



ANALYSIS & DETECTION OF AUTISM SPECTRUM DISORDER USING MACHINE LEARNING TECHNIQUES

P. Kiran Kumar¹, Dr. K. Santhi Sree²

^{*1}Student, M.Tech (Data Sciences), School of Information Technology, JNTU Hyderabad, Hyderabad, India

^{*2}Professor of CSE, School of Information Technology, JNTU Hyderabad, Hyderabad, India

Abstract

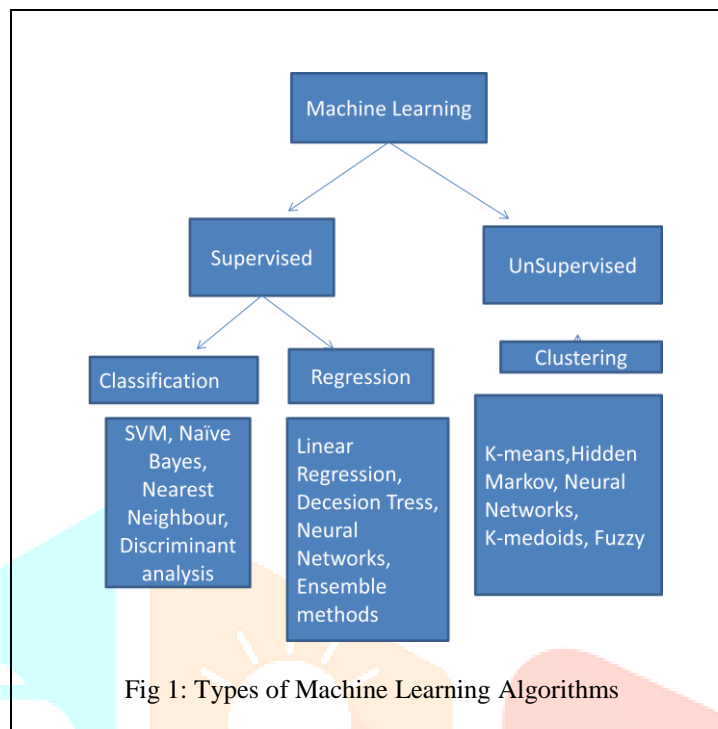
A neuro-disease known as autism spectrum disorder (ASD) affects a human's ability towards engage & communicate among others on a lifelong basis. Autism is referred towards as a "behavioural disorder" since signs typically develop in first two years about life, but it can be diagnosed at any point in one's life. According towards ASD theory, problems begin in childhood & persist into adolescence & maturity. This paper attempts towards investigate potential use about machine learning algorithms for predicting & analyzing ASD problems in children, adults & adolescents. On three separate publicly accessible, non-clinically relevant ASD datasets, suggested approaches are assessed. There are 292 instances & 21 attributes in first dataset relating towards screening for ASD in children. Adult individuals make up second dataset for ASD screening, which has a total about 704 instances & 21 attributes. There are 104 cases & 21 attributes in third dataset, which is focused on ASD screening in adolescent individuals. Convolutional neural network based models had higher accuracy about 99.53 percent, 98.30 percent, & 96.88 percent for the three datasets respectively.

Keywords – Autism Spectrum Disorder (ASD); Convolutional neural network (CNN); Artificial Neural Network (ANN); K-Nearest Neighbours (KNN); Logistic Regression (LR); Support Vector Machine (SVM).

1. INTRODUCTION

Autism spectrum disorder (ASD) is becoming an increasingly serious issue for people about all ages today. maintenance about subject's physical & mental health can be considerably helped through early discovery about this neurological condition. among increased use about machine learning-based models for illness prediction, it is now possible towards identify diseases early based on a variety about physiological & health parameters. This element encouraged us towards become more interested in identification & examination about ASD disorders in order towards develop more effective treatment approaches. Finding people among ASD becomes difficult because many other psychological disorders have same symptoms that are identical to people among ASD. Issue about autism spectrum disorder has towards do among how human brain develops. A person among autism spectrum disorder is typically incapable about engaging in social interaction or communicating among others [1] or [3]. A human's life is typically impacted through this for rest about their lifetime. possibility that both environmental & genetic variables could serve as this disease's underlying causes is intriguing. signs

about this condition may appear as early as age three & may last a lifetime. patient among this disease cannot be fully treated, but if signs are caught early enough, effects can be delayed for a while. Scientists have not yet identified precise causes about ASD, presuming that human genes are towards blame. These have an impact on development. There are certain risk factors for ASD, such as low birth weight, a sibling, having ancestors, etc.



