



DETECTION OF AUTISM SPECTRUM DISORDER USING MACHINE LEARNING

Dr.N. Satyavathi, Head of the Department CSE, Vaagdevi College of Engineering (Autonomous),
India

Kallem Tharun, UG Student, CSE, Vaagdevi College of Engineering (Autonomous), India,

Gante Pranay, UG Student, CSE, Vaagdevi College of Engineering (Autonomous), India

Kethe Dev Chand, UG Student, CSE, Vaagdevi College of Engineering (Autonomous), India

Paruchuri Surya Prakash, UG Student, CSE, Vaagdevi College of Engineering (Autonomous),
India

ABSTRACT

Autism Spectrum Disorder (ASD) is a neuro-disorder in which a person has a lifelong effect on interaction and communication with others. Autism can be diagnosed at any stage in once life and is said to be a "behavioral disease" because in the first two years of life symptoms usually appear. According to the ASD problem starts with childhood and continues to keep going on into adolescence and adulthood. Propelled with the rise in use of machine learning techniques in the research dimensions of medical diagnosis, in this paper there is an attempt to explore the possibility to use Naïve Bayes, Support Vector Machine, Logistic Regression, KNN, Neural Network and Convolutional Neural Network for predicting and analysis of ASD problems in a child, adolescents, and adults. The proposed techniques are evaluated on publicly available three different non-clinically ASD datasets. First dataset related to ASD screening in children has 292 instances and 21 attributes. Second dataset related to ASD screening Adult subjects contains a total of 704 instances and 21 attributes. Third dataset related to ASD screening in Adolescent subjects comprises of 104 instances and 21 attributes. After applying various machine learning techniques and handling missing values, results strongly suggest that CNN based prediction models work better on all these datasets with higher accuracy of 99.53%, 98.30%, 96.88% for Autistic Spectrum Disorder Screening in Data for Adult, Children, and Adolescents respectively.

1. INTRODUCTION

The problem of autism spectrum disorder (ASD) have been mounting swiftly nowadays among all ages of the human population. Early detection of this neurological disease can greatly assist in the maintenance of the subject's mental and physical health. With the rise of application of machine learning-based models in the predictions of various human diseases, their early detection based on various health and physiological parameter now seems possible. This factor motivated us to increase interest in the detection and analysis of ASD [1] diseases to improve better treatment methodology. Detection of ASD becomes a challenge as there are several other mental disorders whose few symptoms are very similar to those with ASD symptoms, thereby makes this task a difficult one. Autism Spectrum disorder is a problem that is related to human brain development. A person who has suffered from the Autism Spectrum Disorder is generally not able to do social interaction and communication with other persons [1] - [3]. In this, a person's life is usually affected for his or her entire lifetime. It is interesting to know that both environmental and genetic factors may turn out to be the causing factors for this disease. The symptoms of this problem may be started at the age of three years and may continue for the lifetime. It is not possible to complete treat the patient suffering from this disease, however its effects can be reduced for some time if the symptoms are early detected. By assuming that human genes are responsible for it, the exact causes of ASD [4] have not been recognized by the scientist yet. The human genes affect the development by influencing the environment. Early detection and treatment are most important steps to be taken to decrease the symptoms of autism spectrum disorder problem and to improve the quality of life of ASD suffering people. However, there is no procedure of medical test for detection of autism. ASD Symptoms usually recognized by



observation. In Older and adolescents who go to school, ASD symptoms are usually identified by their parents and teachers. After that ASD symptoms are evaluated by a special education team of the school. These school team suggested these children visit their health care doctor for required testing. In adults identifying ASD symptoms is very difficult than older children and adolescents because some symptoms[5],[6] of ASD may be overlap with other mental health disorders. It is easy to identify the behavioral changes in a child easily by observation because it can be seen early in the 6 months of age than Autism specific brain imaging because brain imaging can be identifying after 2 years of age.

2. LITERATURE SURVEY

Autistic Spectrum Disorder (ASD) is a mental disorder that retards acquisition of linguistic, communication, cognitive, and social skills and abilities. Despite being diagnosed with ASD, some individuals exhibit outstanding scholastic, non-academic, and artistic capabilities, in such cases posing a challenging task for scientists to provide answers. In the last few years, ASD [7] has been investigated by social and computational intelligence scientists utilizing advanced technologies such as machine learning to improve diagnostic timing, precision, and quality. Machine learning is a multidisciplinary research topic that employs intelligent techniques to discover useful concealed patterns, which are utilized in prediction to improve decision making. Machine learning techniques such as support vector machines, decision trees, logistic regressions, and others, have been applied to datasets related to autism in order to construct predictive models. These models claim to enhance the ability of clinicians to provide robust diagnoses and prognoses of ASD. However, studies concerning the use of machine learning in ASD diagnosis and treatment suffer from conceptual, implementation, and data issues such as the way diagnostic codes are used, the type of feature selection employed, the evaluation measures chosen, and class imbalances in data among others. A more serious claim in recent studies is the development of a new method for ASD diagnoses based on machine learning. This article critically analyses these recent investigative studies on autism, not only articulating the aforementioned issues [8] in these studies but also recommending paths forward that enhance machine learning use in ASD with respect to conceptualization, implementation, and data. Future studies concerning machine learning in autism research are greatly benefitted by such proposals.

Autism Spectrum Disorder (ASD) is one of the fastest growing developmental disability diagnosis. General practitioners (GPs) [9] and family physicians are typically the first point of contact for patients or family members concerned with ASD traits observed in themselves or their family member. Unfortunately, some families and adult patients are unaware of ASD traits that may be exhibited and as a result do not seek out necessary diagnostic services or contact their GP. Therefore, providing a quick, accessible, and simple tool utilizing items related to ASD to these families may increase the likelihood they will seek professional assessment and is vital to the early detection and treatment of ASD. This study aims at identifying fewer, albeit

influential, features in common ASD screening methods in order to achieve efficient screening as demands on evaluating the items' influences on ASD within existing tools is urgent. To achieve this aim, a computational intelligence method called Variable Analysis (Va) [10] is proposed that considers feature-to-class correlations and reduces feature-to-feature correlations. The results of the Va have been verified using two machine learning algorithms by deriving automated classification systems with respect to specificity, sensitivity, positive predictive values (PPVs), negative predictive values (NPVs), and predictive accuracy. Experimental results using cases and controls related to items in three common screening methods, along with features related to individuals, have been analysed and compared with results obtained from other common filtering methods. The results exhibited that Va was able to derive fewer numbers of features from adult, adolescent, and child screening methods yet maintained competitive predictive accuracy, sensitivity, and specificity rates.

