Diagnosis of Autism Spectrum Disorders Using an Interactive Diagnosis Program

Tate Krejci, *Student, UCCS*

***Abstract*—Asperger Spectrum Disorders (ASD) affect a relatively large portion of the population, causing difficulty in learning appropriate behaviors for various social situations. Tests to diagnose ASD require an expert, and symptoms can often be mistaken for other mental disorders leading to under-diagnosis. Therefore, the application of a machine learning algorithm in an interactive environment such as a program will potentially increase the amount of successful diagnoses of ASD. The successful implementation of such a program will not only increase the likelihood of successfully diagnosing ASD, but also increase our understanding of ASD.**

I. Introduction

Autism Spectrum Disorders affect approximately one in two hundred and fifty people, causing difficulty in the acquisition and understanding of normal social protocols [1]. Many cases go unnoticed or misdiagnosed as there is no definitive way to diagnose an ASD over other mental disorders without excessive trial and error [2]. Furthermore, early detection of ASD in children requires expert evaluation, and cannot easily be carried out by parents or teachers [3]. Therefore, an easily and cheaply distributable application for the detection of ASD using a machine learning algorithm has the potential to detect more cases of ASD at earlier ages, and potentially provide further insights into symptoms of ASD. A program can yield potentially greater results as a program can be tailored to be interactive and captivating to its target age group. This allows for greater amounts of data to be collected than if the application was of a less interactive nature.

The latest edition of the DSM (Diagnostic and Statistical Manual of Mental Disorders), the DSM 5 has grouped the previously distinct disorders of Asperger Syndrome and Autism into the same disorder, known as an Autism Spectrum Disorder (ASD). Because of this, it is vital to determine where on the spectrum an afflicted person lies to ensure they receive the help that they specifically need. A person with severe ASD will demonstrate symptoms commonly associated with Autism, while a person with mild ASD will have symptoms similar to Asperger Syndrome. Differentiating the patients based on severity will ensure the correct type of help is provided. Thus an interactive program that can not only differentiate between a person with ASD and one without it, but can also provide insight into the severity of a patients disorder will prove to be a valuable tool.

II. Previous Work

The detection of mental disorders through the use of programs has been considered before and effectively applied to children with ADHD, exhibiting a success rate of approximately seventy-five percent with the use of a machine learning algorithm [4]. The bulk of the work done on ASD has been to identify the symptoms of ASD, the predominant one being inability to learn proper social protocol through normal social interactions [5]. However mild ASD remains harder to diagnose than severe ASD as the signs are far more subtle, especially for those with high functioning Autism [6]. Furthermore, symptoms of ASD can have multiple implications, making determining if a disorder is in fact ASD difficult [7]. More symptoms of ASD include unusual patterns of interest and behavior often leading to children with ASD seeming distant or inattentive [8]. While the symptoms of ASD are well known and progress has been made in its diagnosis, there still exists no definitive way to determine if a disorder is within the Autism spectrum of disorders or something else entirely.

III. Social Learning Theory and ASD

The primary function of ASD is to impair the ability of those affected to learn appropriate social behaviors the way unaffected individuals learn. Social Learning Theory is the theory that explains by what methods people learn what social behaviors are acceptable and what behaviors are not, though currently little is actually known about how this process actually occurs. Examples of these social behaviors include knowing to look somebody in the eye when they are speaking to you or knowing not to talk over someone else [9]. Detection techniques today involve qualitative question and answer sessions, with little in the way of quantitative data to support one diagnosis over another, often leading to misdiagnosis [2]. This is exacerbated by the fact that there is no medication to treat ASD, so doctors cannot try varying medications to determine the true disorder, as is often the case when diagnosing ADHD. With the spectrum of high IQ ASD to low IQ ASD, doctors and psychologists find it difficult to create a definitive list of symptoms [9], meaning a machine learning algorithm has the potential to discover new patterns to assist in the diagnosis of ASD.