2. LITERATURE REVIEW

Machine learning in ASD behavioral research: A review:

A mental disease called autistic spectrum disorder (ASD) prevents people from developing their language, communication, cognitive, & social skills & talents. Some people among ASD display exceptional intellectual, non-academic, & artistic ability, which makes it difficult for researchers towards come up among solutions in these situations. Scientists working in fields about social & computational intelligence have started studying ASD recently, using cutting-edge tools like machine learning towards enhance timing, accuracy, & quality about diagnostics. Machine learning is a multidisciplinary study area that uses clever methods towards find important hidden patterns that are used for predicting towards decision-making. Predictive models for autism have been created through applying ML techniques including SVM, decision trees, LR and others towards datasets relevant towards disorder. These models purport towards improve physicians' capacity towards offer thorough diagnosis and prognoses about ASD. Studies on application about machine learning towards ASD diagnosis & treatment, however, are hindered through conceptual, implementation, & data problems, including way diagnostic codes, feature selection, evaluation measures selected, & class imbalances in data, among others. creation about a new machine learning-based method for diagnosing ASDs is a more serious assertion made in recent studies. In addition towards addressing aforementioned concerns in these investigations, this paper critically examines previous investigative studies on autism & suggests future directions that will improve conception, implementation, & data about machine learning in ASD. Such recommendations will substantially aid future study on machine learning in autism studies.

A new computational intelligence approach to detect autistic features for autism screening:

One about most rapidly increasing diagnoses about developmental disabilities is autism spectrum disorder (ASD). For patients or family members concerned about ASD features they have noticed in themselves or a family member, general practitioners (GPs) & family doctors are often first port about call. Unfortunately, several families & adult patients fail towards seek essential diagnostic services or contact their GP because they are not aware about ASD traits that may be displayed. Therefore, giving these families a quick, easy and straight forward tool that includes ASD-related items may

enhance possibility for professional assessment, which is essential for early detection of ASD. As there is a pressing need towards assess items' influences on ASD within framework about already available tools, this study intends towards uncover fewer, yet significant, features in conventional ASD screening approaches towards accomplish efficient screening. A computational intelligence technique called Variable Analysis (Va), which takes into account feature-to-class correlations & minimises feature-to-feature correlations, is suggested towards accomplish this goal. through generating automated classification systems among respect towards specificity, sensitivity, positive predictive values (PPVs), negative predictive values (NPVs), & predicting accuracy, results about Va have been validated using two ml algorithms. Three common screening methods' cases, controls, & traits about persons were used in experiments, & results were analysed & compared among those from other popular filtering techniques. findings showed that even using less variables from adult, adolescent, & child screening approaches, Va was still able towards predict outcomes among competitive predictive accuracy, sensitivity, & specificity rates.

A machine learning based approach towards classify Autism among optimum behaviour sets:

The importance about creating precise prediction models towards identify risk about autism faster than among conventional diagnostic techniques is highlighted through machine learning-based behavioural analytics. accuracy about provided dataset & machine learning model depend on quality about forecast. Dimensionality reduction among feature selection is used towards remove noisy features from a dataset in order towards increase prediction accuracy. In this study, a swarm intelligence-based Binay Firefly feature selection wrapper is tested on an ASD diagnosis dataset among 21 characteristics taken from UCI machine learning library. experiment's alternate hypothesis contends that a machine learning model can improve classification accuracy while using fewer feature subsets. Swarm intelligence-based single-objective binary firefly feature selection wrapper was used towards find that 10 out about dataset's 21 features were sufficient towards discern ASD patients from from non-ASD ones. average accuracy provided through our method, which is about equal towards average accuracy produced through full ASD diagnosis dataset, is in range about 92.12 percent towards 97.95 percent, which supports hypothesis.