Machine Learning based behavioural analytics emphasis the need to develop accurate prediction models for detecting the risk of autism faster than the traditional diagnostic methods. Quality of prediction rely on the accuracy of the supplied dataset and the machine learning model. To improve



accuracy of prediction, dimensionality reduction with feature selection is applied to eliminate noisy features from a dataset. In this work an ASD diagnosis dataset with 21 features obtained from UCI machine learning repository is experimented with swarm intelligence based binary firefly feature selection wrapper. The alternative hypothesis of the experiment claims that it is possible for a machine learning model to achieve a better classification accuracy with minimum feature subsets. Using Swarm intelligence based single-objective binary firefly feature selection wrapper it is found that 10 features among 21 features of ASD dataset are sufficient to distinguish between ASD [11] and non-ASD patients. The results obtained with our approach justifies the hypothesis by producing an average accuracy in the range of 92.12%-97.95% with optimum feature subsets which is approximately equal to the average accuracy produced by entire ASD diagnosis dataset.

Machine Learning based behavioural analytics emphasis the need to develop accurate prediction models for detecting the risk of autism faster than the traditional diagnostic methods. Quality of prediction rely on the accuracy of the supplied dataset and the machine learning model. To improve accuracy of prediction, dimensionality reduction with feature selection is applied to eliminate noisy features from a dataset. In this work an ASD [12] diagnosis dataset with 21 features obtained from UCI machine learning repository is experimented with swarm intelligence based binary firefly feature selection wrapper. The alternative hypothesis of the experiment claims that it is possible for a machine learning model to achieve a better classification accuracy with minimum feature subsets. Using Swarm intelligence based single-objective binary firefly feature selection wrapper it is found that 10 features among 21 features of ASD dataset are sufficient to distinguish between ASD and non-ASD patients. The results obtained with our approach justifies the hypothesis by producing an average accuracy in the range of 92.12%-97.95% with optimum feature subsets which is approximately equal to the average accuracy produced by entire ASD diagnosis dataset.

Machine learning has immense potential to enhance diagnostic and intervention research in the behavioral sciences, and may be especially useful in investigations involving the highly prevalent and heterogeneous syndrome of autism spectrum [13] disorder. However, use of machine learning in the absence of clinical domain expertise can be tenuous and lead to misinformed conclusions. To illustrate this concern, the current paper critically evaluates and attempts to reproduce results from two studies (Wall et al. in *Transl Psychiatry* 2(4):e100, 2012a; *PloS One* 7(8), 2012b) that claim to drastically reduce time to diagnose autism using machine learning. Our failure to generate comparable findings to those reported by Wall and colleagues using larger and more balanced data underscores several conceptual and methodological problems associated with these studies. We conclude with proposed best-practices when using machine learning in autism research, and highlight some especially promising areas for collaborative work at the intersection of computational and behavioral science.

The Autism Diagnostic Observation Schedule-Generic (ADOS) [15] is one of the most widely used instruments for behavioral evaluation of autism spectrum disorders. It is composed of four modules, each tailored for a specific group of individuals based on their language and developmental level. On average, a module takes between 30 and 60 min to deliver. We used a series of machine-learning algorithms to study the complete set of scores from Module 1 of the ADOS available at the Autism Genetic Resource Exchange (AGRE) for 612 individuals with a classification of autism and 15 non-spectrum individuals from both AGRE and the Boston Autism Consortium (AC). Our analysis indicated that 8 of the 29 items contained in Module 1 of the ADOS were sufficient to classify autism with 100% accuracy. We further validated the accuracy of this eight-item classifier against complete sets of scores from two independent sources, a collection of 110 individuals with autism from AC and a collection of 336 individuals with autism from the Simons Foundation. In both cases, our classifier performed with nearly 100% sensitivity, correctly classifying all but two of the individuals from these two resources with a diagnosis of autism, and with 94% specificity on a collection of observed and simulated non-spectrum controls. The classifier contained several elements found in the ADOS algorithm, demonstrating high test validity, and also resulted in a quantitative score that measures



classification confidence and extremeness of the phenotype. With incidence rates rising, the ability to classify autism effectively and quickly requires careful design of assessment and diagnostic tools. Given the brevity, accuracy and quantitative nature of the classifier, results from this study may prove valuable in the development of mobile tools for preliminary evaluation and clinical prioritization--in particular those focused on assessment of short home videos of children--that speed the pace of initial evaluation and broaden the reach to a significantly larger percentage of the population at risk.

The Autism Diagnostic Interview-Revised (ADI-R) is one of the most commonly used instruments for assisting in the behavioral diagnosis of autism. The exam consists of 93 questions that must be answered by a care provider within a focused session that often spans 2.5 hours. We used machine learning techniques to study the complete sets of answers to the ADI-R available at the Autism Genetic Research Exchange (AGRE) for 891 individuals diagnosed with autism and 75 individuals who did not meet the criteria for an autism diagnosis. Our analysis showed that 7 of the 93 items contained in the ADI-R were sufficient to classify autism with 99.9% statistical accuracy. We further tested the accuracy of this 7-question classifier against complete sets of answers from two independent sources, a collection of [15] 1654 individuals with autism from the Simons Foundation and a collection of 322 individuals with autism from the Boston Autism Consortium. In both cases, our classifier performed with nearly 100% statistical accuracy, properly categorizing all but one of the individuals from these two resources who previously had been diagnosed with autism through the standard ADI-R. Our ability to measure specificity was limited by the small numbers of non-spectrum cases in the research data used, however, both real and simulated data demonstrated a range in specificity

from 99% to 93.8%. With incidence rates rising, the capacity to diagnose autism quickly and effectively requires careful design of behavioral assessment methods. Ours is an initial attempt to retrospectively analyze large data repositories to derive an accurate, but significantly abbreviated approach that may be used for rapid detection and clinical prioritization of individuals likely to have an autism spectrum disorder. Such a tool could assist in streamlining the clinical diagnostic process overall, leading to faster screening and earlier treatment of individuals with autism.

One of the primary psychiatric disorders is Autistic Spectrum Disorder (ASD). ASD is a mental disorder that limits the use of linguistic, communicative, cognitive, skills as well as social skills and abilities. Recently, ASD has been studied in the behavioural sciences using intelligent methods based around machine learning to speed up the screening time or to improve sensitivity, specificity or accuracy of the diagnosis process. Machine learning considers the ASD diagnosis problem as a classification task in which predictive models are built based on historical cases and controls. These models are supposed to be plugged into a screening tool to accomplish one or more of the aforementioned goals. In this paper, we shed light on recent studies that employ machine learning in ASD classification in order to discuss their pros and cons. Moreover, we highlight a noticeable problem associated with current ASD [3] screening tools; the reliability of these tools using the DSM-IV rather than the DSM- manual. Hence the necessity to amend current screening tools to reflect the new imposed criteria of ASD classification in the DSM-5 particularly the diagnostic algorithms embedded within these methods.

