Identifying The Possibility Of Autism Spectrum Disorder Using Machine Learning

Komala D, Veena C J, Sushma C K, Soumya Alaundi, Preksha M

Assistant Professor,Dept. of CSE,DBIT, Banglore,India.

Student,Dept. Of CSE,DBIT,Banglore,India

ABSTRACT: Autism spectrum disorder is a neurological initial disorders that effects an individuals's daily activities. In spite of the way that evaluation of mental unevenness is imaginable at any stage throughout everyday life, it's secondary effects develop in first two decades of life, yet they progress with duration. Mental illness cases have different sorts of troubles similar as difficulties with thought, learning handicaps, motor difficulties, tricky issues, internal ailments near as pressure, debilitation, and different others.

In the current world, there are many ways used to survey them all around, without relating their empathy or express characteristics. The proposed machine learning model incorporates mental handling on a clever schooling. . This approach gives sensible application software by relating the children expressions. A deliberate strategy which combines mental, preliminary and lead mind science is used to start a potential unevenness assessment model, by using which a specific operating system is given to the children. The proposed structure turns out to be a convincing plan in diagnosing Autism.

Keywords: Autism, Spectrum,Random forest,SVM.

1. INTRODUCTION

1.1. Project description

Autism spectrum disorder is a neurological shape that influence a persons capacity to communicate, correspondence and data chops. Despite the way that evaluation of substance leaning to one side to be considerable at anything stage throughout everyday life, it affects all around occur during first 2 times of life and exist. Compound imbalance causes face different kinds of troubles practically identical as difficulties with thought, learning inadequacies, inside medical problems including anxiousness, disablement, etc., Sensory challenges, delicate issues as well as different etc . When some of them will provide well-established caring as well as evaluation for mental irregularity requires gigantic quantum of time and cost. Before disclosure of compound unevenness can provide to a phenomenal with some genuine drug at a starting scene. This could prevent the case's condition from deteriorating more. With diminishing wide stretch expenses related to deferred evaluation.

Hence, a period effective, exact and basic association test gadget is truly critical mentioned which would predict , lack of imbalance characteristics in an individual and recognize whether they bear total compound irregularity evaluation. Its goal of study would stimulate flexibility which might adequate and propose a psychological lopsided model using ML methods. Gauge compound unevenness characteristics of an existent of any age. Thus, this work bases on cultivating a synthetic lopsided imbalance netting action for imagining the ASD qualities among persons nature enough social events 4-11 times, 12-17 times and for people mature enough 18 and further.

2. LITERATURE SURVEY

2.1 Hybrid feature-based anxiety detection in autism using hybrid optimization tuned artificial neural network

Autism is a neuro behavioral disorder that negatively influences the social and communication skills of the individual. Autism is also known as an autism spectrum disorder (ASD) due to its various types of symptoms. The anxiety among the normal person concerns the physician especially, for the ASD person as it creates negative impacts on psychological and physical health. Hence, a dynamic anxiety detection model is developed in this research with the support of an optimized Artificial Neural Network (ANN). Initially, the input EEG signal from the database is pre-processed for the removal of noise and artifacts of the signal using the bandpass filter (BPF) and then the significant features are extracted for the reduction of the complexity. Then, the ANN is utilized for the detection of the anxiety, which is trained using the proposed Trace and Forage optimization (TF) algorithm that is designed through the combination of the characteristics of the rescue search agents and the Finches to obtain the enhanced ability of detection. The analysis is done band-wise using the two standard databases such as DEAP(Dataset for Emotion Analysis using Physiological and Audiovisual Signals) and SEED datasets in terms of metrics, such as specificity, sensitivity, and accuracy. The proposed anxiety detection model achieves an accuracy of 96.67 % with the help of DEAP in accordance with the training percentage.

2.2 Very early detection of autism spectrum disorders based on acoustic analysis of pre-verbal.

With the increasing prevalence of Autism Spectrum Disorders (ASD), very early detection has become a key priority research topic, as early interventions can increase the chances of success. Since atypical communication is a hallmark of ASD, automated acoustic-prosodic analyses have received prominent attention. Existing studies, however, have focused on verbal children, typically over the age of three (when many children may be reliably diagnosed) and as high as early teens. Here, an acoustic-prosodic analysis of pre-verbal vocalizations (e.g., babbles, cries) of 18-month old toddlers is performed. Data was obtained from a prospective longitudinal study looking at high-risk siblings of children with ASD who were also diagnosed with ASD, as well as low-risk age-matched typically developing controls. Several acoustic prosodic features were extracted and used to train support vector machine and probabilistic neural network classifiers; classification accuracy as high as 97% was obtained. Our findings suggest that markers of autism may be present in pre-verbal vocalizations of 18-month old toddlers, thus may be used to assist clinicians with very early detection of ASD. Acoustic-prosodic features encompassing pitch, formant, energy, harmonics, and vocal quality (e.g., breatheness) were extracted from audio recordings obtained from a prospective study that looked at high risk