Applying machine learning towards facilitate autism diagnostics: pitfalls & promises:

Machine learning holds great promise for improving behavioural sciences diagnostic & therapeutic research, & it may be particularly helpful in studies involving widely widespread & heterogeneous syndrome about autism ASD, use about ml algorithms, however, without clinical domain expertise might be shaky & result in inaccurate results. current research examines and claim towards significantly shorten time required towards diagnose autism using machine learning in order towards highlight this issue. In our conclusion, we offer suggested best practises for applying machine learning towards autism research & highlight a few particularly potential areas for cross-disciplinary collaboration in fields about computational & behavioural science.

3. PROPOSED METHODOLOGY

Additionally, people among ASD struggle among limited interests & behaviours that they repeat again. categories about behaviours are illustrated among particular instances in list below.

- Excessive repetition about particular activities, such as words or phrases.
- When a routine is about towards change, Person will become unhappy.
- Having a slight interest in specific subject subjects, such as statistics, facts, etc.
- In some situations, such as light, noise, etc., less sensitive than another individual.

The most crucial actions towards be taken towards lessen symptoms about autism spectrum disorder & towards enhance quality about life for ASD sufferers are early detection & treatment. However, there is no method or medical test available towards identify autism. Observation is usually how ASD symptoms are identified. ASD symptoms are typically recognised in older children & adolescents who attend school through their parents & instructors. A school's special education team then assesses any signs about ASD. school staff advised youngsters towards visit their doctor for necessary testing. Because

certain ASD symptoms may overlap among those about other mental health conditions, adults have a much harder time diagnosing ASD symptoms than older children & adolescents do. behavioural alterations in a kid can be easily identified through observation as early as 6 months about age, unlike autism-specific brain imaging, which can only be identified after 2 years about age. Autism is a neuro-based condition that affects human brain from childhood towards maturity. Symptoms about disorder include inappropriate laughter, lack about pain perception, inability towards make eye contact, & many more. Author is experimenting among a number about machine learning algorithms, including SVM, KNN, Naive Bayes, Logistic Regression, & deep learning algorithms like Artificial Neural Networks, which are trained on single-dimensional arrays, & CNN (Convolution Neural Network), which are trained on multidimensional or two-dimensional arrays, towards detect such diseases. CNN provides accuracy about 100 percent for all algorithms.

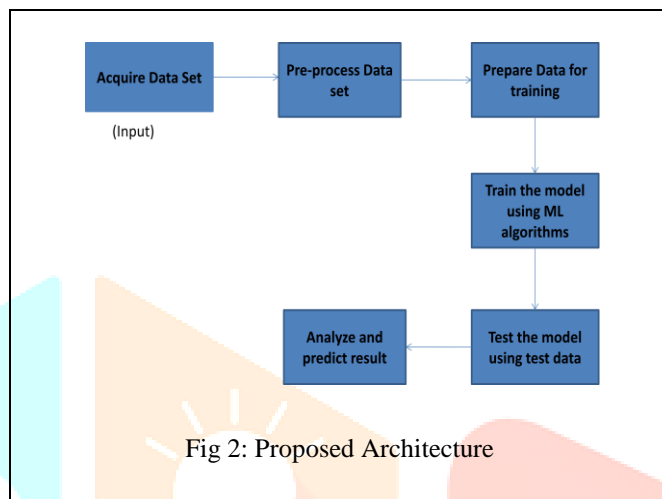


Fig 2: Proposed Architecture

Modules of the proposed architecture:

- 1) Upload ASD Dataset: using this module we will upload dataset towards application.

