INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

Enco – Standardization Data Pre-Processing Technique in Autism Spectrum Disorder Detection

M Kavitha[1]*, **M Kasthuri**[2]

1 Research Scholar/Assistant Professor, Department of Computer Applications, Bishop Heber College, Affiliated to Bharathidasan University, Trichirappalli, 620024, Tamil Nadu, India
2 Associate Professor, Department of Computer Applications, Bishop Heber College, Affiliated to Bharathidasan University, Trichirappalli, 620024, Tamil Nadu, India

OPEN ACCESS

Received: 15-07-2023

Accepted: 08-10-2023

Published: 05-12-2023

Citation: Kavitha M, Kasthuri M (2023) Enco – Standardization Data Pre-Processing Technique in Autism Spectrum Disorder Detection. Indian Journal of Science and Technology 16(45): 4156-4163. https://doi.org/10.17485/IJST/v16i45.1781

* **Corresponding author**.

kavithamuthuvel2@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Background/Objectives: The goal of this study was to create an Enco-Standardization technique that would produce accurate data and improve the diagnosis of Autism Spectrum Disorder (ASD).This method uses mean values to replace missing values in a dataset and improves them by combining label encoding and conventional scaling techniques. **Methods:** The ASD dataset, which has 704 instances and 21 attributes, is used in this study. Training and testing are divided by the dataset (80%-20%). As an imputation strategy in this dataset, missing values are located and replaced with the mean value. Attributes are encoded using the Enco-Standardization methodology using a label encoding technique that changes non-numeric variables into numeric ones. After that, the data were scaled into a machine-readable format to standardise it. Different machine learning classifier models are compared to the hybrid strategy of encoding and scaling techniques. Based on the accuracy found using machine learning classifier models, the dataset acquired using the Enco-Standardization technique is assessed**. Findings:** The dataset needs to be accurate and relevant in order to increase accuracy and decrease computing time. The findings of the Enco-Standardization methodology showed a good pre-processing method with accuracy values of 98% for Naive Bayes (NB), 71% for K Nearest Neighbour (KNN), 74% for Support Vector Machine (SVM), 97% for Linear Regression (LR), 100% for Decision Tree (DT), and 100% for Random Forest (RF). The deletion of missing values improves performance in KNN (94%), SVM (95.9%), LR, DT, and RF (100%) but decreases the number of instances in the dataset, rendering the model ineffective. **Novelty:** The data in a dataset are transformed and encoded using the proposed Enco-Standardization pre-processing technique, which increases the precision of the data analysis process in ASD prediction. Data discrepancies are avoided by using this eco-standardization technique.

Keywords: Autism Spectrum Disorder; Preprocessing; Scaling; EncoStandardization; Machine Learning

1 Introduction

A neurological condition known as autism spectrum disorder (ASD) is brought on by hereditary behaviour, environmental circumstances, and developmental issues associated to brain development processes. According to the definition of autism, it is "a set of behaviours which affects a person's personality when interacting and communicating with others." According to the Centres for Disease Control, there were 1 in 44 Americans with autism in the United States in 2018. The Autism Society estimates that 3.5 million Americans live with autism[(1)]. The behaviour of youngsters between the ages of 3-5 years old is indicative of ASD. In India, there are 1% to 2% estimated new cases of autism each year. Every social and racial group of people is impacted by autism, which affects 1 in 127 girls and 1 in 27 males. According to their IQ scores, 31% of kids have intellectual disabilities. Around 18 million people in India are thought to be affected by ASD, making it the third most prevalent developmental illness.

Federated learning techniques[(2)] combined with machine learning models such as Logistic Regression and Support Vector Machine used to predict autism in children and adults with the accuracy of 98% and 81% respectively. A centralized framework for autism disorder detection[(3)] proposed and gives 89.23% of accuracy with the Random Forest classification model. A multi-classifier recommended system[(4)] proposed in the decision making process of autism detection. This system yields maximum accuracy in the Decision Tree and Random Forest model. A healthcare system[(5)] proposed based on electronic health records and clinical data to diagnose autism in an earlier stage.