2.3 Diagnostic of autism spectrum disorder based on structural brain MRI images using, grid search optimization, and convolutional neural network

In this study, an automatic autism diagnostic model based on sMRI is proposed. This proposed model consists of two basic stages. The first stage is the preprocessing stage, which consists of removing unclear images, identifying the edges of the images by applying the canny edge detection (CED) algorithm, cropping them to the size required by the system, and finally enlarging the images five times with data augmentation. The data augmentation method should not affect the discrimination in the images such as coloring, and also since it is applied to both groups of autism spectrum disorders (ASD) and typical development (TD), it is performed with care not to cause any manipulation in the data. In the second stage, the grid search optimization (GSO) algorithm is applied to the deep convolutional neural networks (DCNN) used in the system to have optimal hyperparameters. As a result, the proposed diagnostic method of ASD based on sMRI achieves an outstanding success rate of 100%. The reliability of the proposed model is validated by testing with five-fold cross-validation, and its superiority is demonstrated by comparing it with recent studies and widely-used pre-trained models.

2.4 Energy-aware EEG-based Scheme for Early-age Autism Detection.

The affordability and miniaturization of sensors create a revolution in wearable wireless solutions deployed to collect physiological parameters to assist in diseases/disorders diagnosis. Electroencephalography (EEG), a recording of the brain's electrical activity, is a promising physiological measure for autism spectrum disorder detection. It can reveal the irregularity of the neural system that is associated with autism. Wireless sensors represent a suitable infrastructure that can be deployed for signal transmission to the processing center. However, streaming EEG signals remotely for classification could shorten the lifetime of the wireless sensor and might question the viability of the application. Therefore, reducing the data transmission might preserve the sensor's energy and increase the wireless sensor network lifetime. This paper proposes the design of a sensor-based scheme for early-age autism detection in children. Automated ASD(Autism Spectrum Disorder) detection can reduce subjectivity in ASD diagnosis and can significantly help physicians in the Diagnosis process. • This research presented a new sensor based ASD detection system based on EEG(Electroencephalogram). By extracting and classifying features locally, the proposed approach is intended to reduce energy consumptions.

3. PROBLEM STATEMENT

Our proposal is to develop a machine learning model that can accurately predict the likelihood of a child being diagnosed with Autism Spectrum Disorder (ASD) based on their behavioral and demographic characteristics. The goal is to improve early detection and diagnosis of ASD, which can lead to better outcomes and earlier intervention for affected children. This project aims to address the challenge of accurately identifying children who may be at risk of ASD, as early intervention can improve outcomes and quality of life for affected individuals and their families.

4. SYSTEM DESIGN

This defined by considering a framework, components, modules, interfaces, and data for a system to meet predetermined requirements is known as functional planning. This introduction of either a framework to idea generation may be viewed as software architecture. Among used here methods for designing PCS(Principal Component Analysis) is object-oriented methodology. Its principle stating for object-oriented modeling is now the UML. It is really utilized a lot for modeling apps, but less so for rising – anti relationships as well as frameworks. Likewise, the process of defining and creating processes to fulfill certain requirements of either the enterprise is known as modelling. Interaction between the mechanisms and just how they interact to administer entire system is included.

4.1. System Architecture

Data Collection:

Collecting the data which is relevant for individuals with Autism Spectrum Disorder (ASD) and without ASD.The data could include socioeconomic information, medical history, behavioral characteristics, etc.

Data Pre-processing:

The collected data should be clean to ensure it is accurate and complete. Transforming the data into a usable format for machine learning algorithms

Feature Selection:

Recognizing the most important features in the data that are predictive of ASD. This can be done using analytical methods, machine learning techniques, or domain expertise.

Model Selection:

Selecting a suitable machine learning algorithm for predicting ASD based on the selected features. This model include classification algorithms such as Decision Trees, Random Forest, Support Vector Machines, etc.