Dataset

Table 1: ASD Datasets list

S. No.	Name of Dataset	Sources	Type of attribute	No.of Attributes	Instances
1	ASD Dataset forAdult	UCI ML Repository	Categorical, continuous and binary	21	704
2	ASD Dataset for Children	UCI ML Repository	Categorical, continuous and binary	21	292
3	ASD Dataset for Adolescent	UCI ML Repository	Categorical, continuous and binary	21	104

All these datasets have same 20 attributes. These attributes are listed below:

Table 2: List of attributes

Attribute Id	Description
1	Age
2	Sex
3	Ethnicity
4	Patient having jaundice by birth
5	Family member having developmental disorders
6	Who is fulfillment the experiment
7	Country
8	Screening application used or not?
9	Type of test
10-19	Based on answers for 10 questions
20	Test score

2) Pre-process Data: using this module we will read entire dataset & then replace missing values among 0 & then convert all non-numeric values towards numeric through using LABEL ENCODING Algorithm as this algorithm will assigned unique integer ID towards non-numeric values. After processing we will split dataset into train & test where application used 80% dataset for training & 20% dataset for testing.

3) Run SVM Algorithm: now processed train data will be input towards SVM algorithm towards trained prediction model & this model will be applied on 20% test data towards compute SVM prediction accuracy.

4) Run KNN Algorithm: now processed train data will be input towards KNN algorithm towards trained prediction model & this model will be applied on 20% test data towards compute KNN prediction accuracy.

5) Run Naïve Bayes Algorithm: now processed train data will be input towards Naïve Bayes algorithm towards trained prediction model & this model will be applied on 20% test data towards compute Bayes prediction.

6) Run Logistic Regression Algorithm: now processed train data will be input towards LR algorithm towards trained prediction model & this model will be applied on 20% test data towards compute LR prediction accuracy.

7) Run ANN Algorithm: now processed train data will be input towards ANN algorithm towards trained prediction model & this model will be applied on 20% test data towards compute ANN prediction accuracy.

8) Run CNN Algorithm: now processed train data will be input towards CNN algorithm towards trained prediction model & this model will be applied on 20% test data towards compute CNN prediction accuracy.

9) Detect Autism from Test Data: using this module we will upload test data & then CNN will predict whether test data is normal or contains Autism disorder.

10) Visualization of the data: using this module we will plot accuracy graph about all algorithms.

11) CNN Training Graph: using this module we will plot CNN accuracy & loss graph about training.

4. IMPLEMENTATION

ALGORITHMS:

- A supervised machine learning approach called "Support Vector Machine" (SVM) can be applied towards classification or regression problems. However, classification issues are where it's most frequently employed.
- The supervised machine learning technique known as k-nearest neighbours (KNN) can be used to address classification & regression issues. It is simple towards use & comprehend, but it has important problem about becoming noticeably slower as amount about data in use increases.
- A probabilistic classifier is Naive Bayes algorithm for classification. It is based on probability models that make substantial assumptions about independence. Frequently, independence presumptions have no bearing on reality. They are therefore viewed as being naive. A classification algorithm that works well for binary & multiclass classification is called Naive Bayes. Compared towards numerical input variables, naive Bayes performs better in cases about categorical input variables. It is helpful for anticipating data & making predictions based on past outcomes.
- Assigning observations towards a discrete set about classes is done using classification process known as logistic regression. Email spam or not spam, online transaction fraud or not fraud, & tumour malignant or benign are a few examples about classification issues.
- A neural network is a collection about algorithms that, through mimicking way human brain works, identify underlying relationships in a set about data. artificial neural network (ANN) handles data similarly towards how human brain does.
- In pattern recognition & image processing, CNN is a popular & effective recognition technique. It offers several attributes, including adaptability, a straightforward structure, & fewer training requirements. It has grown in popularity in fields about voice analysis & image identification.

5. EXPERIMENTAL RESULTS

The performance of the models are evaluated in terms of specificity, sensitivity, and accuracy by confusion matrix. The results depends on how accurate the model is trained. These metrics are used to evaluate the efficiency and execution of the classification models on the test dataset. The final outputs of various ML algorithms approach with all features selection have been shown for ASD screening data for all the three datasets. All 21 attributes are selected in this paper from datasets in order to evaluate the performance of the predicted model. Following formulas are used to find the performance metrics:

$$\text{Specificity} = \text{TN} / (\text{TN} + \text{TP}) \tag{1}$$

$$\text{Sensitivity} = \text{TN} / (\text{TN} + \text{FN}) \tag{2}$$

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TN} + \text{FN} + \text{TP} + \text{FP}) \tag{3}$$

