International Journal of Innovative Research in Engineering
Volume 4, Issue 2 (March-April 2023), PP: 08-11.
www.theijire.com **ISSN No: 2582-8746**

Detecting Autism Spectrum Disorder Using MachineLearning Techniques

Madhanraj T[1], Tharun V[2], Vishnu Kumar R[3], Divya P[4]
[1,2,3,4] *Computer science and engineering, Bannari Amman Institute of Technology, Tamilnadu,India.*

How to cite this paper:
Madhanraj T[1], Tharun V[2], Vishnu Kumar R[3], Divya P[4] "Detecting Autism Spectrum Disorder Using MachineLearning Techniques", IJIRE-V4I02-08-11.

Abstract: *Those with autism spectrum disorders (ASDs) engage in a variety of disruptive behaviors. In most cases, they cannot speak clearly. Instead, they establish a relationship by gesturing and pointing. As a result, one of the most difficult tasks for caregivers is to comprehend their requirements, but early diagnosis of the disease can make it much simpler. The Internet of Things (IoT) and assistive technologies can eliminate the lack of verbal and nonverbal communication. By employing Deep Learning (DL) and Machine Learning (ML) algorithms, the IoT-based systems aid in diagnosis and enhance patient lives. Neuro developmental disorder known as Autism Spectrum Disorder (ASD) has an impact on behavior, social interaction, and communication. Because the symptoms and characteristics can vary widely from person to person, it is referred to as a "spectrum" disorder. Some people with ASD may have trouble communicating verbally and nonverbally, have trouble making and keeping friends, do things over and over again, or have narrow interests. Math, music, and art are just a few of the areas in which some people with ASD excel. The onset of symptoms typically occurs in early childhood, and their severity can range from mild to severe. While the specific reason for ASD isn't known, being a blend of hereditary and ecological factors is accepted. Although there is currently no known treatment for ASD, early intervention and therapy can assist individuals with ASD in acquiring new abilities and enhancing their quality of life.An agglomeration information stream has introduced a very important formulation for information and information engineering because the applications currently offer streaming information. It is challenging to instantiate in order to generate a vast variety of functions known as micro-clusters in tandem with an online method. The collective density of micro-clusters increasesas a result of providing the information of enormous data points through a defined lay. A hypothetical to express agglomeration algorithmic rule that is utilized in a specific offline rate to transform the micro-clusters into the enormous final cluster is the requirement that is currently available.*

Key Word: *Autism spectrum disorder, IOT, Deep and Machine learning, Algorithms, classification.*

I.INTRODUCTION

An active area of research is machine learning-based diagnosis of autism spectrum disorder (ASD). Machine learning can be used to find ASD in a number of different ways. Any machine learning-based approach to ASD detection begins with the extraction of pertinent features from the data. The data can come from a variety of sources, such as MRI images, genetic data, or behavioral data. The process of extracting patterns from the data that are relevant to ASD is known as feature extraction. A machinelearning algorithm can be trained to classify individuals as having ASD or not having ASD after the relevant features have been extracted. Characterization calculations, for example, support vector machines (SVM), choice trees, and arbitrary woodlands, can beprepared utilizing the removed elements. Convolutional neural networks (CNNs) and other forms of deep learning have also been utilized in the detection of ASD. CNNs are able to automatically learn relevant features from raw data, such as MRI images or EEG data, and classify individuals as having ASD or not having ASD. AdaBoost and random forests are examples of ensemble learning methods that can combine the results of multiple machine learning algorithms to improve the accuracy of ASD detection. In general,a promising area of research is the application of machine learning strategies to the detection of ASD. However, it is essential to keep in mind that machine learning models are not a substitute for a trained healthcare professional's clinical diagnosis.

Data Mining Concept:

Data mining is a method for looking for reliable patterns and/or methodical relationships between variables in large amounts of data, typically associated with the industry or marketing, and then validating the result by applying the patterns to original subsets of data. Forecasting is the ultimate goal of data mining, and projecting is the most common type of data mining with the most direct business application. There are three stages to the data mining process: 1) The initial investigation; 2) Validation and verification ofthe representation structure or pattern recognition; and 3) Application (i.e., the application of the representation to novel data in orderto produce a prediction).

