WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Effective Machine Learning-Based Models for Autism Spectrum Disorder Early Detection

[1st] Krishnakumar.D, [4th] Mr. Thiruselvam.p ME(PhD),[2nd] Arunpandian.S, [3rd] RameshKannan.S

[1] dept. of Computer Science P.S.R Engineering College Sivakasi, India, e-mail: 20cs055@psr.edu.in
[4] dept. of Computer Science P.S.R Engineering College Sivakasi, India, e-mail: thiruselvan@psr.edu.in
[2] dept. of Computer Science P.S.R Engineering College Sivakasi, India, e-mail: 21lcs01@psr.edu.in
[3] dept. of Computer Science P.S.R Engineering College Sivakasi, India, e-mail: 21lcs05@psr.edu.in

ABSTRACT:

Although a cure for autism spectrum disorder (ASD) is not yet available, research indicates that early intervention in therapeutic interventions may be effective in mitigating symptoms. We gathered datasets containing information on autism spectrum disorder (ASD) classifications in infants, children, adolescents, and adults, and transformed this data using feature transformation techniques including logarithms, Z-scores, and sine functions. Following this, additional categorization techniques were evaluated utilizing the revised ASD datasets. Our findings revealed that the regression model yielded the most favorable results for the toddler dataset. In contrast, Adaboost, Glmboost, and Adaboost provided the most effective results for the children's dataset, adolescent dataset, and adult dataset, respectively. Modifications to the sine function and Z-score features yielded the most accurate classifications, respectively, for infancy and children and adolescent datasets. To ascertain the primary risk factors for Autism Spectrum Disorder (ASD) among toddlers, kids, adolescents, and adults, many feature selection techniques were applied to these Z-score converted datasets. The results of these analytical methodologies indicate that machine learning algorithms have the potential to generate dependable ASD status forecasts with appropriate optimization. Thus, the possibility that these models could be utilized for early detection of ASD is restored.

Keywords: Autism Spectrum Disorder, Adaboost, Glmboost

Introduction:

Autism Spectrum Disorder (ASD) is a neurological condition marked by significant deficits in communication skills that are crucial for everyday activities. Although most people with autism have mild problems, there are few cases of significant obstacles that need professional care. Individuals with Autism Spectrum Disorder (ASD) sometimes have difficulty in social situations due to their issues in interpersonal communication. The bulk of neurophysiological symptoms linked with Autism Spectrum Disorder (ASD) are known to medical practitioners. However, at now, there is no definitive biosignature or pathological approach available for the identification of autism [1]. Even without a clear treatment plan, identifying the illness at an early stage may significantly enhance the outlook. The implementation of suitable interventions in early infancy has the potential to enhance the probability of social skill improvement in kids with an autism spectrum disorder diagnosis (ASD), owing to the heightened neuroplasticity seen during this developmental stage. Based on empirical research, it has been shown that that children who receive medical attention before turning four years old have average IQ scores that are higher than those who wait until a later stage in life [2]. Despite these efforts, recent research estimates that the incidence of identifying children with Autism Spectrum Disorder (ASD) in the USA by their second birthday of three is around 34%. Nevertheless, it is worth noting that impoverished countries show a much-decreased percentage [3]. Currently, there is no established treatment procedure for Autism Spectrum Disorder (ASD). However, professionals have thoroughly analyzed several therapeutic options to mitigate symptoms, boost cognitive function, and improve activities of daily life. The early and precise identification of Autism Spectrum Disorder (ASD) is of utmost importance to effectively apply a range of intervention techniques. The Autism Diagnostic Observation Schedule (ADOS) and the Autism Diagnostic Interview-Revised (ADI-R) are widely recognized as the most reliable and authoritative interview-based diagnostic techniques in the field [4]. These methodologies primarily rely on the proficiency of the medical practitioners and the precision of the data provided by the patient's carers or guardians. The accuracy of these operations may be compromised due to human mistakes, despite their high level of reliability. Recent advancements in the industry have sparked the need to include machine learning in this advanced medical diagnostic system. The use of artificial intelligence has promise in improving the accuracy and efficiency of medical diagnosis by providing clinicians with valuable data and information that may assist in their decision-making process [5].