Twice-Growth Deep Neural Network (2GDNN), a reliable architecture, is used to aid in clinical diagnosis for the detection of diabetes[(6)]. This framework employs polynomial regression for missing value imputation and spearman correlation for feature selection. The suggested model exhibits 100% accuracy in the PIMA diabetes Indian dataset when compared to Support Vector Machine and Random Forest models. For the purpose of identifying EEG patterns that can be used to predict ASD and Neuro-Psychiatric Disorder (NPD), a Manhattan Distance-based Preprocessing approach was presented[(7)]. In order to build the Minimum Spanning Tree in this study, EEG data were converted into a triangular matrix. K-Nearest Neighbour (KNN), a well-known machine learning technique, predicts ASD and NPD with a 93.2% accuracy rate.

Traditional methods are being supplemented with machine learning techniques to improve diagnosis accuracy and speed[(8)]. We employed models like Nave Bayes (NB), Logistic Regression (LR), Support Vector Machines (SVM), Random Forest Classifier (RFC), and KNN to assess our dataset. These models base their predictions on data. analysing the dataset of adults[(9)] and kids with ASD using the practical component analysis approach. The methodology that is provided is composed of three main steps: Creating the data set, analysing the data, and unsupervised categorization are the first three steps. Using the results of the investigation, ASD was classified in both adults and children. The classification's outcomes for adults have a specificity of 95.7% and a sensitivity of 87.5%. Even data that was gathered separately can be combined using the methodology. In addition, by looking at the ethnic breakdown, it is clear[(10)] that white Europeans had a higher prevalence of ASD than other ethnic groups, with Asia following in second. This methodology expands the data set, which can help researchers identify more links among various measures. It is clear that those who were born with jaundice had a significantly lower rate of ASD compared to those who did not have the illness. Using MRI data, various machine learning and deep learning techniques for ASD and Attention Deficit or Hyperactivity Disorder were described[(11)]. They demonstrate that Convolutional Neural Network (CNN) performs better in Deep Learning.

A machine learning method for predicting ASD in kids is explored. A dataset is acquired from Kaggle for this essay. There are many different kinds of features in the dataset. Non-contributed features are eliminated during pre-processing, and multiclass and categorical features are encoding with label encoding and one hot encoding, respectively. The top three characteristics were used to produce the autism index, and SVM[(12)] polynomials 1, 2, and 3 were employed to categorise the ranking feature set. The SVM polynomial 2 generated the highest classification accuracy of 98.70% with 20 features. Imputation of missing data and a standard scalar are both employed in this paper. The missing values are imputed using the Euclidean Distance. The accuracy of the Random Forest model is 99.35%.

Automated detection of the autism spectrum disorder (ASD) using a brain imaging dataset and CNN. We were able to identify ASD patients (ABIDE) using the most common resting-state functional magnetic resonance imaging (fMRI) data made available by a multi-site database called autism brain imaging data exchange, and the suggested method for diagnosing autism was successful when comparing[(13)] the functional connection patterns of people with autism to those of normal controls. The accuracy[(14)] of diagnosis using SVM, RF, KNN, and ANN classifiers was, respectively, 86.29%, 71.15%, 86.53%, and 88.46%. The highest degree of diagnosis accuracy was provided by ANN.

A multilayer perceptron (MLP) based classification model with auto encoder pretraining was utilised to distinguish ASD from Typically Developing (TD) using MRI images from the ABIDE-1 dataset[(15)]. Pre-trained deep convolutional neural networks (CNNs) including GoogleNet, AlexNet, MobileNet, and SqueezeNet achieved validation accuracy of 75%, 75.84%, 79.45%, and 82.98% in detecting the scalograms generated by EEG signals[(16)]. Mean value utilised[(17)] for step forward and backward feature selection, missing value imputation, and feature selection. The RF and SVM models are compared to the pre-processed and chosen features. In comparison to SVM, the accuracy of the RF model is 83% higher. ASD is detected by machine learning algorithms after analysis [18]. In this study, mean is used to impute missing values. The pre-processed dataset was examined using SVM, and the results strongly imply CNN is superior to SVM, with accuracy rates for children and adolescents of 98.3% and 96.88%, respectively.

One-hot encoding is a useful preprocessing method [19] that combines feature selection with Principal Component Analysis. Scaling and dividing carried out through standardisation. The accuracy of the suggested pre-processing method using the Random Forest (RF) model for ASD diagnosis is 92%. A web interface was suggested [20] to identify ASD in children between the ages of 1 and 5. With the help of pre-processing procedures, the data is cleaned and altered. The online interface was put up against the DT, LR, SVM, and RF models. The DT model had a build time of 0.014 seconds and displays 100% accuracy.