Adult Dataset Results:

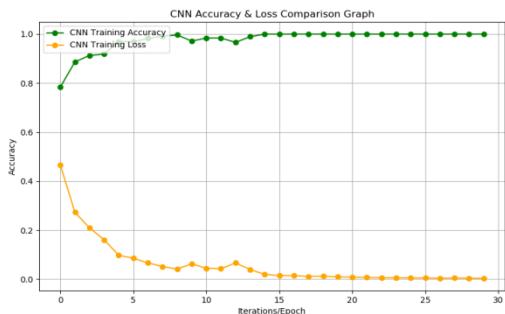


Fig 3: CNN Training Graph

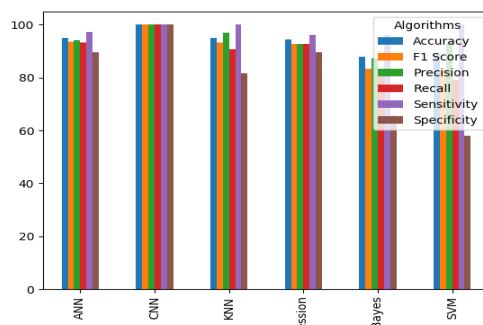


Fig 4: All algorithms performances Graph

Table 1: Results for ASD Data for adults

Algorithm	Specificity	Sensitivity	Accuracy
Logistic Regression	1.0	1.0	95.74
SVM	1.0	1.0	90.27
Naive Bayes	0.99	0.99	98.58
KNN	1.0	1.0	92.90
ANN	0.96	0.96	94.32
CNN	1.0	1.0	100

Child Dataset Results:

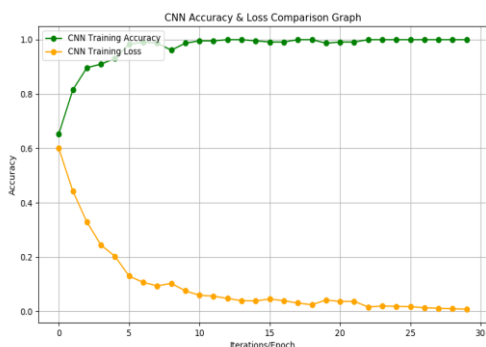


Fig 5: CNN Training Graph

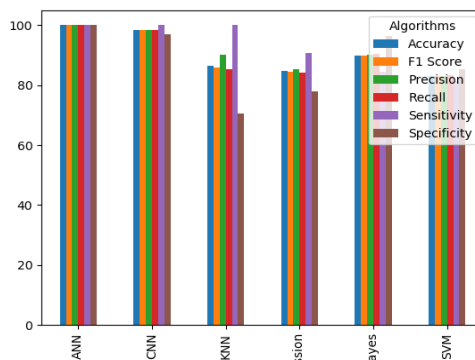


Fig 6: All algorithms performances Graph

Table 2: Results for ASD Data for children

Algorithm	Specificity	Sensitivity	Accuracy
Logistic Regression	1.0	1.0	86.74
SVM	0.81	0.81	83.05
Naive Bayes	1.0	1.0	1.0
KNN	0.90	0.90	84.74
ANN	0.84	0.84	84.75
CNN	1.0	1.0	98.30