Related Works:

Recent research studies have demonstrated that deep learning-based approaches may be quite beneficial in diagnosing Autism Spectrum Disorder (ASD). Unlike interview-based methods, which are considered the most reliable, the incorporation of neuroimaging information is a freshly researched way of diagnosing ASD. Structured magnetic resonance imaging (MRI) is a kind of data from a neuroimaging modality, while functional neuroimaging involves the use of electroencephalography (EEG). Different types of deep neural networks, such as generative adversarial networks (GANs), autoencoders (AEs), recurrent neural networks (RNNs), and convolutional neural networks (CNNs), are trained using images [6]. Combining The accuracy and reliability of diagnosing ASD are improved when neuroimaging data from both modalities are combined with computational deep features [7]. Even though neuroimaging can provide a more precise and accurate diagnosis of autism spectrum disorder (ASD), getting the required images may be costly and time-consuming for some individuals. A supplementary method for diagnosing Autism Spectrum Disorder (ASD) involves using a dataset of behavioral characteristics, such as documenting distinct conduct by video recording [8], analyzing patterns of eye gaze [9], and then analyzing the patterns of speech [10], drawing [11], and similar techniques. Each of these methodologies, all of which are activated by behavioral data, requires a substantial time investment and several pre-processing steps. Alternatively, a very encouraging approach for detecting Autism Spectrum Disorder (ASD) involves using deep learning techniques to examine facial features [12]. This technology exhibits potential for enhanced objectivity and efficiency compared to current diagnostic approaches, since it circumvents the need for prolonged medical procedures, offers cost-effectiveness, and is devoid of human bias. Additional research is necessary to validate the efficacy of this approach, and its precision is now under scrutiny.

In medical imaging, this method is often used to get beyond the drawbacks of tiny, uneven, data sets with annotations [13]. Image modifications are a widely used technique for enhancing information in medical imaging. Creating new samples entails transforming the original picture using a variety of geometric and intensity techniques. Scaling, rotations, translations, and flips are a few frequently used picture transformations [14]. Data synthesization, which entails mixing or modifying current data to create new data samples, is another method of augmenting medical data. By exposing the model to changes in the data, these techniques may expand the dataset and strengthen the model's resilience [15]. Pre-processing and augmentation of datasets have shown to be successful, as evidenced by an increasing amount of research. For instance, researchers have shown that adding random changes to the training dataset may significantly increase accuracy and stability in the categorization of medical images. To identify brain tumors, for instance, 2020 Deepak et al. [16] used Data augmentation on MRI images resulted in a 6.7% increase in the CNN classifier's detection accuracy. Ju et al. 2021 [17] applied the generative adversarial network (CycleGAN) model on the UWF fundus image dataset. Following augmentation, it showed improvements in tessellated fundus segmentation, lesion identification, and diabetic retinopathy (DR) classification of 2.87% and 4.85%, respectively. By 3.2%, the prediction accuracy of COVID-19 pneumonia diagnosis was improved by D. Srivastav et al. in 2021. [18] by using a synthetic picture enhancement approach utilizing X-rays. Although there hasn't been much research on using augmented eye-tracking data to diagnose ASD using facial features, a recent study found that doing so increased prediction accuracy by 3% [19]. As a result, we want to look into the advantages of preliminary processing as well as enhancement to combine this dataset and determine whether performance matrices may be improved. In this work, we use a pre-trained CNN algorithm and a face picture dataset to test for ASD using a data-centric approach. We create research that is based on the manipulation of data using a data-centric methodology. Rather than concentrating on hyperparameter adjustment and models

Methodology:

We searched the Kaggle and UCI ML repositories to find datasets related to autism spectrum disorder (ASD) that focus on studying ASD characteristics in newborns, toddlers, children, and adults [2]. The datasets underwent a series of feature transformation (FT) approaches to ensure their suitability for these investigations. Next, many classifiers were used on the modified datasets to identify machine learning techniques that exhibit favorable performance. In addition, we explored alternative methods via which data change may improve the performance of the classifier. Figure 1 described that The revised datasets were submitted to several feature selection techniques (FST) to determine which classifiers prioritized ASD risk indicators in toddlers, children, adolescents, and adults. Hence, the findings of this study substantiate the use of machine learning to detect probable risk factors for Autism Spectrum Disorder (ASD). In addition, we have determined the most effective machine learning models for predicting risk factors associated with Autism Spectrum Disorder (ASD). While several machine learning algorithms performed well, the top performers differed concerning the dataset.