II.LITERATURE REVIEW

A Framework for clustering evolving data streams:

Charu C. Aggarwal and Jiawei Han..en[1] have proposed a data stream clustering philosophy that is primarily unusual and guided by application-centered requirements. The plan divides the clustering process into an online module that sometimes includes complete review statistics. The VLDB copyright notice, the title of the publication, and the publication's date are included, and it isspecified that replication is authorized by the Very Large Data Base Endowment. The authorization to copy without paying for all or part of the material is also included. A fee and/or specific permissions are required for any other copying or republishing. The forecaster uses the online section to quickly consider the large clusters in the data stream and use a wide range of inputs (such as time horizon or cluster count) to do so. The challenges of selecting, storing, and utilizing this statistical data in a well-organized manner for a rapid data flow become quite challenging. We employ micro-clustering techniques in conjunction with the notions ofa pyramidal occasion surround for this justification. The electrifying, competent, and insightful nature of our methods are demonstrated by our presentation experiments on a variety of real and synthetic data sets.

Data stream clustering: A survey

According to JONATHAN A. SILVA, ELAINE R. FARIA,..en[2], data stream mining is a burgeoning field of study that seeks to extract knowledge from vast quantities of continuously generated data. Numerous data stream clustering methods have been predicted to use unverified knowledge in this situation. However, data stream clustering presents a number of obstacles that must be overcome, such as online commerce with non-stationary, infinite statistics. The essential nature of stream data necessitates the development of algorithms capable of rapid and incremental information substance distribution while appropriately addressing time and memory constraints. As well as a methodical discussion of the major design mechanism of cutting-edge algorithms, this paper provides a review of data stream clustering methods. In addition, this work provides an overview of commonly used new methodologies and addresses the sequential aspect of data stream clustering. Data stream cluster's application in a variety of fields, including stock market research, sensor networks, and network intrusion detection, is illustrated by a number of references. Software correspondence and data repositories are also accessible to numerous practitioners and researchers. Finally, a few important issues and unanswered questions that could be the focus of prospect research are discussed.

Density based clustering over an evolving data stream with noise:

Feng Cao and Martin Ester..en[3] have planned that Clustering is an important task for developing data streams in mining. In addition to the one-pass constraint and incomplete memory, the nature of developing data streams implies the following sources for stream clustering: no presumption regarding the number of clusters, the ability to detect clusters of random shape, and the capacity to leverage outliers Although numerous data stream clustering algorithms have been proposed, none of them provide a solution to the problem of grouping these requirements together. Den Stream, a novel method for finding clusters in a growing data stream, is the subject of this work. The "dense" micro-cluster, also known as the "core-micro-cluster," is introduced to examine the cluster's random shape, while the "possible core-micro-cluster" structure and "outlier micro-cluster structure" are anticipated to continue distinguishing between the possible clusters and outliers. Based on these ideas, a novel pruning strategy is planned to guarantee the weight accuracy of micro clusters with insufficient memory. The efficiency and effectiveness of our method are demonstrated in ourpresentation learn over a variety of real and fake data sets.

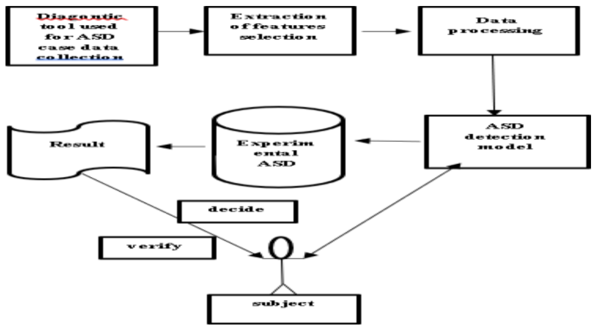
III.EXISTING METHODS

Autism Spectrum Disorders (ASD) can be difficult to diagnose using conventional methods in the current SVM system. The child'soverall psychological well-being is improved when autism spectrum disorders are detected early. Due to the rising number of people with ASD, research on the disorder is now being carried out at a much faster rate than it was in the past. Changes in children's behavior at an early age could be one way to diagnose ASD. Though the precise neuroanatomical nature of these disturbances is stillunclear, structural imaging hypotheses point to disturbances in various brain regions. For the development of biomarkers that couldeventually be used to improve analysis and screen response to treatment, it is essential to depict cerebrum structural contrasts in children with ASD. In this study, we employ machine learning to select a number of conditions that together constitute autism spectrum disorder.