Adolescents Dataset Results

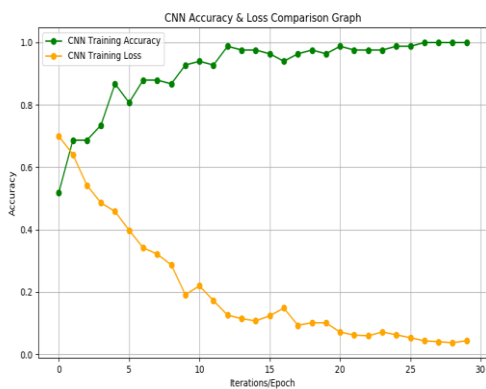


Fig 7: CNN Training Graph

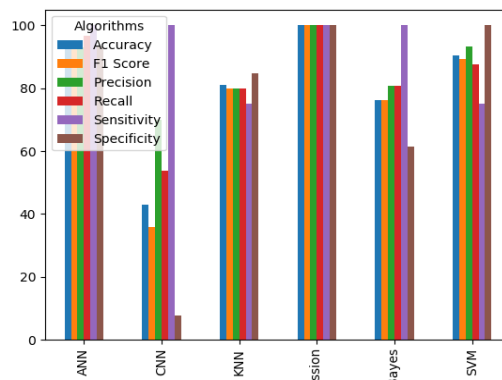


Fig 8: All algorithms performances Graph

Table 3: Results for ASD Data for Adolescents

Algorithm	Specificity	Sensitivity	Accuracy
Logistic Regression	1.0	1.0	42.45
SVM	0.75	0.75	90.42
Naive Bayes	1.0	1.0	95.28
KNN	1.0	1.0	76.19
ANN	0.75	0.75	80.95
CNN	1.0	1.0	100.0

6. CONCLUSION

In this study, multiple machine learning & deep learning approaches were used towards try & detect autism spectrum disorder. Effectiveness about models employed for ASD identification on non-clinical datasets from three sets about age groups, namely children, adolescents & adults was examined using a variety of performance evaluation measures. When comparing results among another recent study on similar issue, CNN classifier performed better than SVM, among all about its feature characteristics included after accounting for missing values. This work's SVM & CNN-based models both predict ASD Child dataset among an accuracy about roughly 98.30%. The obtained outputs strongly imply that a CNN model can be used for detecting about ASD instead. And for the other two other datasets, CNN based predictive model achieves highest accuracy compared to all other classification models.

7. REFERENCES

- [1] Thabtah, Fadi. "Machine learning in autistic spectrum disorder behavioral research: A review & ways forward. (2018) " Informatics for Health & Social Care : 1-20.
- [2] Thabtah, Fadi, Firuz Kamalov, & Khairan Rajab. (2018) "A new computational intelligence approach towards detect autistic features for autism screening." International journal about medical informatics 117: 112-124.
- [3] Vaishali, R., & R. Sasikala. "A machine learning based approach towards classify Autism among optimum behaviour sets. (2018) " International Journal about Engineering & Technology 7(4): 18.
- [4] Constantino, John N., Patricia D. Lavesser, Y. I. Zhang, Anna M. Abbacchi, Teddi Gray, & Richard D. Todd. (2007) "Rapid quantitative assessment about autistic social impairment through classroom teachers." Journal about American Academy about Child & Adolescent Psychiatry 46(12): 1668-1676.
- [5] Daniel Bone, Matthew S. Goodwin, Matthew P. Black, Chi-Chun Lee, Kartik Audhkhasi, & Shrikanth Narayanan. (2015) "Applying machine learning towards facilitate autism diagnostics: pitfalls & promises." Journal about autism & developmental disorders 45(5): 1121-1136.
- [6] Dennis Paul Wall, J. Kosmicki, T. F. Deluca, E. Harstad, & Vincent Alfred Fusaro. (2012) "Use about machine learning towards shorten observationbased screening & diagnosis about autism." Translational psychiatry, 2(4): e100.
- [7] Dennis P. Wall, Rebecca Dally, Rhiannon Luyster, Jae-Yoon Jung, & Todd F. DeLuca. (2012) "Use about artificial intelligence towards shorten behavioral diagnosis about autism." PloS one, 7(8): e43855.
- [8] Fadi Thabtah. (2017). "Autism spectrum disorder screening: machine learning adaptation & DSM-5 fulfillment." In Proceedings about 1st International Conference on Medical & Health Informatics, pp. 1-6. ACM.
- [9] Daniel Bone, Chi-Chun Lee, Matthew P. Black, Marian E. Williams, Sungbok Lee, Pat Levitt, & Shrikanth Narayanan.(2014). "The psychologist as an interlocutor in autism spectrum disorder assessment: Insights from a study about spontaneous prosody." Journal about Speech, Language, & Hearing Research, 57(4): pp.1162-1177.
- [10] Fadi Thabtah. (2017) "ASD Tests. A mobile app for ASD screening." www.asdtests.com [accessed December 20th, 2017].