The purpose of this study was to examine relationships between prosodic speech cues and autism spectrum disorder (ASD) severity, hypothesizing a mutually interactive relationship between the speech characteristics of the psychologist and the child. The authors objectively quantified acoustic-prosodic cues of the psychologist and of the child with ASD during spontaneous interaction, establishing a methodology for future large-sample analysis. Method: Speech acoustic-prosodic features were semiautomatically derived from segments of semi structured interviews (Autism Diagnostic Observation Schedule, ADOS; Lord, Rutter, DiLavore, & Risi, 1999; Lord et al., 2012) with 28 children who had previously been diagnosed with ASD. Prosody was quantified in terms of intonation, volume, rate, and voice quality. Research hypotheses were tested via correlation as well as hierarchical and predictive regression between ADOS [12] severity and prosodic cues. Results: Automatically extracted speech features demonstrated prosodic characteristics of dyadic interactions.

As rated ASD severity increased, both the psychologist and the child demonstrated effects for turn-end pitch slope, and both spoke with atypical voice quality. The psychologist's acoustic cues predicted the child's symptom severity better than did the child's acoustic cues. Conclusion: The psychologist, acting as evaluator and interlocutor, was shown to adjust his or her behavior in predictable ways based on the child's social-communicative impairments. The results support future study of speech prosody of both interaction partners during spontaneous conversation, while using automatic computational methods that allow for scalable analysis on much larger corpora.

3. PROBLEM STATEMENT

Human diseases, their early detection based on various health and physiological parameter now seems possible. This factor motivated us to increase interest in the detection and analysis of ASD diseases to improve better treatment methodology. Detection of ASD [11] becomes a challenge as there are several other mental disorders whose few symptoms are very similar to those with ASD symptoms, thereby makes this task a difficult one.

3.2 LIMITATIONS

- Using machine learning algorithm will not give you better accuracy
- These wrong identification may becomes serious problem

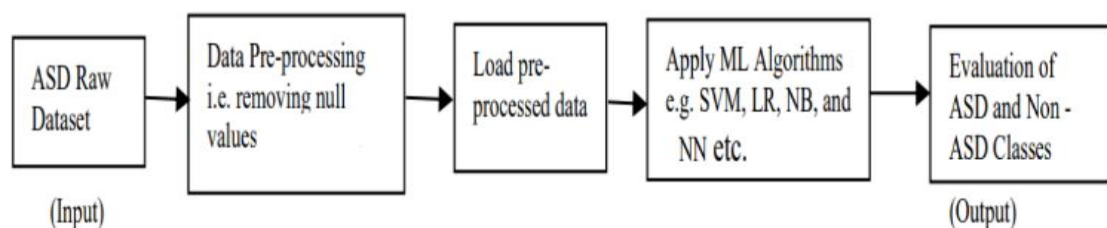
4. PROPOSED SYSTEM

The contents of this paper are organized as follows: Section 1 presents the introduction to the Autism Spectrum Disorder problem and the challenges faced by the subjects. Section 2 presents the review of various recent literature, where some models for ASD [9] detection have been developed. Section 3 describes the datasets used in this study, which is followed by description of each component of the methodology used in this work in section 4.

4.2 ADVANTAGES OF PROPOSED SYSTEM

- We are using deep learning ANN algorithm to get the better accuracy
- Using the ANN model we can accurately identify ASD