IV.PROPOSED SYSTEM

Neurodevelopment syndrome, also known as autism spectrum disorder (ASD), is characterized by impairments in social communication, abnormal behavior, and sensory activities. ASD is diagnosed through observation of frequent movements and social communication abilities. and the person's eye contact. Autism Spectrum Disorder (ASD) is a brain development disorder that affectssocial interaction and communication. Support vector machines (LSTM), linear regression, and laser regression have all been used in a number of studies to classify autism. Linear regression, laser regression, and support vector machines are all examples of machine learning methods that have been used to classify autism. Autism spectrum disorder is diagnosed during medical treatment by professionals who are experienced in the condition. However, the diagnosis of autism spectrum disorder still makes use of screeninginstruments that necessitate direct observation from clinical professionals. The proposed LSTM classification techniques have been developed in a variety of Critical information classification retrieval studies inspired the first query classification method. Comparison of Cancer Classifications.

V.ARCHITECTURE DIAGRAM



Identifying consensus maximization:

A list of concepts that have been assigned to types is contained within the concept model nugget itself. For scoring against other data, you can select any or all of the concepts in that model. Prior to building the model, new fields are added to the data in accordance with the build mode selected on the Model tab of the Text Mining modeling node when you execute a stream containing a Text Mining model nugget. In general, points close to the boundary are critical. The most important issue for most classification algorithmsis how precisely they classify points that are close to the course group limits. The "slam dunk," or easy cases, are those that are far from the class boundaries and have little effect from misclassification. On the other hand, the classification algorithms are capable of effortlessly classifying either a case that is manifestly cancer-free or one that has fully developed cancer. The borderline cases, onthe other hand, are crucial because early detection can save a life and they may exhibit subtle symptoms of cancer. As a result, identifying critical consensus maximization may necessitate the presence of uncertain regions within and around class boundaries.

Finding critical consensus maximization:

There are two phases to how the Find Critical Consensus Maximization algorithm works. It determines the critical consensus maximization for each of the two classes in each phase. The data instances in the boundary set are examined one at a time using thereduced boundary set for each class. The boundary set treats each data instance as a neighborhood's central point. Finding all pointsthat belong to the same class and are within a distance R of the center point creates a neighborhood. Because the objective is to locate critical consensus maximizations that belong to one class but switch to the other class when their attribute values are altered (a totalof two classes are assumed), one class at a time is taken into consideration. The first step in the process of finding critical consensusmaximization is to find an approximate boundary set. The next step is to look at a neighborhood around each boundary point and figure out its CR score.

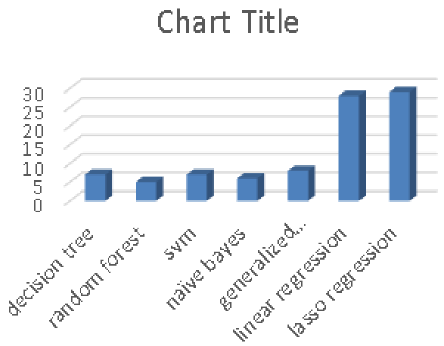
Improving classification accuracy and critical consensus maximization:

Most of the time, classification algorithms are judged on how accurate their predictions are. An algorithm is considered to be accurate if the predictions contain a minimum number of false positives and false negatives. Various data sets were used in the experimental phase to see if critical consensus maximization could help improve classification accuracy. Additional small-scale classification models were derived using the identified consensus maximization. A new classification model is developed and trained for each class by first obtaining a new data set that is a subset of the initial training data set. By reclassifying a subset of the original data records into two new classes, the new data set was created.

A single class is formed by data records belonging to the top k critical consensus Maximization class, such as the "" class. Another class is given to data records that are close to the top k critical consensus maximization but do not belong to the set of critical consensus maximization (this could include instances that belong to both the "" and "_" classes).

