International Journal of Global Research Innovations & Technology (IJGRIT)
ISSN : 2583-8717, Impact Factor: 6.382, Volume 02, No. 02, April-June, 2024, pp 09-13

AN IMPLEMENTATION OF PCS-MIA (PEARSON CHI-SQUARE, MUTUAL INFORMATION AND ANOVA) MODEL TO DETERMINE THE AUTISM SPECTRUM DISORDERS

M.Ramesh*
C Vijayakumaran**

ABSTRACT

Autism Spectrum Disorders, a neurodevelopmental condition, have surged in prevalence among children. Research suggests that early identification and intervention therapies greatly contribute to favourable long-term outcomes. Detecting any ailment in its initial stages can be challenging, as an initial symptom might indicate various other potential conditions. Top of Form Identifying neurodevelopmental diseases presents an added challenge as their symptoms often evade standard tests, relying instead on behavioural observations for diagnosis. Autism, a prime example of such a disorder, may surface in children as early as one to two years old. Early intervention during this critical period can significantly enhance a patient's condition. The primary focus of the proposed project revolves around utilizing machine learning techniques for detecting and diagnosing Autism, especially when symptoms are incomplete or insufficient, with a specific emphasis on rural areas. Consequently, health assistants or less seasoned doctors often handle such cases. Numerous previously developed systems have utilized complete or partial symptom sets for Autism diagnosis, yet they frequently result in misidentification. The proposed framework employs machine learning techniques to detect and diagnose Autism symptoms in young children, specifically when only incomplete sets of symptoms are accessible.

Keywords: *Autism Spectrum Disorders, Neuro Developmental, Neuro-Psychiatric.*

Introduction

Autism can be diagnosed at any age, but symptoms typically appear within the first 24 months of life and evolve over time. Autism spectrum disorders cover a range of neurodevelopmental conditions with varying degrees and manifestations. These challenges usually surface in childhood and involve difficulties in social communication, interaction, and behavioural patterns such as restricted interests and repetitive behaviours. In a densely populated and predominantly youthful country like India, the prevalence of Autism Spectrum Disorder (ASD) has risen. However, there are inconsistencies in assessment methods. They analyze the strengths and limitations of each assessment tool and offer recommendations for their application in clinical practice.Artificial Intelligence is to enhance research in the field of social sciences, particularly in comprehending the intricate and widespread features of autism spectrum disorder. By employing machine learning techniques, researchers see substantial benefits in the objective evaluation of neuro-psychiatric conditions[1].

Machine Learning offers a novel approach to predict Autism Spectrum Disorder (ASD) in children at an early stage. Traditional methods of identifying autistic traits through screening studies are not only costly but also time-consuming. The World Health Organization (WHO) reports a significant rise in the number of ASD diagnoses. These children often struggle with social interaction, leading to delayed development in linguistic, cognitive, repetitive behavioural, speech, and nonverbal communication skills.

---

* Research Scholar, Department of Computing Technologies, SRM Institute of Science and Technology, Kattankulathur, Chennai, Tamilnadu, India.
** Associate Professor, Department of Computing Technologies, SRM Institute of Science and Technology, Kattankulathur, Chennai, Tamilnadu, India.

Literature Review

The IWOA-FRBC model is tested using three ASD benchmark datasets. By accurately simulating each dataset, the IWOA-FRBC model's integrity was proven. Authors says that child, adolescent, and adult datasets are 93%, 95%, and 94% accurate, respectively[2].Comparing accuracy, precision, and false-positive rate by identifying crucial characteristics within the toddler autism dataset. Machine Learning methodologies play a vital role in the analysis and diagnosis of ASD datasets, particularly in predicting ASD.[3]Anovel autism framework employed in deep learning to supplant existing scoring functions. CNNs swiftly categorize psychologically imbalanced data. [4]devised attribute selection and ranking procedures at various stages of ASD datasets, employing classifiers to examine transformed statistics and establish predictive functions. [11] Proposed a novel feature selection methodology to reduce the ASD dataset's size effectively. [12] Presented a CNN-based model for ASD recognition, demonstrating the highest accuracy among various Machine Learning techniques. [5]Variable psychoanalysis assesses feature-to-class correlations while diminishing feature-to-function correlations. [6]Recommended using LASSO for determining the most critical features in supervised machine learning.

The Proposed Methodology

The comprehensive architecture of the proposed Adaptive PCS-MIA Methodology is implemented in Fig.1.