Model Training:

Using the selected machine learning algorithm to train the model based on the collected data and pre-processed data. This involves splitting the data into training sets and testing sets, fitting the model to the training data, and estimate its performance based on the testing data.

Model Evaluation:

Evaluating the performance of the trained model in predicting ASD accurately. This involves using various standard such as accuracy, precision, recall, F1 score, etc.

Deployment:

Accommodate the trained model into a user-friendly application that can be used by healthcare professionals or individuals concerned about the ASD. This could include creating a web or mobile application that provides forecast based on user data input .

Continuous Monitoring and Improvement:

Regularly observing the performance of the distribution model to check its accuracy and reliability. Improving the model as new data become accessible or as new machine learning techniques are evolved.

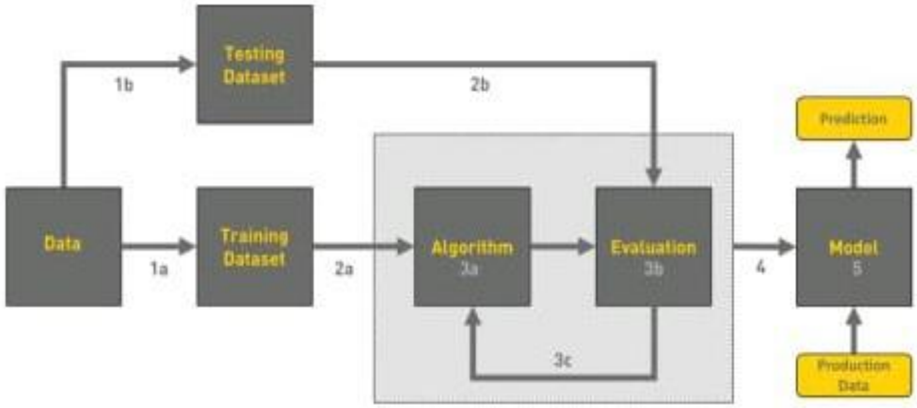

Fig 1 : System Architecture

The following diagram shown in fig-2 indicates the principle steps concerned in the manner of predicting autism spectrum sickness using machine mastering. The procedure begins with input facts collection, where facts is accumulated from various assets, consisting of scientific statistics, surveys, and questionnaires. The information is then preprocessed to remove noise and inconsistencies

.

The following step is feature extraction, wherein applicable features are extracted from the data. Characteristic selection is then executed to pick out the most applicable features. Characteristic scaling is then carried out to scale the functions to a comparable range

The next step is system gaining knowledge of, wherein diverse fashions are trained at the facts. The fashions are evaluated to determine their accuracy and performance. The great version is then selected, and its hyper parameters are tuned to improve its overall performance.

The very last step is version deployment, in which the educated model is deployed in a real-global placing to make predictions. The predictions may be made via a user interface, including a cellular app or a web-primarily based tool.