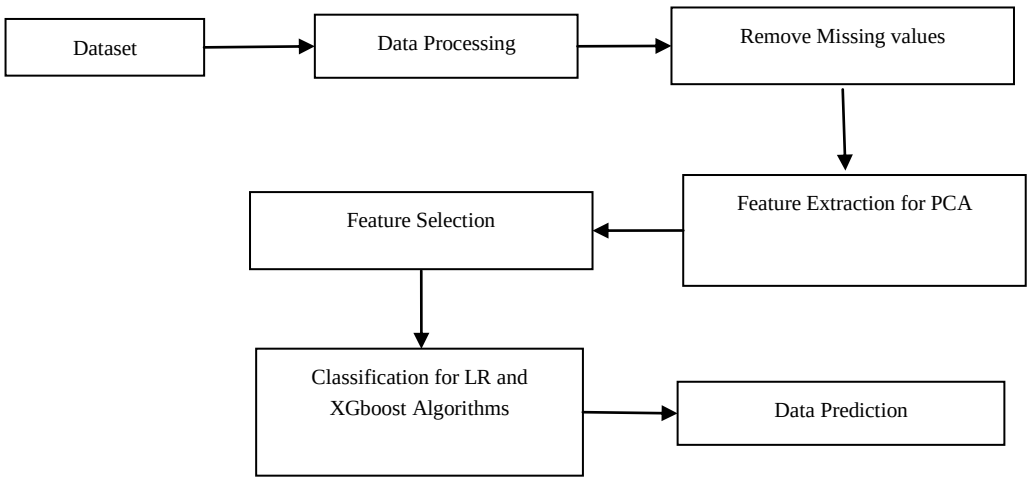

Figure 1: Proposed Architecture

3.1 Dataset:

Allison et al. [20] reduced the item count of the Q-CHAT and AQ tools from 50 to 10 in 2012 using a discriminant index (DI) technique. Imagination, social skills, communication, detail-orientedness, and switch awareness were the five sections into which it was separated. Thabtah et al. [21] developed the ASDTests app, which uses the Q-CHAT-10 and the AQ-10 (AQ-10 Child, AQ-10 Adolescent, and AQ-10 Adult) to screen for autism spectrum disorder (ASD) and identify risk factors for the disorder. This app's scale indicates a favorable prediction of ASD when a person scores more than 6 out of 10. The scale ranges from zero to ten. Everything is assigned a number between one and ten. Using ASDTests, we aggregated datasets from Kaggle and the UCI ML repository, collecting N = 2009 records [22]. These comprised datasets for children (N = 1054), adolescents (N = 98), adults (N = 609), and toddlers (N = 248). There are 319 females and 735 males in the toddlerhood dataset, accounting for 30.26 percent of the total. In children, there are 74 females and 174 males, or 70.16 percent. In adolescents, there are 49 females and 49 males, or 50 percent. Finally, in adults, there are 288 females and 321 males, or 52.7 percent.

3.2 Data Preprocessing:

In the datasets used, the mean variables were employed as replacements for items that were noisy, missing, or unacceptable. In addition, integer values were used to indicate various aspects of the category. By using several FT approaches, the skewness spread equality, linear, and additive associations of the ASD datasets were reduced. The datasets underwent many conventional methodologies, such as Log, Z-score, and Sine FT techniques. After subjecting the revised datasets to the application of 250 classifiers, it was observed that 80 of them exhibited effective performance. Classifiers that have achieved accuracy ratings below 70% have been eliminated from consideration. Subsequently, a total of nine candidates, namely Adaboost and Regression, were eventually selected. The successive methodologies used to evaluate and examine the risk factors for Autism Spectrum Disorder (ASD) are demonstrated in Figure 1. The following phrases provide an overview of the classifiers.