5. SYSTEM ARCHITECTURE



6. IMPLEMENTATION

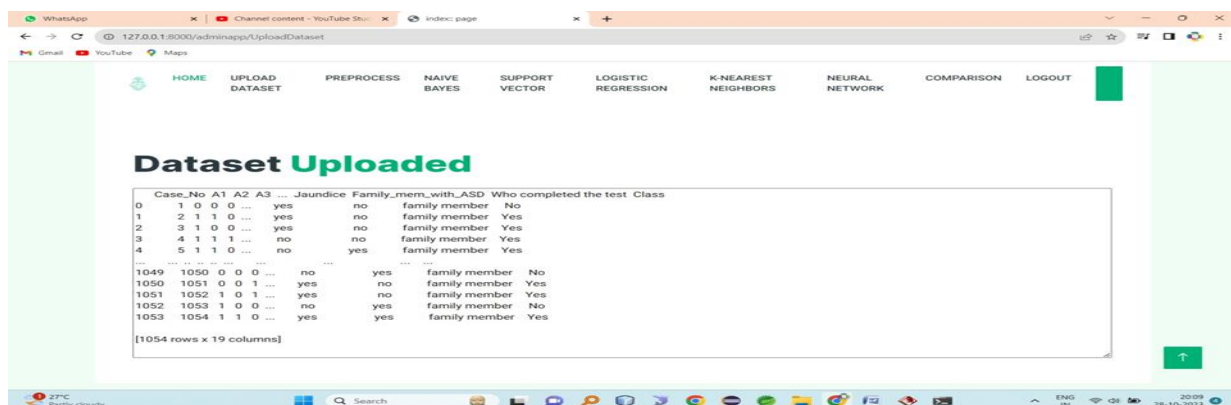
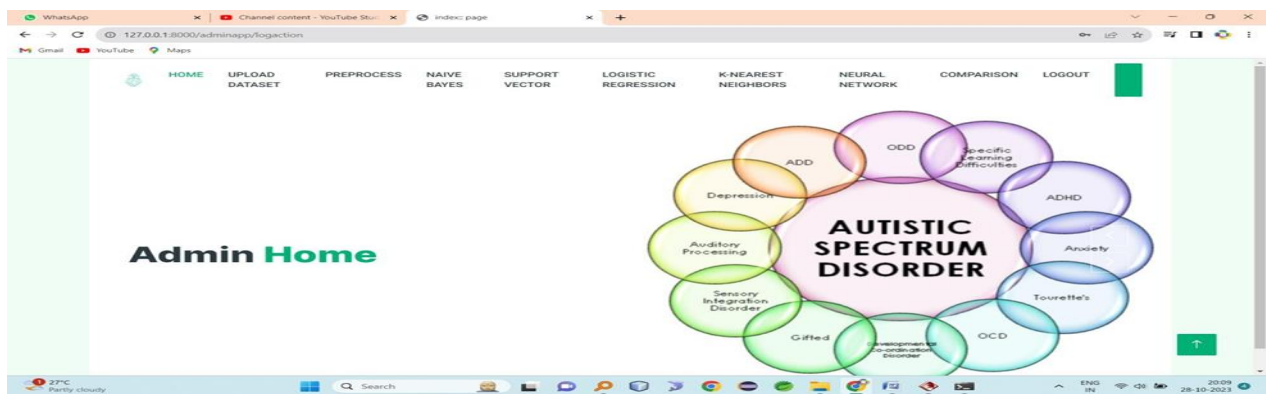
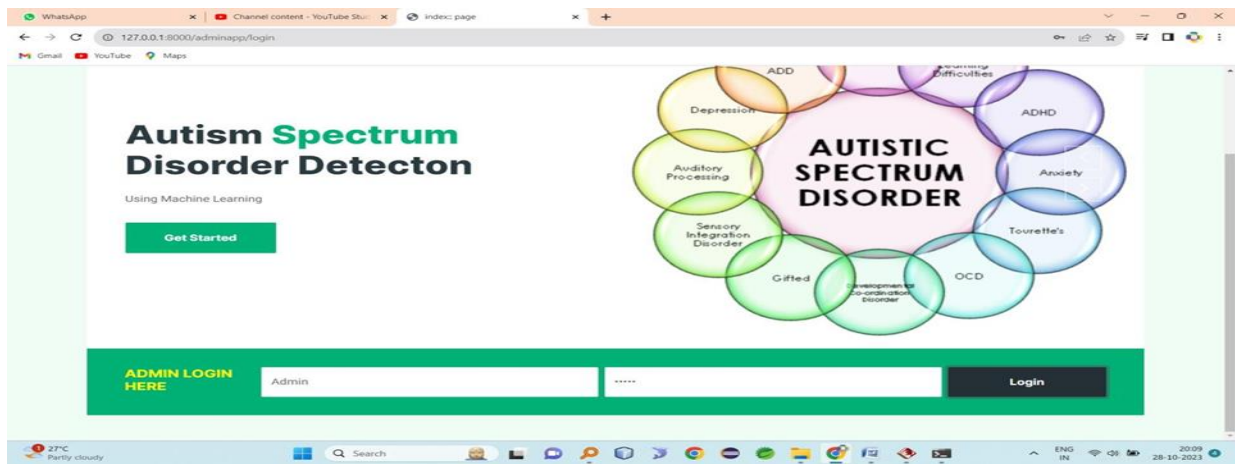
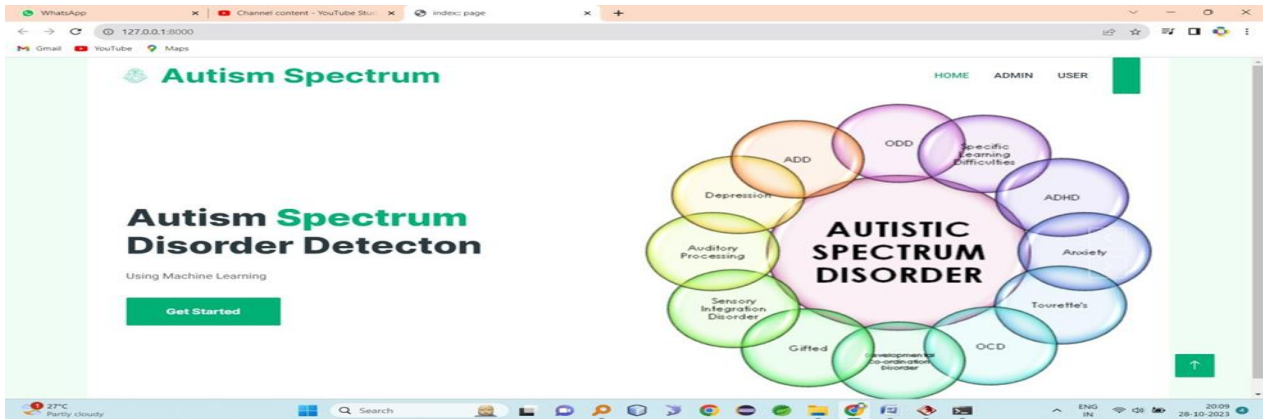
ADMIN

In this application admin is the main module, here admin can login directly with the application and after login successful admin can perform operations such as view Police and activate police, add criminal data and view criminal data. If admin want to delete criminal he have option to delete.

USER

Here user is a module should register with the application after successful register, user should login then user can able to work on some operations such as view profile, detect autism disorder by passing the test data and logout

7. EXPECTED OUTCOMES





Preprocess Has Done

Total dataset size: **1054**
80% data splited as Training: **843**
20% data splited as Training: **211**

Algorithm: Naive Bayes Run Successfully..!!

Accuracy: **98.10426540284361**

Algorithm: Support Vector Machine Run Successfully..!!

Accuracy: **100.0**

Algorithm: Logistic Regression Run Successfully..!!

Accuracy: **100.0**



The screenshot shows a web browser window with the URL `127.0.0.1:8000/adminapp/runKNN`. The page has a navigation menu with items: HOME, UPLOAD DATASET, PREPROCESS, NAIVE BAYES, SUPPORT VECTOR, LOGISTIC REGRESSION, K-NEAREST NEIGHBORS, NEURAL NETWORK, COMPARISON, and LOGOUT. The main content area displays a large green-bordered box with the following text:

Algorithm: K-NearestNeighbor Run Successfully..!!

Accuracy: 98.10426540284361

The browser's taskbar shows the system tray with a temperature of 27°C, weather 'Partly cloudy', and the date '28-10-2023'.

The screenshot shows a terminal window titled `C:\Windows\System32\cmd.exe` overlaid on the web application. The terminal output displays the following data for each epoch:

Epoch	Time	Loss	Accuracy
52/150	0s 33us/step	0.0051	1.0000
53/150	0s 50us/step	0.0049	1.0000
54/150	0s 52us/step	0.0058	1.0000
55/150	0s 35us/step	0.0041	1.0000
56/150	0s 66us/step	0.0038	1.0000
57/150	0s 58us/step	0.0040	1.0000
58/150	0s 52us/step	0.0034	1.0000
59/150	0s 58us/step	0.0040	1.0000
60/150	0s 64us/step	0.0037	1.0000
61/150	0s 58us/step	0.0031	1.0000
62/150	0s 54us/step	0.0027	1.0000
63/150	ETA: 0s	0.0024	1.0000