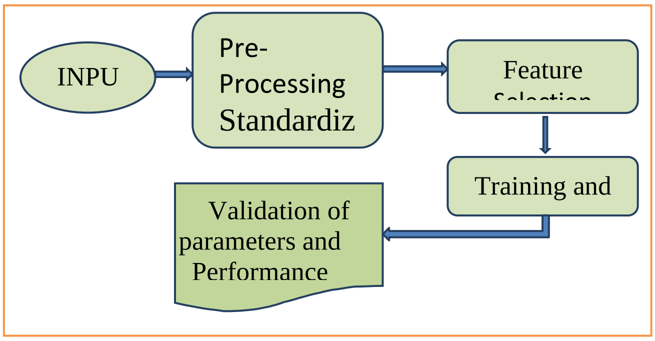

Fig. 1 Architecture of Proposed System

It consists of various stages, namely Pre-processing, Feature selection, Ranking, Feature extraction, Feature subset reduction, Training of Machine Learning (ML) classifiers, and Validation of ML classification accuracy.

- **Pre-Processing**

The data pre-processing stage encompasses a diverse set of activities, such as organizing data, removing outliers, addressing correlations, renovating data, and performing feature extraction for attribute selection.

- **Feature Selection**

Feature selection is a pivotal phase in data analysis and machine learning, focusing on the identification and prioritization of pertinent features or variables within a dataset. The objective is to enhance model performance, trim down dimensionality, and eliminate irrelevant or redundant information. Opting for the most informative features not only improves computational efficiency but also aids in mitigating overfitting risks and elevates the interpretability of the model.

  - Few ASD dataset attributes aid our assessment. Ethnicity, exam taker, and class/ASD features were eliminated for better analysis. Machine learning classifiers may have a goal variable.
  - Imputation of missing values**:** Imputation of missing values in autistic datasets involves replacing them with the label 'unknown' during the preprocessing phase. Following this data preparation step, our goal is to uncover insights that reveal challenges and propose solutions for knowledge classifiers.
  - One-hot encoding is employed to generate dummy features for each distinct category within nominal attributes. Typically utilized for converting non-ordinal categorical variables into numerical data, this method helps represent categorical information in a format suitable for numerical analysis.

- Partitioning ASD datasets into training and test sets, along with attribute selection, is essential for ensuring the reliability of machine learning models. The Scaler plays a vital role in this process by transforming features into a consistent scale, and in this particular system, it utilizes the Standard Scalar() for data scaling.
- The Pearson Correlation Coefficient (PCC) quantifies the strength of the relationship between variables by assessing the correlation between their values.

- **Feature Extraction**

Feature extraction is the process of converting data into a new feature subspace, and Principal Component Analysis (PCA) serves as a valuable technique for achieving this objective. By effectively reducing the dimensionality of intricate datasets, PCA enables the creation of a more concise representation within a lower-dimensional space.

- **Reducing Feature Subset**

Reducing the feature subset involves eliminating non-essential and redundant features, retaining only the relevant ones. The exclusion of irrelevant qualities from an individual's profile does not impact the learning process. "Redundant features" refer to duplicated attributes. In machine learning, the selection of attributes is determined by the nature of the problem being addressed. The minimized feature subset is utilized across various machine learning classifiers with diverse ASD datasets to ascertain the most pertinent features for predicting autistic symptoms and identify those that are less relevant. The outcomes reveal that the proposed technique outperforms existing algorithms in the selection of crucial attributes.

Feature Engineering with PCS-MIA Model

In this proposed adaptive model, the effective data pre-processing process encompasses a wide variety of activities such as data systematization, reducing outlier, correlation, and data renovation. Following this feature engineering with PCS-MIA model is classified into three phases.

The objective of feature selection is to eliminate features that do not contribute significantly to prognosis. Three main techniques for feature selection are filter, wrapper, and embedding. In this study, we employ wrapper-based feature selection methods to create a subset of features for early ASD prediction. The investigation focuses on wrapper techniques, including Sequential Forward Selection (SFS), Sequential Backward Selection (SBS), Sequential Backward Floating Selection (SBFS), Sequential Forward Floating Selection (SFFS), and Recursive Feature Elimination (RFE), along with optimal selection approaches based on classifiers such as Random Forest (RF), Gradient Boosting Classifier (GBC), and Classification And Regression Trees (CART). The RFE-based RF algorithm demonstrated superior performance over other search methods, achieving an average accuracy of 87%.