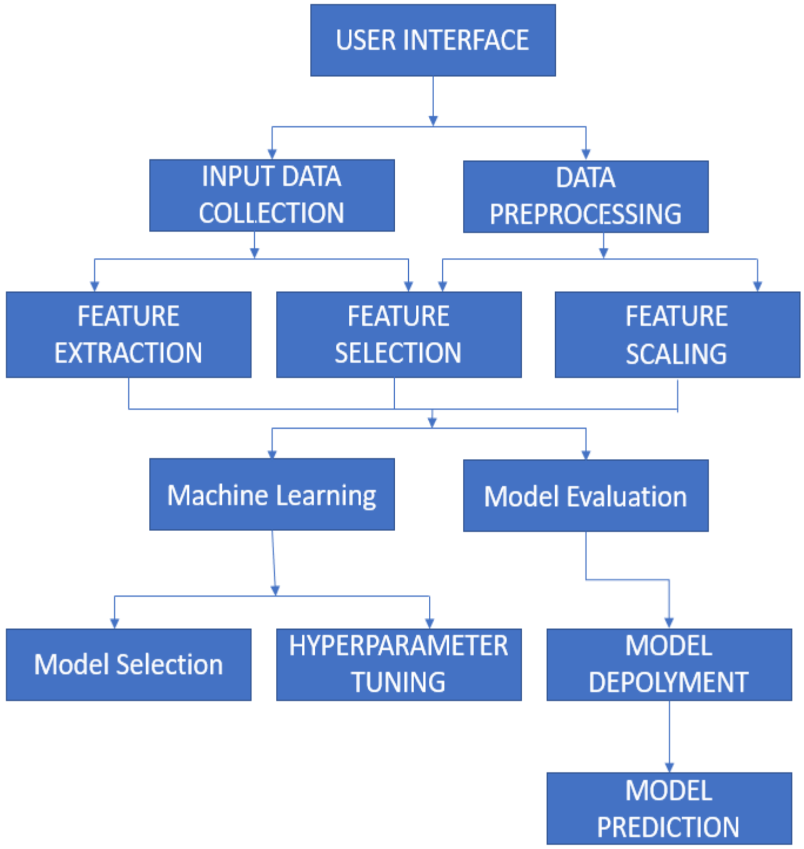

Fig 2 : Context Flow Diagram

Algorithms:

❖ **Support Vector Machine Algorithm**

Support- vector machines (also known as support- vector networks) in machine learning are supervised knowledge models with corresponding knowledge algorithms that dissect data used for regression and type analysis.

Step 1 : Original gene subset input: G = 1, 2, n

The output is a list that is ranked using the R-lowest weight criterion.

Step 2 : Set R = in step one.

Repeat steps 3- 8 in step 2 until G is not empty.

Step 3: is to use G to train the SVM.

Step 4: Computer the Weight Vector using eq (3)

Step 5: Enter the ranking criteria into a computer; Rank = W2

Step 6: Sort the features according to their rankings.

Step 7 :RankNew = type (Rank)

Step 8 :Update the point Rank list in step seven.

Step 9 :Revised R = R G (NEWrank)

Step 10 : Discard the point with the lowest rank. G = G- G Update (Newrank)

Step 11: Finish

❖ **Random Forest Algorithm**

The basic idea behind the approach is that building a compact decision tree with lots of attributes is a simple and inexpensive computational task. However, if we can construct a lot of weak decision trees in parallel small, we can also join the trees to form one.

Random Forest

Precondition A training set S = (x1, y1) (xn, yn), features F, and the number of trees in the timber B. RAND0MFOREST (S, F)

H ← Ø

if I 1, then B do S (i) a bootstrap sample from the RAND0MIZED TREE LEARNING for S

Hi (S(i), F

H H ∪ {hi}

stop for

revert H

halt operation

RAND0MIZED TREELEARNE function (S, F)

Every node:

f really tiny subgroup of F Splitting at the swish point in f

return The taught tree

halt operation.

❖ **Naive Bayes Classifier**

The Naïve Bayesian Classifier approach is based with the so Bayes hypothesis as well as works best once the sources have a higher level of aspect. Naive Bayes can frequently outperform more complex grid designs whilst being far more straightforward. Algorithms, both continuous or categories, can accept any number of factors. Classifiers in cognitive computing are a category of straightforward "probabilistic classifiers" based on the application of bayes' the0rem with strong naïve independence hypotheses between the features.

Fitting a naïve Bayes classifier to binary features

```
Nc = 0, Njc = 0;
for i = 1: N do
c = yi          // Class label of ith example;
Nc: = Nc + 1;
for j = 1: D do
if xij = 1 then
Njc: = Njc  | 1
π̂ = Nc/N ,  θ̂jc  =  Njc/Nc
```

❖ **Logistic Regression**

Early in the 20th century, natural lore began to apply logistic regression. Additionally, it was applied in a wide range of social wisdom operations. When the based variable (target) is categorical, logistic regression is utilized. Similar to direct retrogression, logistic retrogression employs an equation as its representation.

To predict an outcome value, input values(x) are linearly integrated with weights or measure values (referred to as the Greek capital letter Beta) (y). The fact that the affair value being modeled is a double value (0 or 1) rather than a numeric value is a significant distinction from direct retrogression. Here is an illustration of a predictor formula:

y = e^ (b0 + b1*x) / (1 + e^ (b0 + b1*x))

If b0 is really the biases or blocks phrase, b1 is the measure for the single input value, and y is the predicted event (x). The b measure (a constant real number) for each column in your input data must be determined from your training set. The components in the solution (the beta value or b's) are the factual representation of the model that you would store in memory or on a train.

```
if βkj≥O
computer Δvkj by formula(8)with s=1
if βkj+ Δvkj<O (trying to cross over O)
 Δvkj ← -βkj
 endif
endif
if βkj ≤O
computer Δvkj by formula (8)  with s= -1
        if βkj+ Δvkj> O (trying to cross over O)
                  Δvkj ← -βkj
  endif
endif
```

❖ **Decision Tree**

A controlled fluency technique family the Deep Learning technique. The decision tree algorithm can be used to solve retrogression and brace concerns, unlike other supervised literacy techniques. By learning straightforward decision rules derived from prior data, a Decision Tree can be used to create a training model that can be used to predict the class or value of the target variable (training data). In Decision Trees, we begin at the tree's base when predicting a class marker for a record. We contrast the traits on the record with those on the core attribute.

```
Procedure 1: CREATE A TREE
```

Step 1 : second: repeat

Step 2 : maxGain is 0 at position 3.