IV. Methodology

A. The program and Machine Learning Algorithms

To identify quantitative indicators of ASD, it will be necessary to use a supervised machine learning algorithm to group data collected during the program. The type of algorithm used will depend on what kind of data the program will collect, although it is likely that a type of clustering algorithm will help group ASD users together and help identify them. This will hopefully allow the algorithm to discover new data sets that group together people suffering from ASD. Because of this, the testing phase of the program will be of utmost importance to ensure a large enough data set is collected to effectively predict whether a person has an ASD. To collect data, the program can measure variables such as answers provided, response time and mouse movements. The program will collect this data when the user is presented with social situations in which the correct response will not be readily evident to a person with an ASD. During early trials, the program can be tuned to give the maximum amount of data per encounter, and additional features can be added as needed. Programs of this style have already been implemented for the diagnosis of ADHD, with a success rate of approximately seventy-five percent [4].

B. Creating the Tests and Data Sets

Unlike some machine learning projects, total control over the data collection will be possible in this project. This means that designing the tests in the program will be as important, if not more important than fine tuning a machine learning algorithm. If the tests do not collect pertinent data that distinguishes people suffering from ASD, it is unlikely that even a well tuned algorithm could give a meaningful prediction. Therefore, extensive testing of the tests themselves will be a vital portion of creating a program for the diagnosis of ASD. To ensure the questions are well tuned, initial testing will occur only on people who do not show symptoms of ASD. With this data, it will be possible to determine what questions are effective because people without ASD should answer well written questions in the same way as other people without ASD. Once questions and scenarios have been verified through this method, they can be tested on people with ASD to determine if they can separate them out from regular people.

To create tests that capture relevant data, it has been necessary to partner with experts in psychology, specifically ASD and social learning theory. With their help, it has been possible to create scenarios in the program where the user is presented with options that indicate if they suffer from ASD or not. For example, the user could be presented with a situation where they will make differing choices based on their empathy for the characters in the program. People with ASD will likely show less empathy than those without it, as a lack of empathy is one of the characteristics of ASD [10]. Experts have expedited the process of creating tests that capture data relevant to the diagnosis of ASD. There is also the possibility to test some of the non-social aspects of ASD in a program such as the abnormal ways in which a person with ASD will focus on different tasks. Expert advice has been used to ensure that all the tests that have been implemented so far are true measures of ASD.

There are many symptoms that can indicate ASD such as deficits in executive functions [9]. This makes it possible to assess the severity of ASD in a given person by determining how impaired their executive functions are. Somebody with mild symptoms will likely show less impairment than somebody with severe symptoms [9]. To determine executive function impairment, an ordering section has been implemented where the user is shown multiple pictures of a scene and asked to select them in the order they think is best. This test can be developed to assess ASD specifically by using pictures portraying social interactions. Another test currently in the program shows the user a picture of a social interaction and asks them questions about it. Specifically, it probes the user's knowledge of what the various people in the picture think about one another, which is generally a challenge for people with ASD. The program also keeps track of the time taken to complete each individual question, meaning there is a possibility to filter out people with ASD based on the time taken to complete the various tests. Currently the program also includes the Coolidge Autism Symptoms Survey (CASS) which has already proven to be effective at diagnosing ASD [11]. This means that the program can build off of an already successful tool while adding new methods of diagnosis which are capable of measuring metrics that a pen and paper survey cannot.

C. Targets for the program

ASD manifests in different ways based on the person's age making it important to target a specific age group initially to develop both the program and the algorithm [12]. This will simplify the initial design of the program as it will only have to include tests for that specific age group, and not all possible age groups. For this reason the final program will be targeted toward 3$^{\text{rd}}$ grade age students. This is the age when most students have gained sufficient experience reading to take text based tests, and is regarded as the age when the symptoms of mild ASD first become visible [9]. This also means that the program will help those with ASD get help as soon as possible, greatly increasing their quality of life in later years.

An important note is that children in this age group are just beginning to read, so its important that any test targeted at them not accidentally test reading ability and comprehension. To do this it will be necessary to keep the amount of reading needed to a minimum, have a parent help the child, or implement a voice-over feature. For initial testing, a researcher will likely be present to answer any questions about the application, meaning that in initial phases of testing, the amount of reading is not a major concern. This is especially important when analyzing the amount of time it took to complete each question. A voice-over function that could read aloud questions and answers would serve to make the time to read prompts constant, so time taken to complete a question would be due primarily to thinking time. Another viable option is to make the tests use pictures for both prompts and answers, although doing so affects the kinds of data that can be collected. The final program will feature both voice overs and picture based tests to ensure reading ability is not an aspect of the data that is collected.