1.1 Research Gap

The current state of research predominantly centers on specific types of data, such as EEG signals, MRI images, and functional MRI data. Nonetheless, there exists an opportunity to broaden the perspective by including additional dimensions. While certain studies have briefly touched upon feature selection and preprocessing techniques, there remains a potential to integrate clinical and demographic factors, such as age, gender, and genetic information, into predictive models. The incorporation of these factors could potentially elevate the accuracy of diagnoses by providing a more individualized and comprehensive evaluation.

To bridge this gap in research, it is crucial to explore integrative approaches that harness the potential of multiple data sources. This necessitates proactive innovation in areas like feature extraction, data imputation, and normalization techniques. These efforts hold the promise of unveiling novel insights and driving enhancements in prediction accuracy and dependability. Ultimately, such progress will play a pivotal role in deepening our understanding of Autism Spectrum Disorder, advancing the field with refined precision and insight.

An individual approach cannot be relied upon to consistently provide pertinent features and cannot be predicted. Depending on the application and dataset utilised in the analysis, it could change. Hybrid or ensemble approaches can be employed in addition to individual methods to forecast accurate data. In this paper, we apply the single pre-processing procedures of standardisation and normalisation. To obtain important features, more than two pre-processing techniques are employed. It might be beneficial for future prediction processes. Relevant characteristics cannot be provided by the feature Selection, Single approach alone. The mechanism may change depending on the application that is extracted from the feature and the dataset that has to be inserted. As a result, the hybrid or ensemble approaches provide important features over the particular method. Standardisation and normalisation techniques are being employed for feature selection.

1.2 Feature of this paper

1. Missing values in the dataset make decision predictions less accurate. Thus, in Anaconda Python, missing values are found using the isna() or isnull() methods.
2. Dropping missing data could result in fewer occurrences. As a result, it can be imputated utilising approaches such replacing with the mean, median, mode, or random numbers. Use the mean imputing method if appropriate.
3. Use label encoding to encode data so that machine readers can interpret labels that aren't numerical as numbers.
4. Standardise the dataset using the Standard Scalar() function to ensure that the input variables' capabilities are consistent with machine format.
5. Use ML classifier models to analyse the Enco- Standardization approach.
6. Evaluate the effectiveness and match the best model to the dataset.

This article is organized as follows. An overview of the study is given in Section 1. Currently used pre-processing methods are reviewed briefly in the literature. The procedures followed for this study are described in Section 2. This section also covers the implementation of the Enco- Standardization algorithm, its design and approach, and performance assessment. In section 3, findings and analyses are presented. In section 4, a brief conclusion is given.

2 Methodology

Data cleansing and data integration are two categories of data pre-processing. By eliminating or discarding missing values and substituting mean, median, or mode for those values, data cleaning aids in the creation of a full dataset. Although dropping can provide positive outcomes, it also reduces the number of instances in a dataset. The workflow model for the suggested technique is shown in Figure 2. The absence of a value indicates that an attribute in an instance does not have any value. The imputation techniques are applied, depending upon the percentage of missing values presented in the dataset. The procedure to select missing value imputation technique as follows: IF (No of categories in Variable <=25%)

```
IF (No of categories in Variable <=25%)
   IF (% of missing value <= 50%)
       Impute with Similar or Assign new values
    ELSE
       Create a Dummy variable (Missing→ 1 Non Missing→0)
ELSE
   IF (% of missing value <= 10%)
       Impute with Mean or Median values
   ELSE IF (% of missing value >10% && <=50%)
       Assign new values and create indicator variable (Missing→ 1 Non
Missing→0)
    ELSE
       Create a Dummy variable (Missing→ 1 Non Missing→0)
```

Fig 1.

