dataset by the optimization method and feature selection method. In this work, the classification of the feature is designed using Symmetrical Uncertainty and the optimized feature set can be resolute with the guidance of the Cultural Algorithm.

Keywords: Autism Spectrum Disorder, Data Mining, Feature Selection, Optimization technique, Symmetrical Uncertainty, Cultural Algorithm.

—————————————————— ◆ ——————————————————

1 INTRODUCTION

ASD screening is the method by which the autistic indications of an individual can be firmed [1][2]. This is a critical stage of ASD analysis as autism can't be recognized by traditional medical approaches like body check-ups or blood tests. And there are numerous kinds of autism screening devices that contain straight observation, structured and semi-structured inquiries and discussions [3]. As a result of a deficiency inconsistent procedures in broadcasting children for autism, in several circumstances, the indications become perceptible only subsequently they grow into adults. Thus, the part of a feasible screening tool for classifying the threat of ASD at the initial phase is immense. In effect, ASD screening methods depend on a modest domain professional, in addition to an enormous number of queries that respondents have to reply, therefore these methods have been assessed by researchers for being extensive and subjective [4][5]. And so, evolving detection structures that can be mined by robotic techniques could be a favourable direction. This method of learning is known as data mining and classically employs a chronological dataset to determine operative hidden structures for the refining plan and the result progressions [6], [7]. Contemporary studies in autism research, mainly focused on ASD diagnosis, for instance, [8][9] and others, showed that machine learning methods and data mining could enrich the exactness and competence of the investigative stage. On the other hand, there has been little advancement in exploring the data mining methods within autism screening is owing to the unapproachability of datasets. With the progression of mobile technology, the contemporary dataset associated with behavioural features of autism has been suggested by [10].

————————————————————
- R.Abitha , Asst.Professor, WCC and Research Scholar ,PG & Research Dept of Computer Science, Presidency College, University of Madras, Chennai, E-mail: abirae2000@yahoo.co.in
- Dr S.Mary Vennila ,Head, Associate Professor, PG & Research Dept of Computer Science, Presidency College, University of Madras, Chennai,
  E-mail: vennilarhymend@yahoo.co.in

2   RELATED WORKS

Sudha, V. Pream, and M. S. Vijaya [11] employed supervised machine learning techniques to construct a model to identify syndromic ASD by classifying mutations that underlie these phenotypes, and supervised learning algorithms, namely support vector machines, decision trees and multilayer perceptron, are used to explore the results.Al-Mejibli, Intisar Shadeed, et al [12] studied the behavior of SVM in regarding to the used attributes of dataset with different kernel functions. It analyzed the influence of various datasets descriptions on efficiency of (SVM)classification. Thabtah, Fadi, Firuz Kamalov, and Khairan Rajab [13] aimed at identifying fewer, albeit influential, features in common ASD screening methods in order to achieve efficient screening as demands on evaluating the items' influences on ASD within existing tools is urgent. To achieve this aim, a computational intelligence method called Variable Analysis (Va) is proposed that considers feature-to-class correlations and reduces feature-to-feature correlations. Thabtah, Fadi, and David Peebles [14] proposed a new machine learning method called Rules-Machine Learning that not only detects autistic traits of cases and controls but also offers users knowledge bases (rules) that can be utilized by domain experts in understanding the reasons behind the classification. Tejwani, Ravi, et al [15identified autism using machine learning techniques and resting-state brain imaging data, leveraging the temporal variability of the functional connections (FC) as the only information. We estimated and compared the FC variability across brain regions between typical, healthy subjects and autistic population by analysing brain imaging data from a world-wide multi-site database known as ABIDE (Autism Brain Imaging Data Exchange). Wang, Mingliang, et al [16] proposed a multi-center low-rank representation learning (MCLRR) method for ASD diagnosis, to seek a good representation of subjects from different centers. Specifically, we first choose one center as the target domain and the remaining centers as source domains. We then learn a domain-specific projection for each source domain to transform them into an intermediate representation space. Li, Genyuan, Olivia Lee,

and Herschel Rabitz [17] presented a new method based on scores produced in Support Vector Machine (SVM) modeling combined with High Dimensional Model Representation (HDMR) sensitivity analysis. The new method effectively and efficiently identifies the key causative metabolites in FOCM/TS pathways, ranks their importance, and discovers their independent and correlative action patterns upon ASD.