VI.RESULT AND DISCUSSION

In recent years, a number of studies have focused on using machine learning to identify autism spectrum disorder. Some examples of these studies' outcomes are as follows: Electroencephalography (EEG) data was used by researchers in a study that was published in the Journal of Medical Systems to classify people with Autism Spectrum Disorder and typically developing people. The EEG datawere used to detect ASD with an accuracy of 85%, according to the study. Researchers used machine learning methods to find patterns in the functional connectivity of brain networks in people with autism spectrum disorder in another study that was publishedin the journal Scientific Reports. A study published in the Journal of Autism and Developmental Disorders used machine learning techniques to predict Autism Spectrum Disorder in infants at high familial risk for the disorder. The study found significant differences in the patterns of brain connectivity between people with ASD and people who are normally developing. ASD could bepredicted with an accuracy of 80% using a combination of behavioral and neural biomarkers, according to the study. Based on environmental risk factors like exposure to pesticides and air pollution, a study in the International Journal of Environmental Research and Public Health used machine learning techniques to predict autism spectrum disorder. Based on environmental risk factors, the study predicted ASD with 70 percent accuracy. In general, these studies suggest that methods based on machine learningcan be useful in identifying potential biomarkers or risk factors for autism spectrum disorder and detecting the disorder itself. However, additional research is required to confirm these findings and develop diagnostic tools for ASD that are more accurate andreliable.



VII.CONCLUSION

Autistic Spectrum Disorder has an impact on the patients' communication and behavior because it is a developmental disability. Thedisease's incidence can be greatly reduced by early detection. The ASD diagnosis process is implemented using the LINEAR and LASSO REGRESSION models in this work. Data and feature engineering, model training, and model testing are all part of this application. Our work has achieved high performance on the Autism Screening Classification data set as a result of proper work on these parts, as evidenced by high values for precision, recall rate, and F 1 score. It demonstrates that this machine learning-based method may aid in the diagnosis of ASD in practice. In addition, in order to facilitate comprehension of the data and feature engineering in the visualization portion, we present a number of feature distribution images. In conclusion, machine learning methods have shown promise for identifying potential biomarkers or risk factors for autism spectrum disorder (ASD). Various kinds of data,such as EEG data, patterns of brain connectivity, behavioral and neural biomarkers, and environmental risk factors, have been analyzed using these methods. Although these studies' findings are encouraging, it is essential to keep in mind that clinical diagnosisand evaluation by trained professionals should always come first. Multiple assessments and evaluations, including behavioral observation, medical history, and standardized tests, are all part of the complicated process of making an ASD diagnosis. However, in cases where there is a high level of uncertainty or complexity, machine learning can be a useful tool for assisting clinicians in making diagnoses that are more accurate and efficient. Additionally, the use of machine learning techniques in the detection and characterization of ASD is an exciting area of research that has the potential to significantly improve our understanding and management of this complex disorder. Additionally, the identification of potential biomarkers or risk factors for ASD can inform thedevelopment of more targeted and personalized interventions for individuals with ASD.

References

[1] "A framework for clustering evolving data streams," by C. C. Aggarwal, J. Han, J. Wang, and P. S. Yu, in Proceedings of the International Conference on Very Large Data Bases (VLDB '03), 2019, pages 81–92.

[2] Data stream clustering: A study by J. A. Silva, E. R. Faria, R. C. Barros, E. R. Hruschka, A. C. P. L. F. d. Carvalho, and J. A.Gama, A survey," volume of ACM Computing Surveys 46, no. 1, pp. 13:1–13:31, Jul. 2019.

[3] "Density-based clustering over an evolving data stream with noise," by F. Cao, M. Ester, W. Qian, and A. Zhou, in Proceedingsof the 2006 SIAM International Conference on Data Mining. Pages SIAM, 2019, 328–339.

[4] "Density-based clustering for real-time stream data," by Y. Chen and L. Tu, is published in the Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. USA: New York, NY: ACM, 2019, pp. 133–142.

[5] "Stream data clustering based on grid density and attraction," by L. Tu and Y. Chen, in ACM Transactions on Knowledge Discovery from Data, vol. 3, no. 3, pp. 1–27, 2019.