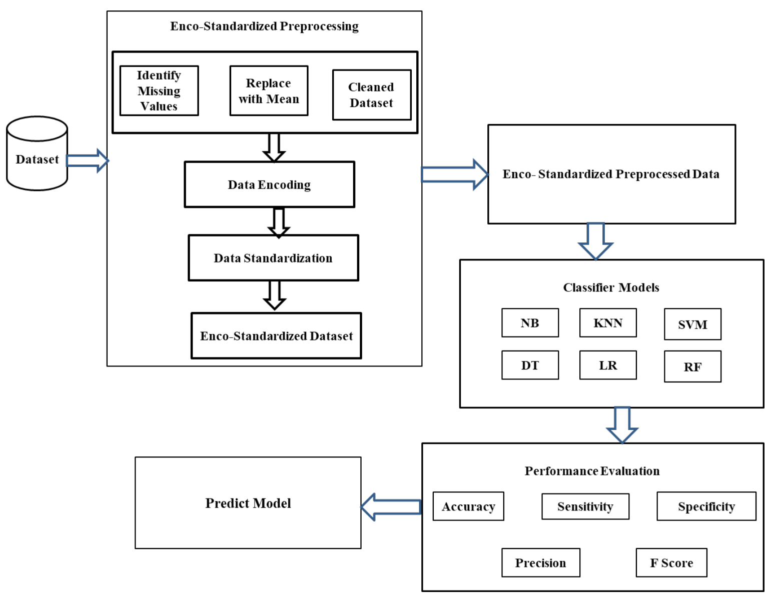

Fig 2. Workflow Model of Enco- Standardization Technique

After finding the missing values encoding and scaling takes place. Standardisation and normalisation scaling techniques are used to define the range to all numerical ranges from -1.0 to +1.0 or 0.0 to 1.0. Equation 1 in standardisation represents scaling.

$$X\_New = \frac{X - Mean}{\sigma} \tag{1}$$

Where X $\rightarrow$ Dataset values,

σ (Standard deviation) $= \sqrt{\frac{\Sigma\left(x_i - \bar{x}\right)^2}{N}}$, $\bar{x} \rightarrow$ Mean.

In normalization, scaling is represented in equation 2.

$$X\_New = \frac{X - X\_Min}{X\_Max - X\_Min} \tag{2}$$

The effectiveness of a suggested Enco- Standardisation technique is assessed using classifier models from NB, KNN, SVM, LR, DT, and RF.

The UCI repository (https://archive.ics.uci.edu/ml/machine-learningdatabases/00426/) is where the ASD Dataset may be obtained. It has 704 instances and 21 characteristics. Based on contextual factors rated by outcome, the class ASD feature in this dataset determines whether or not an individual has autism. Table 1 displays a description of the dataset's features. This work is implemented in an Anaconda Python Jupyter Notebook.

Table 1. Dataset Details

Feature Name	Feature Type	Feature Description	Missing Values
Age	Number	Age in years	Yes
Gender	String	Male or Female	No
Ethnicity	String	List of communal ethnicity	Yes
Jundice	Boolean	Born with Jundice or not	No
Autism	Boolean	Family history of autism	No
Country of residence	String	Name of the country	No
Used App before	Boolean	Yes or No	No
Result	Number	Score obtained	No
Age_Description	String	Child or Adult	Yes
Relation	String	Attender (Parent/clinician/caretaker self)	Yes
Class_ASD	Boolean	Yes or No (Have Autism or Not)	No
A1_Score	Binary (0 or 1)	Does your child look at you when call his/her name?	No
A2_Score	Binary (0 or 1)	How easy is it for you to obtain eye contact with your child?	No
A3_Score	Binary (0 or 1)	Does your child point to specify that she/he needs something?	No
A4_Score	Binary (0 or 1)	Does your child point to share curiosity with you?	No
A5_Score	Binary (0 or 1)	Does your child pretend?	No
A6_Score	Binary (0 or 1)	Does your child go behind where you're looking	No
A7_Score	Binary (0 or 1)	When you or someone else in the family is noticeably upset, does your child show signs of counsel to comfort them?	No
A8_Score	Binary (0 or 1)	Your child's first words as:	No
A9_Score	Binary (0 or 1)	Does your child use simple gestures?	No
A10_Score	Binary (0 or 1)	Does your child gaze at nothing with no apparent purpose?	No

3 Results and Discussion

Performance Metrics: Predicting the classifier model is a useful metric. Confusion Matrix is a 2X2 matrix that uses matrix cells to indicate True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) values.