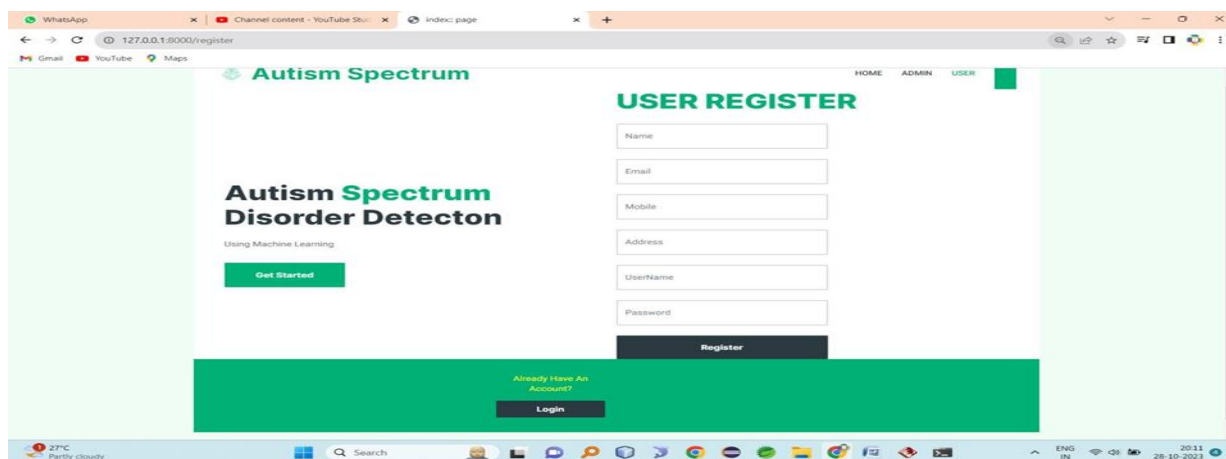
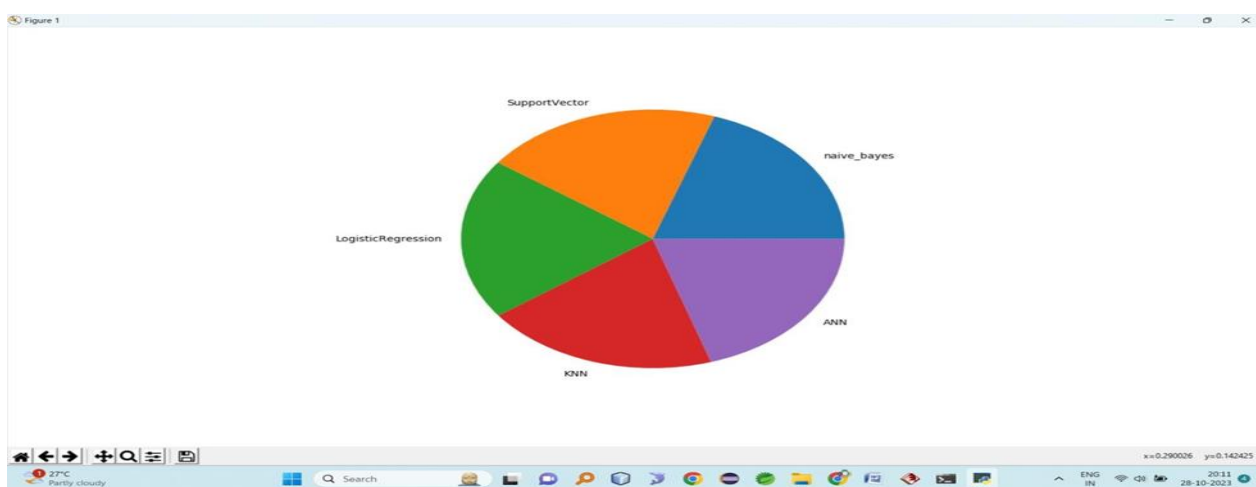
The background shows the same web application interface as the previous screenshot, with the success message partially visible.