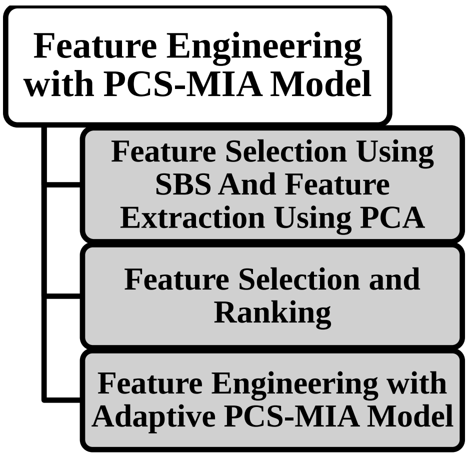

Fig. 2: Feature Engineering with PCS-MIA Model

- **Feature Selection Using SBSAnd Feature Extraction Using PCA**

The primary objective of this study is to achieve early detection of ASD in affected individuals. Feature engineering is employed to extract data features for predictive modelling. The proposed technique incorporates sequential backward selection (SBS) to diminish the dimensionality of an initial function subset, while Principal Component Analysis (PCA) is applied to reduce the complexity of an aggregated ASD dataset. The research scrutinizes and trims down features across three ASD datasets categorized by age. Support Vector Machine (SVM), Random Forest (RF), and K-Nearest Neighbours (KNN) are utilized for classifying the reduced feature set. The overall model performance is assessed based on accuracy and sensitivity, revealing that Random Forest exhibits a more precise categorization of ASD data.

- **Feature Selection and Ranking**

The authors propose a novel hybrid approach for selecting features, the Feature-RF-GBC-RFE model, incorporating Random Forest (RF) and Gradient Boosting Classifier (GBC). Feature selection involves identifying a subset of the dataset's most relevant characteristics. This study analyzes and trims down features in an ASD dataset categorized by age. The reduced feature set is then evaluated using Random Forest, Support Vector Machine (SVM), Gradient Boosting Classifier, AdaBoost, and the hybrid RF-GBC-RFE feature selection method. Model accuracy and sensitivity are used to predict overall performance. The hybrid RF-GBC-RFE feature selection strategy demonstrates more accurate data classification. The structure of the hybrid RF-GBC-RFE feature selection is illustrated in Fig. 2, and Table 3 presents the most significant features in the reduced final feature set

- **Feature Engineering with Adaptive PCS-MIA Model**

The study introduces an adaptive PCS-MIA feature engineering model designed to assist doctors, psychologists, and learning disability mentors in analyzing autism. The model incorporates effective data preprocessing and dimensionality reduction techniques, utilizing feature selection methods (Filter + wrapper-based) and extraction methods such as PCA with Random Forest to identify a subset of significant features based on their score and ranking. Through meticulous feature selection, the model demonstrates the capability to predict early signs of autism. The research involves the analysis of both real-time and four standard autism datasets. The findings indicate that the proposed technique can identify highly diverse traits, and the Matthews correlation coefficient (MCC) consistently yields higher scores compared to Cohen's kappa accuracy.

Performance Evaluation

Recall signifies the percentage of relevant data retrieved, while precision is the ratio of correct predictions to the total predictions. The F1 score is the weighted harmonic mean of accuracy and recall. Accuracy is the percentage of correctly categorized data over the total. The error rate is obtained by subtracting accuracy from 100%. MCC is a binary classification rate that performs well when the binary predictor accurately predicts most positive and negative occurrences in the training dataset. Cohen's Kappa measures agreement between two observers ranking the same group on a two-class scale. It's noted that MCC is considered more accurate than Cohen's kappa in this context.

Table 1: Performance Metrics for Classifying Data

F1	$\mathbf{2} * \frac{\boldsymbol{P} * \boldsymbol{R}}{\boldsymbol{P} + \boldsymbol{R}}$
Precision(P)	$\frac{tp}{tp + fp}$
Recall(R)	$\frac{tp}{fn + tp}$
ERR	$\frac{\mathrm{fp} + \mathrm{fn}}{fp + fn + tp + tn}$
Acc	1-ERR
MCC	$\frac{tp * tn - fp * fn}{\sqrt{(tp + fp)(tp + fn)(tn + fp)(tn + fn)}}$ Worst case=-1; Best case =1
Cohen's Kappa	$\frac{2 * (tp.tn - fp.fn)}{(tp + fp)(fp + tn) + (tp + fn)(fn + tn)}$

Conclusion and Future Work

In drawing clinical conclusions, the application of machine learning represents a significant leap forward, utilizing readily available information as a tool for development and progress. The Adaptive PCS-MIA model surpasses previous machine learning techniques and successfully categorizes five autism datasets. The proposed adaptive model incorporates exploratory data analysis (EDA), filter and wrapper selection, and a scoring and rating feature subset extraction method based on Random Forest (RF). Notably, the Matthews correlation coefficient (MCC) emerges as a more reliable metric compared to F1 and Cohen's kappa.