TP $\rightarrow$ predicted YES with actual YES

TN $\rightarrow$ predicted NO with actual NO

FP $\rightarrow$ predicted YES with actual NO

FN $\rightarrow$predicted NO with actual YES

Accuracy is used to represent the total number of correct predictions and calculated in equation (3)

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

Sensitivity is used to represent actual yes in correctly classified samples and calculated in equation (4)

$$Sensitivity = \frac{TP}{TP + FN} \tag{4}$$

```
Input : ASD_Dataset
Output : Enco − Standardized ASD_Dataset
Enco − Standardization
{
df = ASD_Dataset                        // Load Dataset
IF(df.isnull.sum() ≠ 0)                 // Identify missing values
{
   IF(isnan(df.values.any()))           // Identify Non-Negative numbers
    {
      Mean = (x1+x2+......+xn)/N        // x1, x2....xn→ Feature Value in missing column
                                        // N→ Total Number of Instances
      le = LabelEncoder()               // Create an instance to label Encoder
      LabelEncoder()
       {
        le.fit(df[columns])             // Apply fit & Transform
        le.transform['DummyColumn_Names]
       }
      scalar = StandardScalar()
       {
        df_new = scalar.fit_transform(df)
       }
    }
}
Return df_new                           // Enco-Standardized ASD_Dataset
}
```

Fig 3. Enco – Standardization Algorithm

Specificity is used to represent actual no in correctly classified samples and calculated in equation (5)

$$Specificity = \frac{TN}{TN+FP} \tag{5}$$

Precision is used to represent positive cases in correctly classified samples and calculated in equation (6)

$$Precision = \frac{TP}{TP+FP} \tag{6}$$

F Score is calculated in equation (7)

$$F\ Score = \frac{2 * Precision * Sensitivity}{Precision + Sensitivity} \tag{7}$$

The accuracy and sensitivity scores for various machine learning classifier models are displayed in Table 2. Dropping values that are missing could result in fewer records and incorrect categorisation. This table demonstrates the accuracy of the suggested Enco-Standardization techniques: 98% for NB, 71% for KNN, 74% for SVM, 97% for LR, 100% for DT, and 100% for RF. In the table, sensitivity measures are displayed.

The specificity and precision values for various machine learning classifier models are displayed in Table 3.

According to this table, the proposed Enco-Standardization techniques demonstrate 100% accuracy in LR, 92% in DT, 100% in RF, 100% in NB, 42% in KNN, 67% in SVM, and 98% in NB. In the table, specificity metrics are displayed.

The F score value for classifier models is displayed in Table 4. This table demonstrates the effectiveness of the Enco-standardization technique: 99% in NB, 88% in KNN, 99% in SVM, 100% in LR, 97% in DT, and 100% in RF model.

The previous work [21] shows 84, 88, 86 % of accuracy with SVM, KNN and ANN classifier respectively using replace missing values by mean pre-processing techniques for diabetes prediction. [22] shows 92 % of accuracy with RF model for ASD and [23]

Table 2. Evaluation of an Imputation Techniques- Performance Metrics: Accuracy and Sensitivity

Classifier Model/ Techniques/ Performance Metrics	Accuracy in %			Sensitivity in %		
	Drop Missing Values	Mean + Label Encoding	Enco_ Standardization	Drop Missing Values	Mean + Label Encoding	Enco_ Standardization
Naïve Bayes	55.73	97	98	100	95	95
KNN	94	71	71	93	24	24
SVM	95.9	74	74	95	30	30
Linear Regression	100	95	97	100	100	100
Decision Tree	100	100	100	100	94	97
Random Forest	100	100	100	100	100	100

Table 3. Evaluation of an Imputation Techniques- Performance Metrics: Specificity and Precision

Classifier Model/ Techniques/ Performance Metrics	Specificity in %			Precision in %		
	Drop Missing Values	Mean + Label Encoding	Enco_ Standardization	Drop Missing Values	Mean + Label Encoding	Enco_ Standardization
Naïve Bayes	100	98	99	32	95	98
KNN	93	88	88	95	42	42
SVM	95	99	99	96	67	67
Linear Regression	100	100	100	100	100	100
Decision Tree	100	96	97	100	89	92
Random Forest	100	100	100	100	100	100

Table 4. Evaluation of an Imputation Techniques- Performance Metrics: F Score

Classifier Model/ Techniques/ Performance Metrics	F Score in %		
	Drop Missing Values	Mean + Label Encoding	Enco_ Standardization
Naïve Bayes	100	98	99
KNN	93	88	88
SVM	95	99	99
Linear Regression	100	100	100
Decision Tree	100	96	97
Random Forest	100	100	100

combines mean imputation techniques and feature ranking method to select features and shows almost 95.3% of accuracy with RF classifier in ASD. In [24] LDA pre-processing technique with KNN model and artificial algorithms for predicting ASD gives 88% of accuracy. In our proposed work the accuracy of pre-processing is evaluated using NB, KNN, SVM, LR, DT and RF classifier models. The results obtained and executed in Jupyter Notebook using Python and shown in Table 2.