3 FEATURE SELECTION

Data Pre-processing does removal errors, maintains the consistency and accountability of missing values in most of the real-world data which are incomplete with aggregate and missing values, noisy data with errors, outliers and duplication records. The quality of mining depends on quality data for quality decision making. Hence this paperwork addresses the importance of pre-processing tasks through normalizing the attributes in the given dataset. The feature selection is a vital procedure that commonly smeared in structure recognition, machine learning, and data [18]. The key profits of feature selection techniques are:

- Reducing computation time
- Improving classification performance and
- A better understanding of the data in various mining applications

In this research paper, a cultural algorithm-based feature selection technique is recommended to regulate the extremely pertinent and low redundant features in the dataset to mend the classification exactness and lessen the calculation time. In advance, the optimization of the feature set, filter-based feature selection algorithm called Symmetrical Uncertainty is employed to rate the structures of the dataset.

3.1 SYMMETRICAL UNCERTAINTY

The symmetrical uncertainty (SU) [19] between target concept and features are employed to capture the best features for classification. The elements with higher SU values have a higher weight. SU measures the relationship among A, B variables based on the information theory. It was calculated as follows

$$SU(A, B) = 2 \frac{I(A, B)}{H(A) + H(B)}$$

Estimating I(A, B) as the MI among A, B. H(..) as an entropy function for A, B features. The SU indicates the normalized range value [0,1] as the correction factor value is 2. If SU value is 1, then the information of one feature is predictable. If SU value is 0, then A, B are not associated.

3.2 CULTURAL ALGORITHM

Reynolds [20][21]-[8][9] hosted Cultural Algorithms (CA) as an advanced method that is derived from the cultural growth method in nature. It comprises of trust and population spaces and a set of communication channels between these galaxies to regulate the eminence of the pooled information and its brand. The fundamental pseudo-code of the CA structure is presented in Fig.1.The figure displays how the leading stages of CA are executed in each group. The Obj() function generates the characters in the population space and the Accept() function elites the greatest individuals that are habituated to evaluate the trust space knowledge by the function Update(). The Influence()

function adventures the roulette wheel selection to pick one knowledge source to attain the development of the future generation. A CA is a knowledge-based evolutionary automated classification. Its fundamental data is to incorporate knowledge tools into predictable advanced computational systems. Its reproductions are two phases of improvement: the belief space level and the population space level. The two space levels are linked together by a firm communication procedure composed of an acceptance function and an influence function, which are indicated here as Accept() and Influence(), respectively. The acceptance function is experienced to accumulate the consideration of selected individuals from the population; then the belief space can be modified by an update function, epitomized here as the update(); next, the influence function can make use of the problem-solving knowledge in the belief space to aid the development of the population component. In the population space, like conventional evolutionary population simulations, individuals are primarily evaluated by a generation function objective (). Innovative individuals are created by a generation function generate (). At that point, a selection function select(), is deployed to select the population for the next generation. Cultural algorithm has the features: (1) Dual evolutionary inheritance: In the population space and belief space are innate parent data; (2) Population space evolution is sheltered by the belief space knowledge to guide; (3) Support the population space and belief space grading; (4) Facilitating the adaptive evolution of two space; (5) Different space evolution can be carried out at different speeds; (6) Upkeep a combination of dissimilar algorithms to explicate the issue; (7) "Cultural" variation can be expressed in different techniques within a protocol
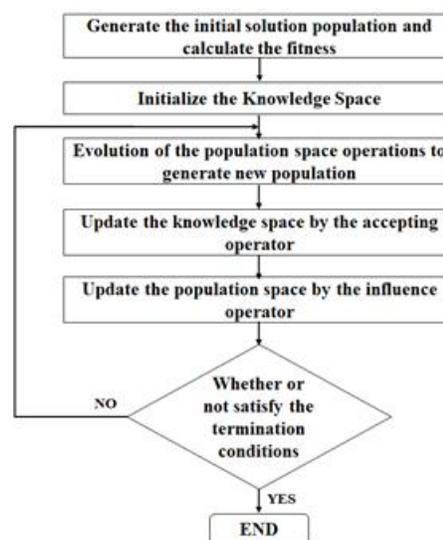