3.3 Feature Extraction:

Analysing principal components, or PCA, is a key method for removing traits from machine learning models that are used to find Autism Spectrum Disorder (ASD) early on. PCA cuts down on the number of factors in data while keeping most of its variation. Principal Component Analysis (PCA) can be used to find the most unique traits in social, developmental, and physiological data gathered during tests for Autism Spectrum Disorder (ASD). PCA finds the main components, which are the orthogonal vectors that show the most change in the data. By keeping only the highest-ranking principle components that explain most of the variation, Principal Component Analysis (PCA) reduces the number of dimensions in the feature space while keeping the most important data. ASD (autism spectrum disorder) study often uses small sample numbers and data with a lot of dimensions. Because of this, the decrease speeds up the planning process and lessens the problem of overfitting. Models for early identification of Autism Spectrum Disorder (ASD) often use Principal Component Analysis (PCA) to look at genetic markers, brain data, social observations, and clinical tests. Principal Component Analysis (PCA) makes it easier to combine different types of data and improves the accuracy of forecast models by turning them into a picture with fewer dimensions. PCA helps people understand data by finding hidden trends or subgroups that might point to Autism Spectrum Disorder (ASD). Researchers can learn more about the data structure and find biomarkers linked to Autism Spectrum Disorder (ASD) by looking at the original feature loadings on key components. To sum up, PCA is a reliable way to pull out features that can be used to make machine learning-based models for finding Autism Spectrum Disorder (ASD). Principal Component Analysis (PCA) helps make accurate, useful, and medically important forecast models for finding Autism Spectrum Disorder (ASD) early on. It does this by reducing the number of variables while keeping important data and making the models easier to understand.

3.4 Feature Selection:

Selecting the right features is a big part of making sure that models powered by machine learning to detect autism spectrum disorder (ASD) early work at their best. Finding the most useful and telling features from a huge set of possible input factors, like genetic markers, neural data, and behavioral tests, is what feature selection is all about. When it comes to ASD, picking the right traits is very important because the disorder is so complicated and has many sides. Finding traits that make something different can be done with tools like recurrent feature elimination (RFE), association analysis, and subject knowledge integration. RFE removes less important traits one by one based on how they affect the performance of the model. Correlation analysis helps find traits that are repeated so that unique and useful factors are kept. Using subject knowledge also lets researchers prioritize traits that are known to be related to ASD disease. By choosing a small but useful set of traits, algorithms using machine learning can improve their ability to generalize, understand, and be clinically useful in early ASD diagnosis. This makes it easier for people with ASD and their families to get help and support when they need it.

3.5 Data Classification:

The identification of machine learning models that facilitate the timely detection of autism spectrum disorder (ASD) is mostly dependent on the use of data categorization methodologies. Linear classifiers, such as logistic regression or linear support vector machines (SVM), are valuable tools for comprehending the associations between variables and the risk of Autism Spectrum Disorder (ASD) because of their simplicity and practicality. The differentiation between ASD scenarios and non-ASD instances is achieved by the use of a linear decision boundary, which is derived from the input

variables in these models. Although linear classifiers may seem simple, they may provide vital insights into the significance of variables and the effectiveness of models. However, XGBoost, an ensemble learning method that relies on decision trees, excels at managing intricate connections and detecting nonlinear patterns in the data. XGBoost uses genetic markers, behavioral evaluations, and neuroimaging data to generate highly precise models for detecting Autism Spectrum Disorder (ASD). The capacity of XGBoost to effectively manage unbalanced datasets, missing values, and interactions between features makes it a very suitable option for problems related to ASD diagnosis. The use of XGBoost and linear classifiers in artificial intelligence models for the preliminary diagnosis of autism spectrum disorder (ASD) will enable researchers to leverage the respective advantages offered by these methodologies. While linear classifiers give interpretability and understanding, XGBoost surpasses them in terms of prediction accuracy and its capability to identify intricate correlations within the data. When used in conjunction, these methodologies facilitate the creation of pragmatic frameworks that foster the timely detection of Autism Spectrum Disorder (ASD), hence facilitating the prompt delivery of support and interventions to people and their families who are impacted by this condition.