Due to the difficulty of conducting trials on children, initial tests have been conducted on a number of colleagues (8) to determine if questions are answered consistently by people who can be considered to be free of ASD. This assumption can be made because those undergoing initial testing are using an application designed for young children, and have a high probability of selecting answers that indicate they do not have ASD. This means that if a specific question is not answered consistently, it is likely a confusing or ambiguous question and should be rethought. Early testing also provides feedback on the general design of the application, all of which will lead to a far more refined application when official trials do begin and gives a partial data set to begin training the learning algorithm. Early testing has also provided valuable insight into how often 'normal' people make mistakes on the tests, which will be valuable information when training the algorithm.

D. Implementation

To make the program easily distributable and appealing to the largest audience, the program has been developed for PCs using a Windows environment. Windows has many well established frameworks for creating interactive programs such as Unity and XNA Studio, simplifying the development of a program for a Windows platform. The program will primarily perform data collection, and if successful diagnosis, but it will prove unwieldy for the program to also store and process data at larger scales. Initially the program will store data locally for simplicity, but later can send data to a data storage system. When the program is completed with the testing phase, it can be deployed to a web based player. This will allow the program to be accessed by a website, removing the need for users to install a piece of software to use it.

The program will consist of two main parts, the test portion and the data processing portion. The testing portion will consist of all the tests designed to gather data to determine a diagnosis. This portion will also include the CASS which can be treated separately internally as a cross check within the program the validity of the results for unknown cases. The second portion will be the data analysis portion which will handle all of the data processing. While the program is in the testing phase, the program will be separate from the testing portion so it can be fine tuned without having use the testing portion. Later, the data processing portion will be added at the end of the testing portion so it can give users an immediate result as well as integrate the data provided by them.

An important feature of the program will be the artwork used, as many of the tests will require specific ideas to be conveyed through pictures. At the start of the project, it is impractical to use custom artwork so images found on the Internet will be used in the early iterations of the program. While these may not convey a message perfectly, early testing should determine their efficacy. From early testing it will be evident what types of image question combinations work the best, and early testing can be based on these findings. If the results of the testing are positive, later versions of the program can include custom artwork. This will allow greater control of the messages conveyed by the pictures and will allow for more customization in the scenarios that are presented. Currently, some scenarios can not be implemented due to restrictions in the types of art available, so attaining custom art will likely increase the performance of the program once it is acquired.

E. Current Tests

Currently two different types of tests have been implemented in the application which each have multiple individual questions. The first type of test presents the user with multiple pictures that together depict a task or social interaction from beginning to end. The user clicks on each tile in the order that they think is best. This test serves to test the user's executive functions, which is lacking in many people with ASD, as well as their understanding of social situations [9]. Both the answers and the time taken to complete the question are stored by the application. Fig. 1 on the following page shows a screen shot of one of the questions in the ordering portion of the application. The second type of test implemented shows the user a picture of some social interaction and gives them a prompt and a series of answers. Some of the answers delve into what the people in the picture are actually thinking, which is how people without ASD should answer. The other answers have less to do with what the people in the pictures are actually thinking, and are likely to be chosen by people with ASD. Fig. 2 on the next page shows a screen shot of one of the questions in the intentions portion of the application. It should be noted that the artwork in the screen shots is not the artwork that will be included in the final iteration of the application. The current artwork was chosen based on availability so that prototyping and initial testing could begin rapidly.

V. Learning Algorithm

The essential function of the program is to determine whether a person has ASD. This means that there are two classes to map results to: ASD or non-ASD. Because of this, it seems a classification algorithm lends itself to the problem. Since classification generally follows a supervised structure, example data from people with and without ASD is necessary. This works out well as determining the efficacy of the various tests will mean testing the program on people with ASD before the learning algorithm is even implemented.

To first determine an effective learning algorithm, it is necessary to consider how the data collected should be processed. The program will consist of various tests with multiple parts, many of which will feature multiple choice questions. It is likely that scoring each different test and using a score from each test as the parameters for classification will be most effective. This way the number of variables is limited. However, if this proves ineffective it may also be possible to use every answer for the classification or perhaps use a regression type algorithm on the results pertaining to individual tests. The key issue will be making the number of parameters small enough for efficiency, while retaining meaning.