Simulation results demonstrate the accuracy of the PCS-MIA model across various age groups and real-time datasets, with accuracy rates of 98%, 96%, 100%, and 97% for toddlers, children, adolescents, and adults, respectively. The study highlights the efficacy of the adaptive PCS-MIA model in predicting autism spectrum disorder, emphasizing that infants with cognition deficits, impaired sociability, and jaundice exhibit a higher risk. Challenges in understanding ASD in developing nations are attributed to cultural factors, while limited resources lead to prolonged untreated periods for ASD patients. The adaptive PCS-MIA model proves valuable in predicting autism early and facilitating intervention before abnormal behavioural and cognitive patterns manifest. These findings contribute to enhancing the access of individuals with ASD to emotional support networks.

References

1. K. Umamaheswari and L. Parthiban, "IMPROVED WHALE OPTIMIZATION ALGORITHM BASED FEATURE SELECTION WITH FUZZY RULE BASE CLASSIFIER FOR AUTISM SPECTRUM DISORDER DIAGNOSIS," *Int. J. Adv. Res. Eng. Technol.*, vol. 11, no. 11, pp. 41–55, 2020, doi: 10.34218/IJARET.11.11.2020.005.
2. S. R. Shahamiri and F. Thabtah, "Autism AI: a New Autism Screening System Based on Artificial Intelligence," *Cognit. Comput.*, vol. 12, no. 4, pp. 766–777, Jul. 2020, doi: 10.1007/s12559-020-09743-3.
3. C. Radhika and N. Priya, "Prediction Of Learning Disability Of The Children Using Adaptive Effective Feature Engineering Techniques."
4. M. Li *et al.*, "An automated assessment framework for atypical prosody and stereotyped idiosyncratic phrases related to autism spectrum disorder," *Comput. Speech Lang.*, vol. 56, pp. 80–94, Jul. 2019, doi: 10.1016/j.csl.2018.11.002. "13." .
5. A. A. Abdullah, S. Rijal, and S. R. Dash, "Evaluation on Machine Learning Algorithms for Classification of Autism Spectrum Disorder (ASD)," *J. Phys. Conf. Ser.*, vol. 1372, no. 1, 2019, doi: 10.1088/1742-6596/1372/1/012052.
6. S. Raj and S. Masood, "Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques," *Procedia Comput. Sci.*, vol. 167, no. 2019, pp. 994–1004, 2020, doi: 10.1016/j.procs.2020.03.399.
7. S. Padmapriya and S. Murugan, "A Novel Feature Selection Method for Pre-Processing the ASD Dataset."
8. T. Akter *et al.*, "Machine Learning-Based Models for Early Stage Detection of Autism Spectrum Disorders," *IEEE Access*, vol. 7, pp. 166509–166527, 2019, doi: 10.1109/ACCESS.2019.2952609.
9. K. Vakadkar, D. Purkayastha, and D. Krishnan, "Detection of Autism Spectrum Disorder in Children Using Machine Learning Techniques," *SN Comput. Sci.*, vol. 2, no. 5, Sep. 2021, doi: 10.1007/s42979-021-00776-5.
10. K. S. Oma, P. Mondal, N. S. Khan, M. R. K. Rizvi, and M. N. Islam, "A Machine Learning Approach to Predict Autism Spectrum Disorder," in *2nd International Conference on Electrical, Computer and Communication Engineering, ECCE 2019*, 2019, doi: 10.1109/ECACE.2019.8679454.
11. D. Bone, M. S. Goodwin, M. P. Black, C. C. Lee, K. Audhkhasi, and S. Narayanan, "Applying Machine Learning to Facilitate Autism Diagnostics: Pitfalls and Promises," *J. Autism Dev. Disord.*, vol. 45, no. 5, pp. 1121–1136, May 2015, doi: 10.1007/s10803-014-2268-6.
12. J. O. Healthcare Engineering, "Retracted: Prediction and Analysis of Autism Spectrum Disorder Using Machine Learning Techniques," *Journal of healthcare engineering*, vol. 2023. p. 9815989, 2023, doi: 10.1155/2023/9815989.
13. C. Radhika1 and N. Priya2, "Advance Hybrid RF-GBC-RFE Wrapper-Based Feature Selection Techniques for Prediction of Autistic Disorder," *J. Algebr. Stat.*, vol. 13, no. 1, pp. 503–510, 2022.