Figure 1: *Flowchart of Cultural Algorithm*

4 PROPOSED CULTURAL ALGORITHM BASED FEATURE SELECTION (CA-FS) METHOD

CA is an adaptive mutation technique performing a heuristic search, inspired by the evolution process of genetics with the belief space. A population, containing the challenging results, is preserved, which experiences selection, crossover, and transformation to advance and congregate to the finest solution. A parallel search is achieved on the

solution space to discover an optimal solution without getting trapped in a confined optimum. CA can create favourable results for feature selection over a high-dimension space attributable to its sturdiness to the fundamental search space size and multivariate disseminations. In general, the necessities for probing an optimal solution in the entire feature space comprise a search engine with a preliminary condition, state-space, and a termination condition. Given n number of features, the size of the search space is 2n-1. As every feature has two possible states: "1" or "0", an n bit string will have 2n possible combinations. Assume τ features, which are not important to decision making in terms of the values of their SU, be removed. The length of a binary string becomes n-τ. Even in the reduced search space (2n-т), a brute-force search for a large space of 2n-т is still infeasible. Of course, such space reduction is worthy of the CA Wrapper search. The following are phases intricate in the recommended Cultural Algorithm based Feature Selection Technique.

STAGE 1: INITIALIZATION OF POPULATION: CA maintains a diverse population $x_{1\ldots n} = \langle x_1 \ldots x_n \rangle$ of n individuals xi, the candidate solutions. The fitness of these individuals is assessed by computing an objective function $F(x_i)$ that is to be optimized for a specified problem. These individual clarifications are signified as 'chromosomes', which protect the complete variety of probable results.

STAGE 2: SELECTION OPERATOR: Selection is the method of assessing the aptness of the individuals and picking them for imitation. There are various methods to achieve selection. Some commonly used methods include Elitist Selection, Hierarchical Selection, Rank Selection, Roulette-Wheel Selection, and Tournament Selection. This work has used Tournament selection to select sufficiently good individuals for mating.

STAGE 3: CROSSOVER OPERATOR: Crossover operator creates two offspring from the two selected parent chromosomes by exchanging part of their genomes. Crossover is the process of extracting the best genes from parents and reassembling them into potentially superior offspring. The modest practice of crossover is called a Single-point crossover. And the further types are Two-Point Crossover, Uniform crossover. This work has used a single-point crossover.

STAGE 4: MUTATION OPERATOR: Mutation preserves a genetic assortment of the population from one generation to DNAs to the next and upsurges the view of the algorithm to produce more fit individuals. Using a small mutation probability, at each position in the string, a character at this position is changed randomly. Mutation of bit strings flips the bits at arbitrary points with a small prospect. This work has used uniform mutation.

STAGE 5: BELIEF SPACE: This belief space assurance the finest fit members are conceded on to the next generation. The pre-eminent individual or a set proportion of suitable members endures to the next generation. Normative Knowledge, Domain-Specific Knowledge, Situational Knowledge, Temporal Knowledge, and Spatial Knowledge are the types of Belief space in CA. Besides, Normative knowledge belief space has also employed in this work. Influence function and acceptance function is utilized for accepting the belief space among the other generation.

STAGE 6: TERMINATION CONDITION: Three possible termination standards could be deployed for the CA: A

substantial result has been achieved, a predefined determined number of generations has been stretched, the population has transformed to a definite level of genetic discrepancy. The algorithm convergence is delicate to the mutation prospect: a very high transformation rate inhibits the exploration from congregating, however very low rate effects in untimely conjunction of the search. The termination standards for this work is the maximum number of generations = 20 to 50.

STEP BY STEP PROCEDURE FOR CULTURAL ALGORITHM BASED FEATURE SELECTION METHOD

Step 1: Measure Symmetrical Uncertainty of individual features from the dataset.

Step 2: Rank the features in the dataset according to their importance: $F = (f_1 > f_2 > f_3 \ldots)$

Input: Top m features set f_r and class label C.

   Output: S

   Step 3: S← null

   Step 4: Procedure CA

   Input: PopSize Ps, GenSize, GenomeLength N, ProbMutation Pm

   Output: The best individuals in all generations.

   Step 4.1: Population Initialization: Ps *N.

   Step 4.2: Retain f_1 from f_r

   Step 4.3: Ps← Random binary chromosome

   Step 4.4: for each chromosome do

   Step 4.5: compute fitness according to classifier ANN.