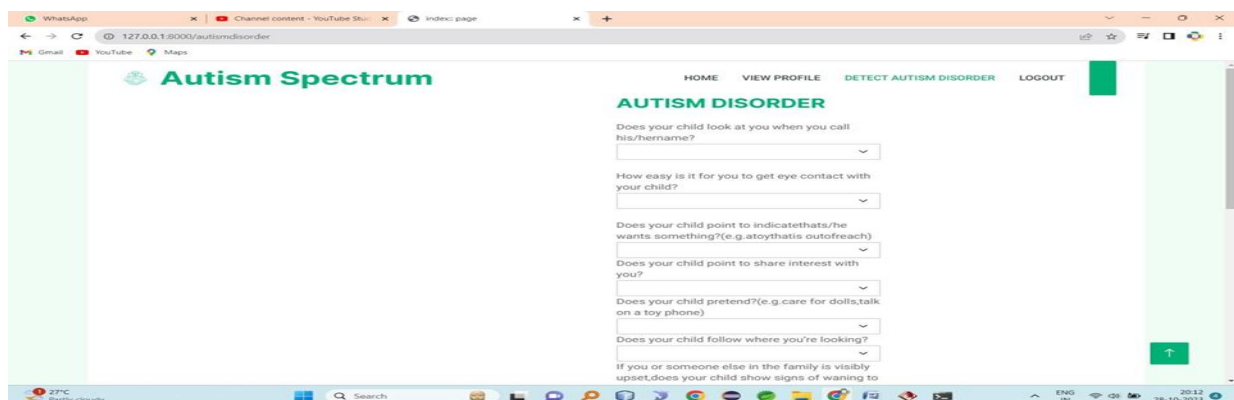
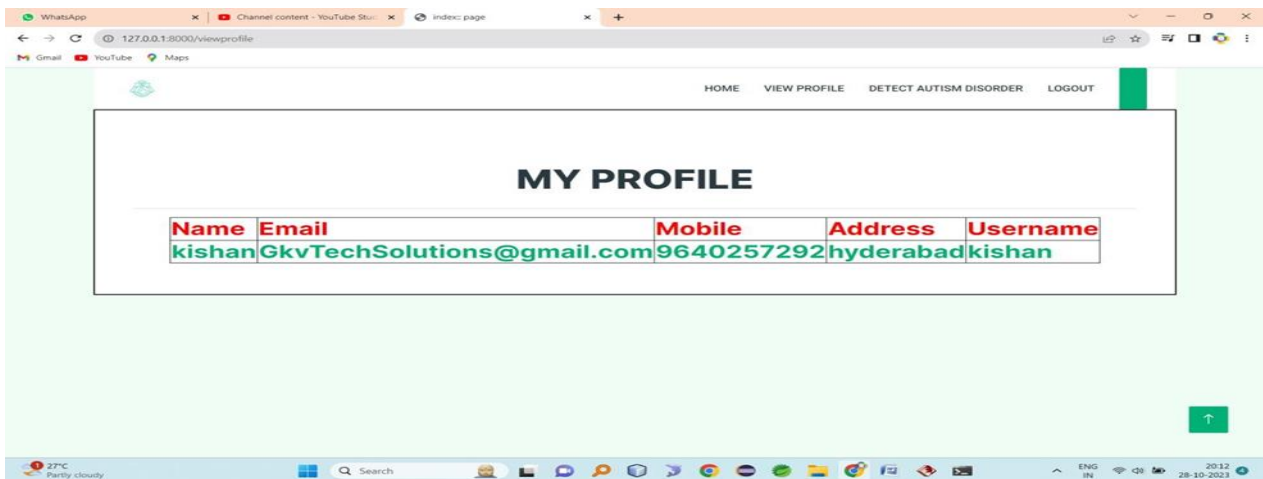
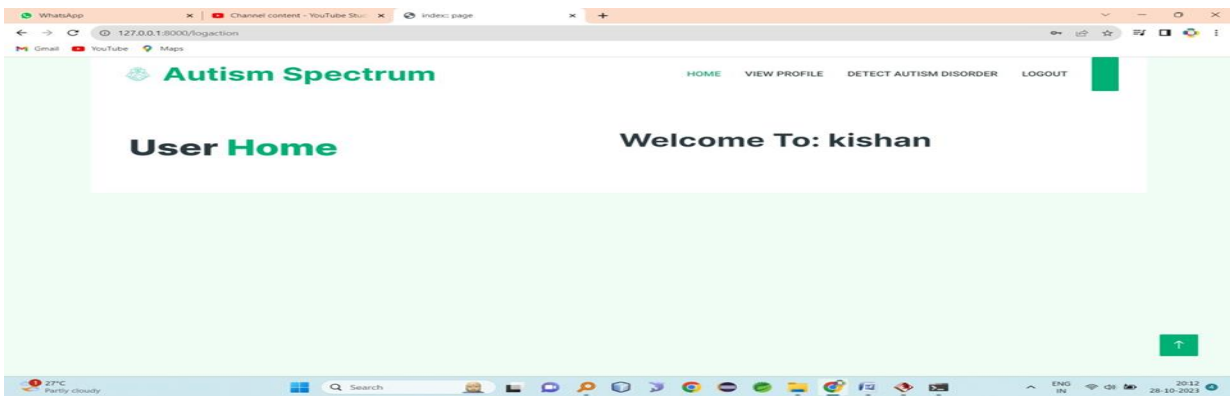
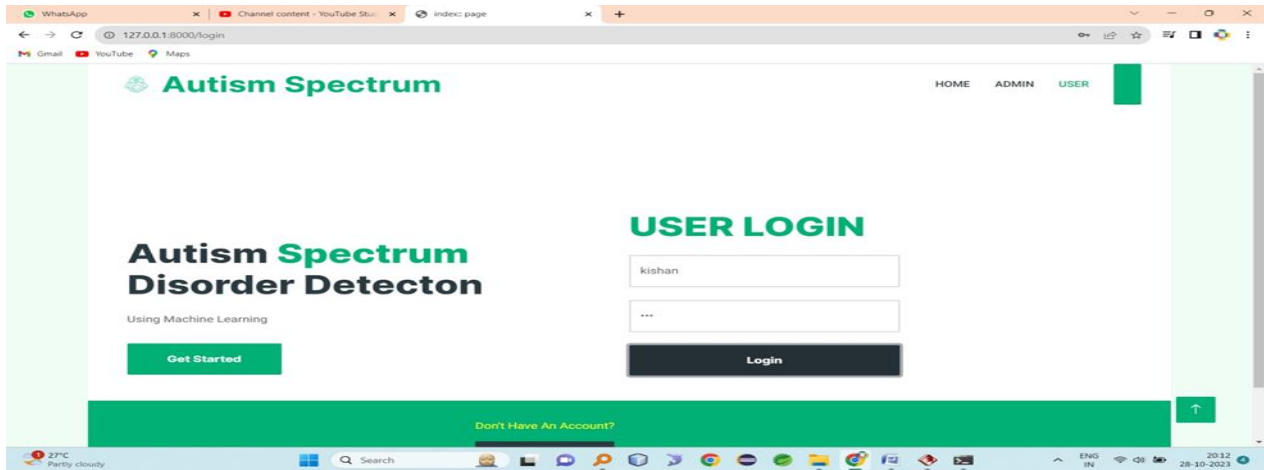
The screenshot shows a web browser window with the URL `127.0.0.1:8000/adminapp/runANN`. The page has the same navigation menu as the previous screenshot. The main content area displays a large green-bordered box with the following text:

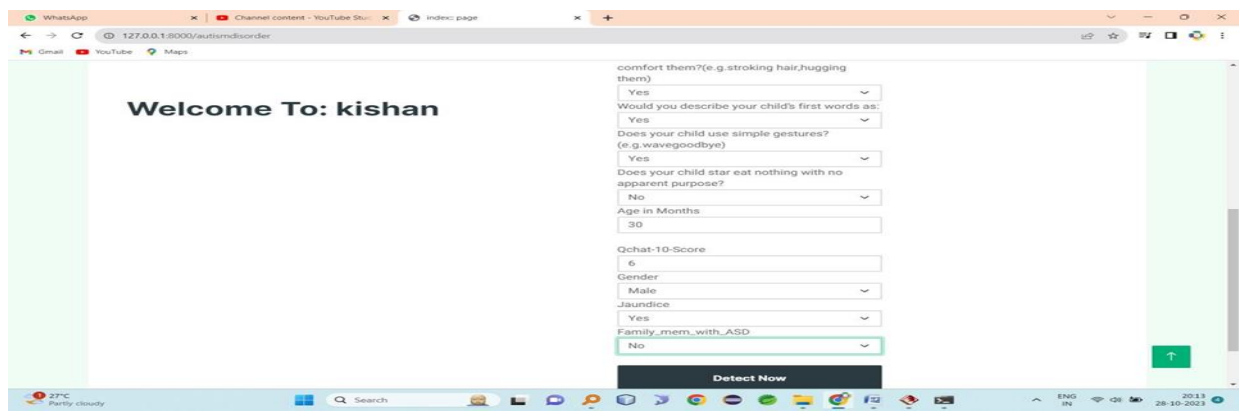
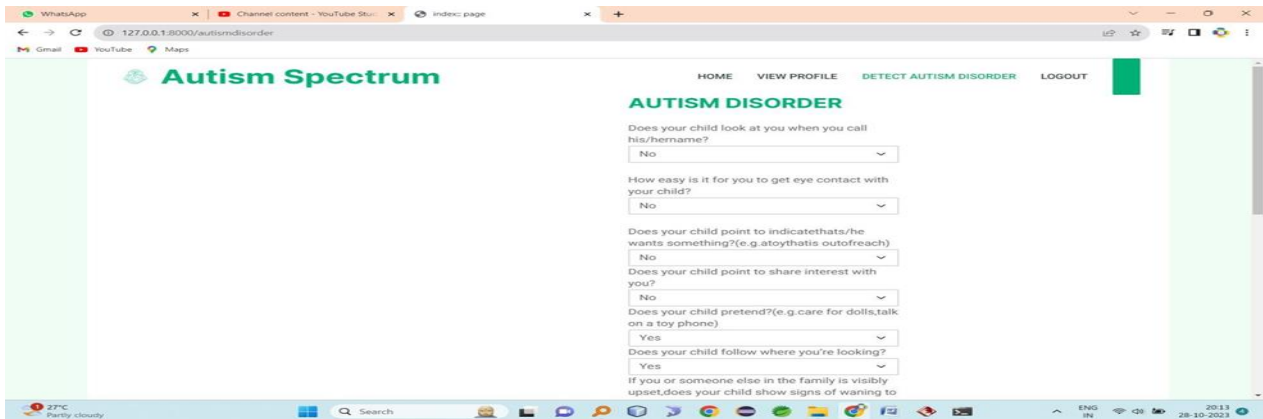
Algorithm: Artificial Neural Network Run Successfully..!!