Step 3 : split A null in position four.

Step 4 : Number five: e Entropy (features)

Step 5 : Do for every Attribute in S.

Step 6 : Information Gain

Step 7: Gain (a,e)

Step 8: If gain exceeds maximum gain,

Number

Step 9: maxGain gain

Step 10: splitA a

Step 11: end if

Step 12: The end of

Step 13: Dividing (S, split A) until each division has been handled, day 14

Step 14: finish the process

5. CONCLUSION

The approach bear three outcome. Initially, a model for predicting sickness traits was created. The anticipated model, that uses the datasets, will predict sickness with 97 percent efficiency by implementing the CART model simply in the case of teenager, minor, and grown up, separately. This outcome exhibit improved quality scrutiny compared to the opposing current screening sickness strategy. Additionally, the proposed model would forecast sickness traits for notice age groups, and existing techniques find challenging to understand. For said real dataset, the findings indicated only fair detection accuracy (77 to 85 percent). The limited scope of the large database was the primary cause of this inferior outcome.

To use the appliance in order to someone to predict the illness traits easily, a simple user add supporting the predicted prediction model has been built for finished users. Since most of the running works largely seen on advance and check the system or plan of prediction models or techniques and have not succeeded in creating any mobile applications for A number of notable existing works were viewed by end users as having been extended to a certain extent. For different age groups, this analysis offers an effective and solid method of addressing sight syndrome. Identifying syndrome traits in children and teenagers is difficult, so the classification of the traits is typically postponed because it is very labour-intensive and time-consuming. With the aid of an autism screening tool, a person can receive guidance at an advance stage that could stops the condition from obtain bad and lower costs attached with a late diagnosis.

REFERENCES

1. American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). Arlington, VA: American Psychiatric Publishing.

2 Biederman, J., Petty, C. R., Monuteaux, M. C., Fried, R., Doyle, A. E., Seidman, L. J., ... & Faraone, S. V. (2010). Adult psychiatric outcomes of girls with attention deficit hyperactivity disorder: 11-year follow-up in a longitudinal case-control study. American Journal of Psychiatry, 167(4), 409-417.

3. Coghill, D., Seth, S., & Matthews, K. (2014). A comprehensive assessment of memory, delay aversion, timing, inhibition, decision making and variability in attention deficit hyperactivity disorder: advancing beyond the three-pathway models. Psychological Medicine, 44(9), 1989-2001.

4. DuPaul, G. J., & Stoner, G. (2014). AUTISM DISORDER PREDICITON in the Schools, Third Edition: Assessment and Intervention Strategies. Guilford Publications.

5. Gao, X., & Huang, Y. (2019). Feature selection based on correlation coefficient and maximum relevance minimum redundancy for AUTISM DISORDER PREDICITON detection. Journal of Medical Imaging and Health Informatics, 9(6), 1171-1177.

6. Hossain, M. S., & Kabir, E. (2018). Application of SVM and ANN in prediction of heart disease using feature selection. Journal of Medical Systems, 42(7), 117.

7. Kooij, S. J., Bejerot, S., Blackwell, A., Caci, H., Casas-Brugué, M., Carpentier, P. J., ... & Asherson, P. (2010). European consensus statement on diagnosis and treatment of adult AUTISM DISORDER PREDICITON: The European Network Adult AUTISM DISORDER PREDICITON. BMC Psychiatry, 10(1), 67.

8. Kuhn, M., Johnson, K., & Applied Predictive Modeling, M. E. (2013). Applied predictive modeling (Vol. 26). Springer.

9. R Core Team. (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

10. Salam, M. A., Miah, S., & Islam, M. S. (2019). Diagnosis of AUTISM DISORDER PREDICITON using machine learning techniques: a systematic review. Journal of Medical Systems, 43(4), 81.

11. Vapnik, V. N. (2013). The nature of statistical learning theory. Springer science & business media.