   Step 4.6: end for

   Step 4.7: repeat

   Step 4.8: Select parents p1,p2 from the population based on the fitness

   Step 4.9: for all new children do

   Step 4.10: retain f_1 from f_r

   Step 4.11: Crossover p1,p2;

   Step 4.12: Mutate each gene in new child chromosome with probability Pm;

   Step 4.13: Update knowledge space with influence and acceptance function

   Step 4.14: end for

   Step 4.15: Evaluate the fitness of new individuals according to the classifier.

   Step 4.16: Replace the least-fit population with new best individuals

   Step 4.17: Stopping criteria

Step 5: End Procedure

5 RESULT AND DISCUSSION

5.1 DESCRIPTION OF THE DATASET
The autistic child dataset was composed of several special schools in and around Chennai. ASD composed of 75 features that are positioned to classify the ASD people.

5.2 NUMBER OF FEATURES OBTAINED
Table 1 represents the number of features attained by the existing feature selection methods like Cultural Algorithm, Symmetrical Uncertainty. And recommended Cultural Algorithm based Feature Selection technique. The novel dataset collected of 75 characteristics, IG feature selection methods give 51 features, SU gives 49, GR gives 52, the optimization methods like PSO gives 48, ACO gives 49, CA

gives 46 ,SSU-PSO gives 47 [22]and the proposed CA-FS

Table 1: *Number of features obtained by the proposed CA-FS method, SU and CA feature selection techniques*

Sl.No	FS Methods	Features Obtained
1	Original Dataset	75
2	Symmetrical Uncert ainty	49
3	Information Gain	51
4	Gain Ratio	52
5	Particle Swarm Optimization	48
6	Ant Colony Optimization	49
7	Cultural Algorithm	46
8	Proposed CA-FS Method	42

5.3   PERFORMANCE ANALYSIS OF THE FEATURE SELECTION METHODS USING CLASSIFICATION METHODS

method gives only 42.

This present research explores the classification methods like Artificial Neural Network and Support Vector Machine is measured for assessing the novel dataset and feature selection methods like GR, IG, SU, ACO, PSO and proposed CA-FS technique. The performance metrics like Accuracy, Sensitivity or Recall, False Positive Rate ,Specificity, Precision, False Negative Rate, F-Measure and the error metrics like Mean Squared Error (RMSE), Mean Absolute Error (MAE), Root Relative Squared Error (RRSE) and Root Relative Absolute Error (RAE) are used .Table 2 contributes to the performance metrics measured in this paper.

Table 2: *Performance Analysis*

Feature Selection Methods	Classification Techniques									
	ANN					SVM				
	Accuracy	TPR	TNR	Precision	F-Measure	Accuracy	TPR	TNR	Precision	F-Measure
Original Dataset	85.31	0.5672	0.6104	0.5578	0.5535	55.62	0.4589	0.5488	0.4589	0.4213
Symmetrical Uncertainty	87.67	0.6012	0.7187	0.6812	0.6387	51.67	0.5412	0.6926	0.5412	0.5131
Information Gain	70.54	0.5714	0.7029	0.6728	0.6179	49.65	0.5317	0.6787	0.5317	0.5133
Gain Ratio	71.28	0.5832	0.7104	0.6673	0.6223	51.96	0.5478	0.6627	0.5478	0.5234
PSO	69.89	0.5897	0.7064	0.6542	0.6201	49.63	0.5318	0.6522	0.5318	0.5758
ACO	70.89	0.5613	0.6926	0.6448	0.6001	52.98	0.5271	0.6409	0.5271	0.5160
CA	88.75	0.6289	0.7883	0.7039	0.6642	61.32	0.5976	0.7166	0.5976	0.5553
CA-FS	89.41	0.8914	0.8184	0.8952	0.8932	69.25	0.7589	0.7424	0.7589	0.6957

6.CONCLUSION

Autism Spectrum Disorder (ASD) is one of the growing neurodevelopment conditions worldwide with many individuals undetected, making early screening crucial for individuals, family members and physicians. Most of the existing ASD methods consist of a large set of questions covering communication, social and repetitive behaviours and rely on domain expert rules with a basic scoring function to detect autistic traits. Before going for classification, feature selection is most important step in reducing the overfitting, improving the accuracy, and to reduces the training time. In this research work, Cultural Algorithm based Feature Selection (CA-FS) is proposed to select the most relevant features to the predictive model. The proposed CA-FS method gives least number of features than the existing other feature selection techniques. The evaluation of the CA-FS method is done with the classification techniques ANN and SVM. From the result obtained, the proposed CA-FS method gives more accuracy, TPR, Precision, Specificity, and F-Measure with using ANN than the other techniques with ANN and SVM and also it gives the minimal FPR, FNR for proposed CA-FS with ANN.