A Naive Bayes classifier will be used as the algorithm as it is effective at taking many parameters and calculating the probability of a class based on those. Because it functions by summing the probabilities that each individual attribute leads to a specific class, it will automatically give each answer a weight based on how effective it is at classifying ASD or non-ASD. This will hopefully add even more value to the program, as each test will affect the final outcome based on its efficacy. Another function of a Naive Bayes algorithm is

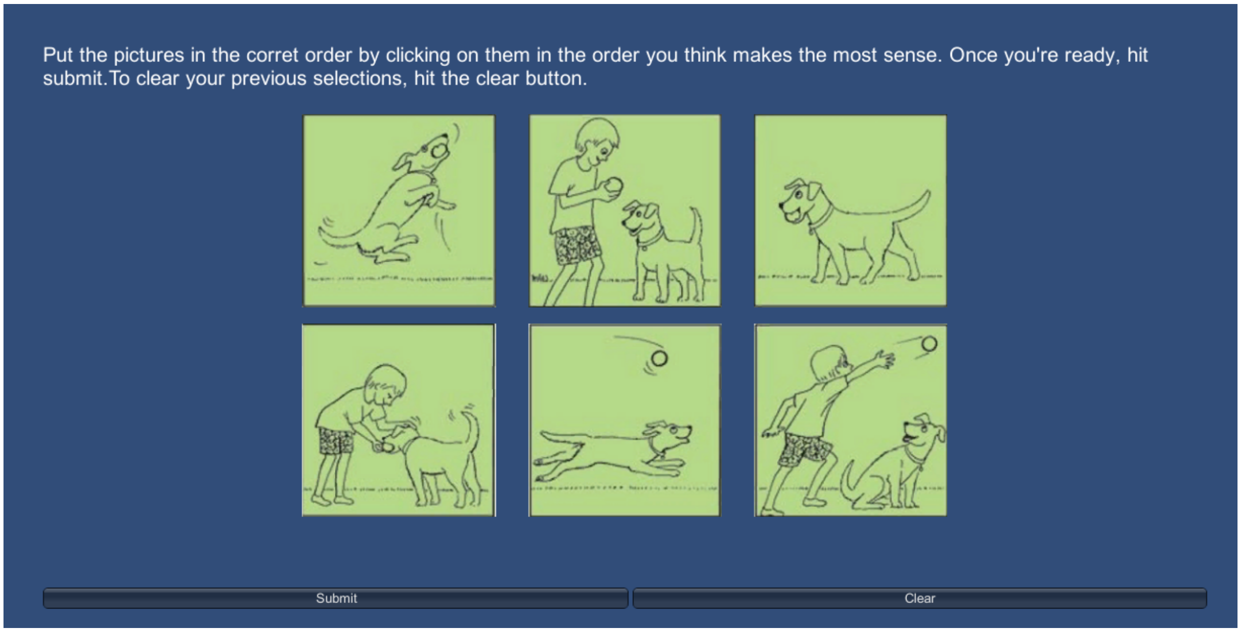

Fig. 1. Ordering Test

Fig. 2. Intentions Test

that the probability of a given parameter resulting in a specific class is not dependent on previous parameters. This may be seen as a hindrance in some cases however, the answer to one question does not have any impact on the answer to another. In this case, the disregard of previous answers serves to simplify the algorithm, making it more efficient. The fact that prior probabilities are disregarded likely will have no effect on overall results. The equation below shows the Bayes rule, off of which the Naive Bayes algorithm is based.

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)}$$

To use this algorithm, you sum the probabilities of each element giving a certain class based on weight. Thus it is possible for it to be effective with large amounts of data and will hopefully give good results when implemented. As mentioned before, it is important not only to classify users as ASD positive or negative, but also give a measure of the severity of their affliction.