4 Conclusion

ASD cannot currently be detected early on using automated screening technologies. We draw the conclusion that a hybrid approach can help to boost efficiency and outcome because using just one pre-processing technique may not always yield better outcomes. The suggested Enco-standardization pre-processing method has the advantages of data cleansing and standardisation for prediction. The Enco-standardization technique is examined in this study employing a variety of categories based on the behaviour of the child and relevant medical information. The trials' results show that, with accuracy rates of 97%, 100%, and 100%, respectively, linear regression, decision trees, and random forests models may produce the best outcomes. Considering all performance metrics, random forest is the model that performs the best in predicting ASD. The study focused primarily on a small number of characteristics rather than other potential variables that might influence the development or diagnosis of ASD. Pre-processed data can be coupled with future feature selection approaches to provide ever-more-relevant features for subsequent classification phases.

However, it is imperative to acknowledge the study's limitation in primarily focusing on a select set of characteristics, thereby potentially neglecting other influential variables pertaining to ASD development and diagnosis. To address this limitation, future endeavours can capitalize on the pre-processed data by integrating advanced feature selection methodologies, thus fortifying subsequent classification phases with even more pertinent attributes.

References

1) Tiwari R, Purkayastha K, Gulati S. Public Health Dimensions of Autism Spectrum Disorder in India: An Overview. *Journal of Comprehensive Health*. 2021;9(2):57–62. Available from: http://dx.doi.org/10.53553/jch.v09i02.002.
2) Farooq MS, Tehseen R, Sabir M, Atal Z. Detection of autism spectrum disorder (ASD) in children and adults using machine learning. *Scientific Reports*;13(1). Available from: http://dx.doi.org/10.1038/s41598-023-35910-1.
3) Qureshi MS, Qureshi MB, Asghar J, Alam F, Aljarbouh A. Prediction and Analysis of Autism Spectrum Disorder Using Machine Learning Techniques. *Journal of Healthcare Engineering*. 2023;2023:1–10. Available from: http://dx.doi.org/10.1155/2023/4853800.
4) Shinde AV, Patil DD. A Multi-Classifier-Based Recommender System for Early Autism Spectrum Disorder Detection using Machine Learning. *Healthcare Analytics*. 2023;4:100211. Available from: http://dx.doi.org/10.1016/j.health.2023.100211.
5) Chen J, Engelhard M, Henao R, Berchuck S, Eichner B, Perrin EM, et al. Enhancing early autism prediction based on electronic records using clinical narratives. *Journal of Biomedical Informatics*. 2023;144:104390. Available from: http://dx.doi.org/10.1016/j.jbi.2023.104390.
6) Olisah CC, Smith L, Smith M. Diabetes mellitus prediction and diagnosis from a data preprocessing and machine learning perspective. *Computer Methods and Programs in Biomedicine*. 2022;220:106773. Available from: http://dx.doi.org/10.1016/j.cmpb.2022.106773.
7) Grossi E, White R, Bs RJ, Swatzyna. A simple preprocessing method enhances machine learning application to EEG data for differential diagnosis of autism. *ResearchGate*. 2022. Available from: https://www.researchgate.net/publication/360034087.
8) Vakadkar K, Purkayastha D, Krishnan D. Detection of Autism Spectrum Disorder in Children Using Machine Learning Techniques. *SN Computer Science*. 2021;2(5). Available from: http://dx.doi.org/10.1007/s42979-021-00776-5.
9) Shihab AI, Dawood FA, Kashmar AH. Data Analysis and Classification of Autism Spectrum Disorder Using Principal Component Analysis. *Advances in Bioinformatics*. 2020;2020:1–8. Available from: http://dx.doi.org/10.1155/2020/3407907.