REFERENCES

[1] Heinsfeld, Anibal Sólon, et al. "Identification of autism spectrum disorder using deep learning and the ABIDE dataset." NeuroImage: Clinical 17 (2018): 16-23.

[2] Kumazaki, Hirokazu, et al. "Relationship between odor identification and visual distractors in children with autism spectrum disorders." Journal of autism and developmental disorders 48.7 (2018): 2590-2592.

[3] Sudha, V. Pream, and M. S. Vijaya. "Machine Learning-Based Model for Identification of Syndromic Autism Spectrum Disorder." Integrated Intelligent Computing, Communication and Security. Springer, Singapore, 2019. 141-148.

[4] Andrews, Derek Sayre, et al. "Using pattern classification to identify brain imaging markers in autism spectrum disorder." Biomarkers in Psychiatry. Springer, Cham, 2018. 413-436.

[5] Bishop-Fitzpatrick, Lauren, et al. "Using machine learning to identify patterns of lifetime health problems in decedents with autism spectrum disorder." Autism Research 11.8 (2018): 1120-1128.

[6] Rotholz, David A., et al. "Improving early identification and intervention for children at risk for autism spectrum disorder." Pediatrics 139.2 (2017): e20161061.

[7] Brown, Hilary K., et al. "Association between serotonergic antidepressant use during pregnancy and autism spectrum disorder in children." Jama 317.15 (2017): 1544-1552.

[8] Adelman, Chana R., and Thomas Kubiszyn. "Factors that affect age of identification of children with an autism spectrum disorder." Journal of Early Intervention 39.1 (2017): 18-32.

[9] Howsmon, Daniel P., et al. "Classification and adaptive behavior prediction of children with autism spectrum disorder based upon multivariate data analysis of markers of oxidative stress and DNA methylation." PLoS computational biology 13.3 (2017): e1005385.

[10] Lord, Catherine, et al. "Autism spectrum disorder." The Lancet 392.10146 (2018): 508-520.

[11] Sudha, V. Pream, and M. S. Vijaya. "Machine Learning-Based Model for Identification of Syndromic Autism Spectrum Disorder." Integrated Intelligent Computing, Communication and Security, Springer, Singapore, 2019. 141-148.

[12] Al-Mejibli, Intisar Shadeed, et al. "Performance Evaluation of Kernels in Support Vector Machine." 2018 1st Annual International Conference on Information and Sciences (AiCIS). IEEE, 2018.

[13] Thabtah, Fadi, Firuz Kamalov, and Khairan Rajab. "A new computational intelligence approach to detect autistic features for autism screening." International journal of medical informatics 117 (2018): 112-124.

[14] Thabtah, Fadi, and David Peebles. "A new machine learning model based on induction of rules for autism detection." Health informatics journal (2019): 1460458218824711.

[15] Tejwani, Ravi, et al. "Autism classification using brain functional connectivity dynamics and machine learning." arXiv preprint arXiv:1712.08041 (2017).

[16] Wang, Mingliang, et al. "Low-Rank Representation for Multi-center Autism Spectrum Disorder Identification." International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2018.

[17] Li, Genyuan, Olivia Lee, and Herschel Rabitz. "High efficiency classification of children with autism spectrum disorder." PloS one 13.2 (2018): e0192867.

[18] Kira, Kenji, and Larry A. Rendell. "A practical approach to feature selection." Machine Learning Proceedings 1992. Morgan Kaufmann, 1992. 249-256.

[19] Singh, Bharat, Nidhi Kushwaha, and Om Prakash Vyas. "A feature subset selection technique for high dimensional data using symmetric uncertainty." Journal of Data Analysis and Information Processing 2.04 (2014): 95.

[20] Guo, Yi-nan, et al. "Interval multi-objective quantum-inspired cultural algorithms." Neural Computing and Applications 30.3 (2018): 709-722.

[21] Reynolds, Robert G. "The cultural algorithm: Culture on the edge of chaos." Culture on the edge of chaos. Springer, Cham, 2018. 1-11.

[22] Abitha R ,Mary Vennila S. "A Swarm Based Symmetrical Uncertainty Feature Selection Method For Autism Spectrum Disorders." 3rd International Conference on Inventive Systems and Control. IEEE,2019.