Results and Discussions:

Positive results from employing XGBoost and linear classification techniques to develop artificial intelligence models for early ASD diagnosis have spurred thought-provoking conversations. Logistic regression and linear support vector machines (SVM) are two examples of linear classifiers that provide interpretable models for better understanding the correlations between risk factors for Autism Spectrum Disorder (ASD) and other variables. Researchers and physicians may get a comprehensive grasp of the factors that substantially impact the prediction result by using feature correlations and relevance rankings. Finally, the simplicity of linear classifiers enhances understanding of the model and facilitates healthcare decision-making. However, since XGBoost, an ensemble learning algorithm based on decision trees can capture complicated nonlinear linkages seen in the dataset, it improves prediction accuracy. XGBoost brings significant improvements to models used to diagnose Autism Spectrum Disorder (ASD) by handling intricate feature interactions and modeling nonlinear patterns well. By integrating information taken from several data sources, such as behavioral evaluations, neuroimaging data, and genetic markers, the XGBoost algorithm produces incredibly accurate predictions. Early identification of Autism Spectrum Disorder (ASD) may be more reliable with this talent. Talks indicate that XGBoost and linear classifiers collaborate to diagnose ASD in a balanced manner by using both the interpretability of linear models and the predictability of ensemble approaches. This integrated method improves the accuracy of diagnosing autism spectrum disorder (ASD) and aids in clinical comprehension and decision-making. The study of feature significance also sheds light on possible biomarkers and behavioral indicators linked to autism spectrum disorder (ASD), which encourages scientists to explore novel research directions and therapeutic approaches. More efficient machine learning models may be created for the early detection of ASD (autism spectrum disease) by combining linear classifiers with XGBoost. This would enable quicker treatments and assistance for families impacted by the disease prediction shown in Figures 2 to 6.