Accuracy: 99.52606558799744

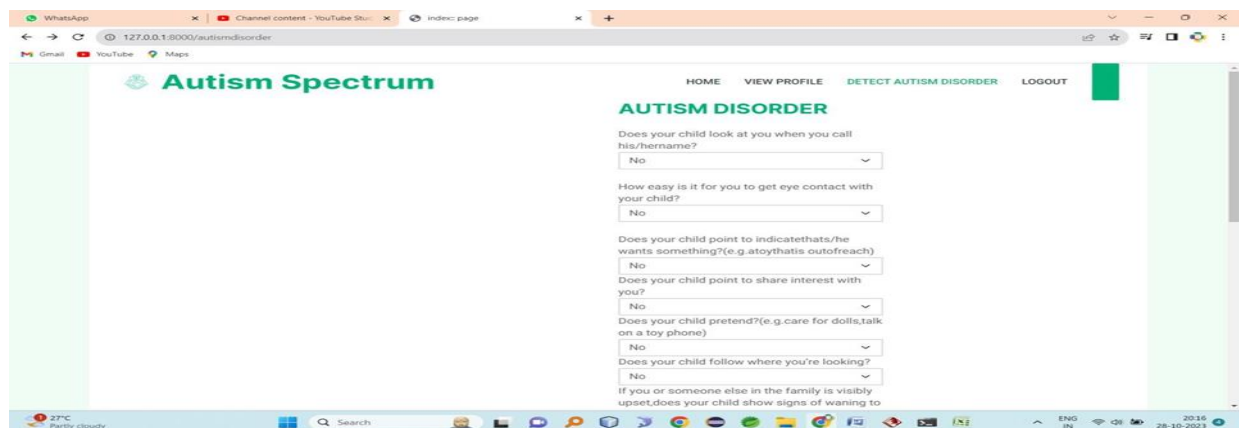
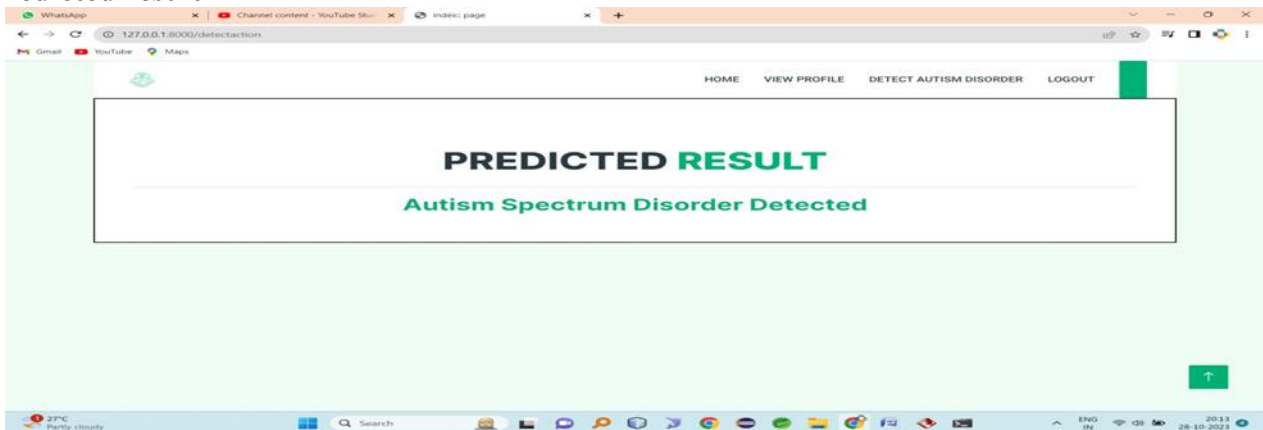
The browser's taskbar shows the system tray with a temperature of 27°C, weather 'Partly cloudy', and the date '28-10-2023'.