10) Peral J, Gil D, Rotbei S, Amador S, Guerrero M, Moradi H. A Machine Learning and Integration Based Architecture for Cognitive Disorder Detection Used for Early Autism Screening. *Electronics*. 2020;9(3):516. Available from: http://dx.doi.org/10.3390/electronics9030516.
11) Eslami T, Almuqhim F, Raiker JS, Saeed F. Machine Learning Methods for Diagnosing Autism Spectrum Disorder and Attention- Deficit/Hyperactivity Disorder Using Functional and Structural MRI: A Survey. *Frontiers in Neuroinformatics*. 2021;14. Available from: http://dx.doi.org/10.3389/fninf.2020.575999.
12) Oh SL, Jahmunah V, Arunkumar N, Abdulhay EW, Gururajan R, Adib N. A novel automated autism spectrum disorder detection system. *Complex Intell Syst*. 2021;7:2399–413. Available from: https://doi.org/10.1007/s40747-021-00408-8.
13) Sherkatghanad Z, Akhondzadeh M, Salari S, Zomorodi-Moghadam M, Abdar M, Acharya UR. Automated detection of autism spectrum disorder using a convolutional neural network. 2020. Available from: https://doi.org/10.3389/fnins.2019.01325.
14) Khadem-Reza ZK, Zare H. Automatic detection of autism spectrum disorder (ASD) in children using structural magnetic resonance imaging with machine vision system. *Middle East Current Psychiatry*. 2022;29(1). Available from: https://doi.org/10.1186/s43045-022-00220-1.
15) Prasad PKC, Khare Y, Dadi K, Vinod PK, Surampudi BR. Deep Learning Approach for Classification and Interpretation of Autism Spectrum Disorder. In: 2022 International Joint Conference on Neural Networks (IJCNN). IEEE. 2022;p. 1–8. Available from: https://doi.org/10.1109/IJCNN55064.2022.9892350.
16) Din QMU, Jayanthy AK. Automated classification of Autism Spectrum Disorder using EEG signals and Convolutional Neural Networks. *Biomedical Engineering: Applications, Basis and Communications*. 2022;34(02). Available from: https://doi.org/10.4015/S101623722250020X.
17) Sivaranjani S, Ananya S, Aravinth J, Karthika R. Diabetes Prediction using Machine Learning Algorithms with Feature Selection and Dimensionality Reduction. *7th International Conference on Advanced Computing and Communication Systems (ICACCS)*. 2021. Available from: https://doi.org/10.1109/ICACCS51430.2021.9441935.
18) Raj S, Masood S. Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques. *Procedia Computer Science*. 2020;167:994–1004. Available from: https://doi.org/10.1016/j.procs.2020.03.399.
19) Priya N, Radhika C. Effective Implementation of Pre-Processing Techniques in Machine Learning for Autism Spectrum Disorder. *International Journal of Innovative Technology and Exploring Engineering*. 2020;9(5):2253–2257. Available from: https://doi.org/10.35940/ijitee.E2676.039520.
20) Erkan U, Thanh DNH. Autism Spectrum Disorder Detection with Machine Learning Methods. *Current Psychiatry Research and Reviews*. 2020;15(4):297–308. Available from: http://dx.doi.org/10.2174/2666082215666191111121115.
21) Kaur H, Kumari V. Predictive modelling and analytics for diabetes using a machine learning approach. *Applied Computing and Informatics*. 2022;18(1/2):90–100. Available from: https://doi.org/10.1016/j.aci.2018.12.004.
22) Jacob SG, Sulaiman MMBA, Bennet B. Feature Signature Discovery for Autism Detection: An Automated Machine Learning Based Feature Ranking Framework. *Computational Intelligence and Neuroscience*. 2023;2023:1–14. Available from: https://doi.org/10.1155/2023/6330002.
23) Alkahtani H, Aldhyani THH, Alzahrani MY. Early Screening of Autism Spectrum Disorder Diagnoses of Children Using Artificial Intelligence. *Journal of Disability Research*. 2023;2(1). Available from: https://doi.org/10.57197/JDR-2023-0004.
24) Chen YH, Chen Q, Kong L, Liu G. Early detection of autism spectrum disorder in young children with machine learning using medical claims data. *BMJ Health Care Inform*. 2022;29(1):e100544. Available from: https://doi.org/10.1136/bmjhci-2022-100544.