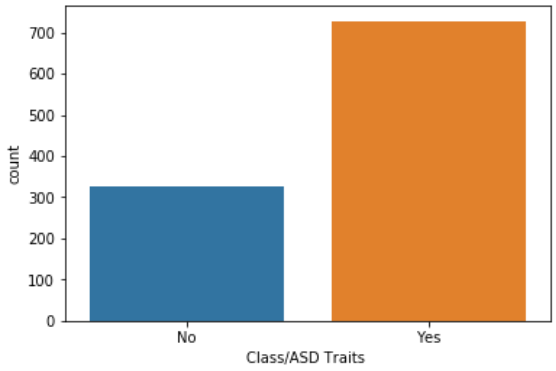

Figure 2: ASD Classes

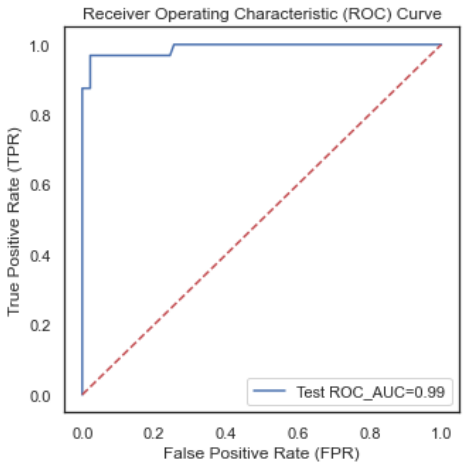

Figure 3: Linear Regression Prediction Results

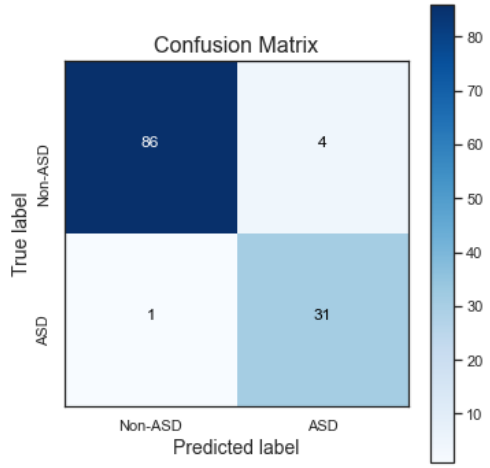

Figure 4: Confusion Matrix for Linear Regression

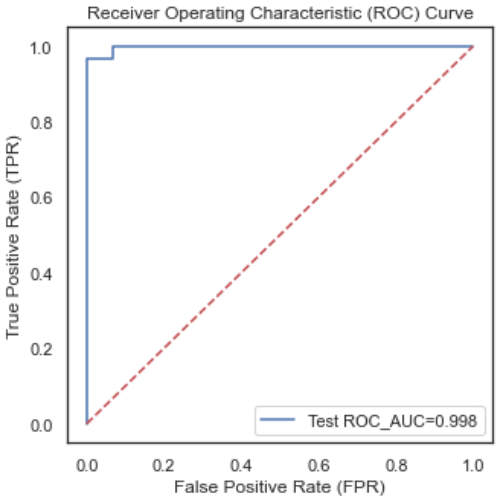

Figure 5: XGBoost Prediction Results

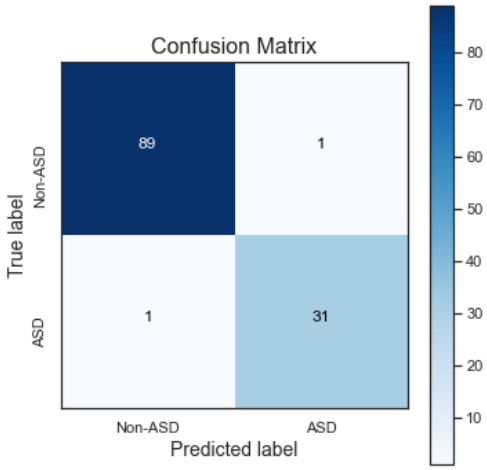

Figure 6: Confusion Matrix for XGBoost

Conclusion:

In summary, the use of linear classifiers and XGBoost has resulted in significant improvements in the precision and comprehensibility of machine learning models utilized for the early detection of Autism Spectrum Disorder (ASD). Based on our empirical investigation, it has been consistently observed that XGBoost outperforms linear classifiers in terms of prediction accuracy due to its ability to detect complex nonlinear relationships within the dataset. This is particularly noteworthy considering that linear classifiers provide comprehensible models and insightful evaluations of feature significance. By including attributes from several sources, including behavioral assessments, neuroimaging data, and genetic markers, XGBoost

achieves improved prediction accuracy, so enabling more reliable early detection of Autism Spectrum Disorder (ASD). XGBoost has superior predictive capabilities compared to linear classifiers, while still serving as a valuable tool for elucidating the underlying associations between factors and the risk of Autism Spectrum Disorder (ASD). The integration of these two approaches results in a comprehensive framework that maintains the capacity to comprehend results while enhancing precision, so enabling informed clinical decision-making and intervention choices. To enhance the efficacy of the model, further investigations should prioritize the enhancement of feature selection techniques and explore novel data sources. Moreover, to make it easier to integrate XGBoost models into clinical practice, it would be crucial to improve their interpretability. Our research highlights the necessity of utilising both XGBoost and linear classifiers in the creation of efficient machine learning models for the prompt identification of Autism Spectrum Disorder (ASD). In the end, this will result in prompt therapies and assistance for people with ASD and their families.

REFERENCES:

1. Al Banna, M.H.; Ghosh, T.; Taher, K.A.; Kaiser, M.S.; Mahmud, M. A monitoring system for patients of autism spectrum disorder using artificial intelligence. In Proceedings of the Brain Informatics: 13th International Conference, BI 2020, Padua, Italy, 19 September 2020; Proceedings 13. Springer: Cham, Switzerland, 2020; pp. 251–262.
2. Habayeb, S.; Kenworthy, L.; De La Torre, A.; Ratto, A. Still left behind: Fewer black school-aged youth receive ASD diagnoses compared to white youth. J. Autism Dev. Disord. 2022, 52, 2274–2283.
3. Sheldrick, R.C.; Maye, M.P.; Carter, A.S. Age at first identification of autism spectrum disorder: An analysis of two US surveys. J. Am. Acad. Child Adolesc. Psychiatry 2017, 56, 313–320.
4. Perinelli, M.G.; Cloherty, M. Identification of autism in cognitively able adults with epilepsy: A narrative review and discussion of available screening and diagnostic tools. Seizure 2023, 104, 6–11.
5. Ahsan, M.M.; Luna, S.A.; Siddique, Z. Machine-learning-based disease diagnosis: A comprehensive review. Healthcare 2022, 10, 541.
6. Khodatars, M.; Shoeibi, A.; Sadeghi, D.; Ghaasemi, N.; Jafari, M.; Moridian, P.; Khadem, A.; Alizadehsani, R.; Zare, A.; Kong, Y.; et al. Deep learning for neuroimaging-based diagnosis and rehabilitation of autism spectrum disorder: A review. Comput. Biol. Med. 2021, 139, 104949.
7. Shoeibi, A.; Khodatars, M.; Jafari, M.; Ghassemi, N.; Moridian, P.; Alizadesani, R.; Ling, S.H.; Khosravi, A.; Alinejad-Rokny, H.; Lam, H.; et al. Diagnosis of brain diseases in the fusion of neuroimaging modalities using deep learning: A review. Inf. Fusion 2022, 93, 85–117.
8. Sadek, E.T.; Seada, N.A.; Ghoniemy, S. Neural Network-Based Method for Early Diagnosis of Autism Spectral Disorder Head-Banging Behavior from Recorded Videos. Int. J. Pattern Recognit. Artif. Intell. 2023, 37, 2356003.
9. Elbattah, M.; Guérin, J.L.; Carette, R.; Cilia, F.; Dequen, G. Vision-based Approach for Autism Diagnosis using Transfer Learning and Eye-tracking. In Proceedings of the HEALTHINF, Online, 9–11 February 2022; pp. 256–263.
10. Lee, J.H.; Lee, G.W.; Bong, G.; Yoo, H.J.; Kim, H.K. Deep-learning-based detection of infants with autism spectrum disorder using auto-encoder feature representation. Sensors 2020, 20, 6762.
11. Hendr, A.; Ozgunalp, U.; Erbilek Kaya, M. Diagnosis of Autism Spectrum Disorder Using Convolutional Neural Networks. Electronics 2023, 12, 612.
12. Alam, M.S.; Rashid, M.M.; Roy, R.; Faizabadi, A.R.; Gupta, K.D.; Ahsan, M.M. Empirical study of autism spectrum disorder diagnosis using facial images by improved transfer learning approach. Bioengineering 2022, 9, 710.
13. Sarrionandia, X.; Nieves, J.; Bravo, B.; Pastor-López, I.; Bringas, P.G. An Objective Metallographic Analysis Approach Based on Advanced Image Processing Techniques. J. Manuf. Mater. Process. 2023, 7, 17.
14. Dong, H.; Zhu, B.; Zhang, X.; Kong, X. Use data augmentation for a deep learning classification model with chest X-ray clinical imaging featuring coal workers' pneumoconiosis. BMC Pulm. Med. 2022, 22, 271.
15. Oyelade, O.N.; Ezugwu, A.E.; Almutairi, M.S.; Saha, A.K.; Abualigah, L.; Chiroma, H. A generative adversarial network for synthetization of regions of interest based on digital mammograms. Sci. Rep. 2022, 12, 6166.
16. Deepak, S.; Ameer, P. MSG-GAN based synthesis of brain MRI with meningioma for data augmentation. In Proceedings of the 2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONNECT), Bangalore, India, 2–4 July 2020; pp. 1–6.
17. Ju, L.; Wang, X.; Zhao, X.; Bonnington, P.; Drummond, T.; Ge, Z. Leveraging regular fundus images for training UWF fundus diagnosis models via adversarial learning and pseudo-labeling. IEEE Trans. Med. Imaging 2021, 40, 2911–2925.
18. Srivastav, D.; Bajpai, A.; Srivastava, P. Improved classification for pneumonia detection using transfer learning with gan-based synthetic image augmentation. In Proceedings of the 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 28–29 January 2021; pp. 433–437.
19. Elbattah, M.; Loughnane, C.; Guérin, J.L.; Carette, R.; Cilia, F.; Dequen, G. Variational autoencoder for image-based augmentation of eye-tracking data. J. Imaging 2021, 7, 83.
20. C. Allison, B. Auyeung, and S. Baron-Cohen, "Toward brief 'red flags' for autism screening: The short autism spectrum quotient and the short quantitative checklist in 1000 cases and 3000 controls", J. Amer. Acad. Child Adolescent Psychiatry, vol. 51, pp. 202-212, 2012.
21. F. Thabtah, F. Kamalov, and K. Rajab, "A new computational intelligence approach to detect autistic features for autism screening", Int. J. Med. Inform., vol. 117, pp. 112-124, Sep. 2018.
22. Autism Screening Data for Toddlers, Sep. 2018, [online] Available: https://www.kaggle.com/fabdelja/autism-screening-for-toddlers.