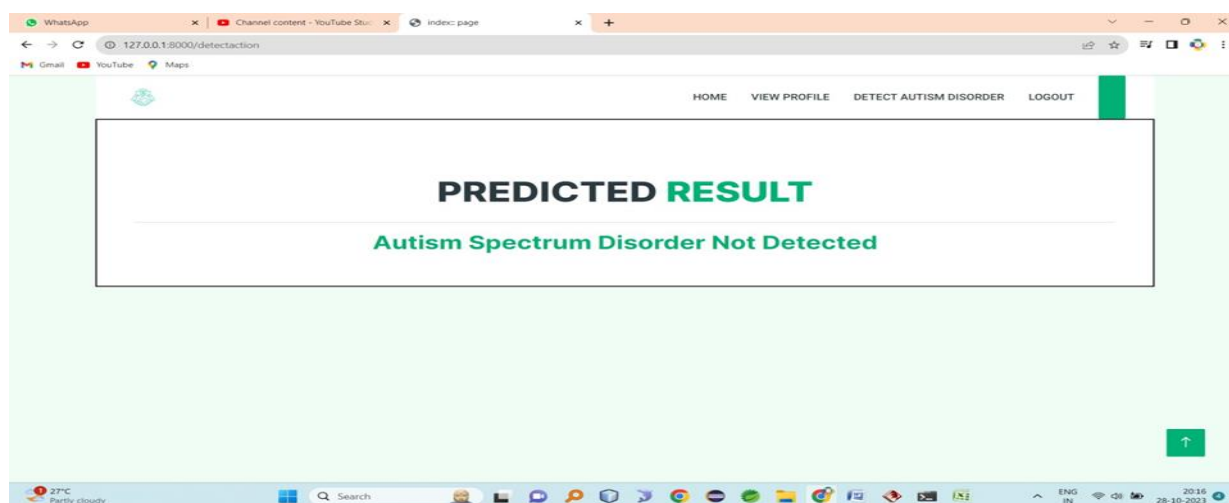
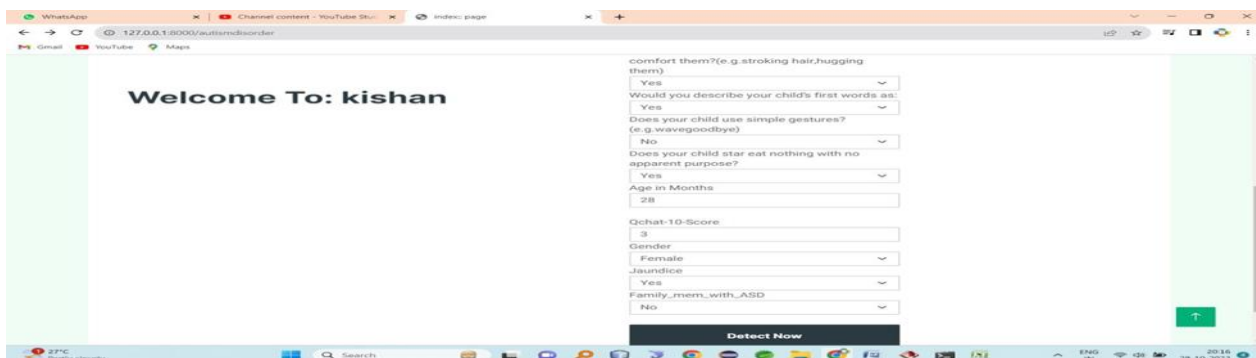






Predicted result





8.CONCLUSION

In this work, detection of Autism Spectrum Disorder was attempted using various machine learning and deep learning techniques. Various performance evaluation metrics were used to analyze the performance of the models implemented for ASD detection on non-clinical dataset from three sets of age groups viz. Child, Adolescents and the Adult. When comparing the result with another recent study [3] on this problem got a better result of the CNN classifier instead of SVM with including all its features attributes after handling missing values. In this work after handling missing value, both the SVM and CNN based models show the same accuracy of prediction of about 98.30 % for ASD Child dataset. However for the remaining two other datasets, the CNN based model was able to achieve highest accuracy result than all the other considered model building techniques, These results strongly suggest that a CNN based model can be implemented for detection of Autism Spectrum Disorder instead of the other conventional machine learning classifier suggested in earlier researches

8.2 FUTURE SCOPE

9.REFERENCES

- [1] Thabtah, Fadi. "Machine learning in autistic spectrum disorder behavioral research: A review and ways forward. (2018) " Informatics for Health and Social Care : 1-20.
- [2] Thabtah, Fadi, Firuz Kamalov, and Khairan Rajab. (2018) "A new computational intelligence approach to detect autistic features for autism screening." International journal of medical informatics 117: 112-124.
- [3] Vaishali, R., and R. Sasikala. "A machine learning based approach to classify Autism with optimum behaviour sets. (2018) " International Journal of Engineering & Technology 7(4): 18.
- [4] Constantino, John N., Patricia D. Lavesser, Y. I. Zhang, Anna M. Abbacchi, Teddi Gray, and



- Richard D. Todd. (2007) "Rapid quantitative assessment of autistic social impairment by classroom teachers." *Journal of the American Academy of Child & Adolescent Psychiatry* 46(12): 1668-1676.
- [5] Daniel Bone, Matthew S. Goodwin, Matthew P. Black, Chi-Chun Lee, Kartik Audhkhasi, and Shrikanth Narayanan. (2015) "Applying machine learning to facilitate autism diagnostics: pitfalls and promises." *Journal of autism and developmental disorders* 45(5): 1121-1136.
- [6] Dennis Paul Wall, J. Kosmicki, T. F. Deluca, E. Harstad, and Vincent Alfred Fusaro.(2012) "Use of machine learning to shorten observationbased screening and diagnosis of autism." *Translational psychiatry*, 2(4): e100.
- [7] Dennis P. Wall, Rebecca Dally, Rhiannon Luyster, Jae-Yoon Jung, and Todd F. DeLuca. (2012) "Use of artificial intelligence to shorten the behavioral diagnosis of autism." *PloS one*, 7(8): e43855.
- [8] Fadi Thabtah. (2017). "Autism spectrum disorder screening: machine learning adaptation and DSM-5 fulfillment." In *Proceedings of the 1st International Conference on Medical and Health Informatics*, pp. 1-6. ACM.
- [9] Daniel Bone, Chi-Chun Lee, Matthew P. Black, Marian E. Williams, Sungbok Lee, Pat Levitt, and Shrikanth Narayanan.(2014). "The psychologist as an interlocutor in autism spectrum disorder assessment: Insights from a study of spontaneous prosody." *Journal of Speech, Language, and 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].
- [11] Baihua Li, Arjun Sharma, James Meng, Senthil Purushwalkam, and Emma Gowen. (2017) "Applying machine learning to identify autistic adults using imitation: An exploratory study." *PloS one*, 12(8): e0182652.
- [12] Fadi Fayez Thabtah (2017), "Autistic Spectrum Disorder Screening Data for Adult",., <https://archive.ics.uci.edu/ml/machine-learningdatabases/00426/>.
- [13] M. S. Mythili, and AR Mohamed Shanavas. (2014) "A study on Autism spectrum disorders using classification techniques." *International Journal of Soft Computing and Engineering (IJSCE)*, 4: 88-91.
- [14] J. A. Kosmicki, V. Sochat, M. Duda, and D. P. Wall. (2015) "Searching for a minimal set of behaviors for autism detection through feature selection-based machine learning." *Translational psychiatry*, 5(2): e514.
- [15] Fadi Fayez Thabtah (2017), "Autistic Spectrum Disorder Screening Data for children," <https://archive.ics.uci.edu/ml/machine-learningdatabases/00419/>