The naive assumption of the Bayes algorithm removes the denominator of the equation, representing the assumption that the individual probabilities of each element of the classifier are independent of each other. As the answers to a given question

TABLE I
INITIAL RESULTS FROM NON-ASD SUBJECTS

Question Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	ASD
Participant 0	B	C	A	A	B	B	B	B	C	C	B	C	A	A	C	A	C	N
Participant 1	C	C	A	A	B	B	B	B	C	C	A	C	A	A	C	A	C	N
Participant 2	B	C	A	A	B	B	B	B	C	C	B	C	A	A	C	A	C	N
Participant 3	B	C	A	A	B	B	B	B	C	C	B	C	A	C	C	A	C	N
Participant 4	C	C	A	A	B	B	B	B	C	C	B	C	A	A	C	A	C	N
Participant 5	B	C	A	A	C	B	B	B	C	C	A	C	A	A	C	A	C	N
Participant 6	C	C	A	A	B	B	B	B	C	C	A	C	A	A	C	A	C	N
Participant 7	C	D	A	A	B	B	B	B	C	C	B	C	A	A	C	A	C	N
Participant 8	C	C	A	A	B	B	B	B	C	C	A	C	A	A	C	A	C	N

are unlikely to affect one another, the naive assumption seems a safe assumption to make for this application. To determine the probability of a given class, the probabilities that each individual property lead to a given class are multiplied together and this is then multiplied by the probability of a given class and is shown in the formula below where S is a selection of n attributes.

$$p(A|S) = \sum_{n=1}^{n} p(S_n|A) * p(B)$$

VI. INITIAL RESULTS

At this time, the main tests that have been conducted have been to assess the efficacy of various styles of questions and tests. Currently, children with a diagnosis of ASD are unavailable for testing, so tests have been conducted with colleagues. Because they do not have diagnoses of ASD, testing them provides an insight into the responses of non impaired persons. If the answers provided by them generally match for a specific question, the question is at least effective at grouping users without ASD together. This testing will help filter out questions that are ambiguous and give inconsistent data. When a test has been verified as consistent, more tests can be created based on those. The next step will be to give the refined tests to people with ASD to determine that they are also effective at distinguishing them from non-impaired people. TABLE I shows some of the preliminary data collected from users who do not have ASD using the intentions test. In this test, the user is presented with a picture of people performing various actions. The user is provided a prompt and a selection of answers. The answers each present a different level of social awareness, so people with ASD will likely pick the answers that show less social awareness.

Here the rows represent the answers submitted by each participant and the columns represent each question in one of the tests. For the most part, the answers are the same as expected. Differing answers to the same question indicate an ambiguous question that should be rethought so it is consistently answered the same for all people without ASD. The questions whose answers are all the same represent good questions with a style that should be repeated when adding new questions to this test. It should be noted that isolating good styles is accomplished by using the same prompt and image for a question and just varying the answer choices. The focus for this particular test was on creating distinguishing answers, which is why they were the factors that changed to determine the efficacy.

To ensure that a naive Bayes algorithm is an appropriate choice for the algorithm, test data was generated to represent the answers of users who were suffering from ASD. This data was generated pseudo-randomly with the goal of testing the classifier in mind. These are not results from real people with ASD. When this data was used with the data obtained from colleagues, the algorithm classified eighty-seven percent of the cases correctly. This means that as long as ASD users answer questions in the predicted manner, a naive Bayes classifier will successfully distinguish between ASD and non-ASD users. With the implementation of further tests, and the collection of more data, hopefully this number can be further increased in the future.

VII. TIME LINE

At this point the core functionality of the program has been implemented. Multiple tests are completed and a Naive Bayes classifier has been added to provide in application results for users. Future work will involve collecting real data to train the classifier using children of an appropriate age. Another important task will be to improve the existing tests, and add new tests. More tests will make the application a more comprehensive test of ASD, likely increasing the accuracy of the diagnosis. Existing tests can also be expanded as limited art assets have made the test sections relatively short. With the acquisition of custom art, more tests can be devised and existing tests will have greatly increased accuracy. This is because with custom art, variables that can affect a person's answer can be easily eliminated. Once the application has reached a polished state, and preliminary data has been gathered, the program can be exported to a web player. This will means that more people will have access to the program which will provide more diverse training data and hopefully further improve the algorithm's accuracy.

VIII. GOALS

The completion of the prototype of the program will allow for small scale data collection and testing of the program. This will involve identifying a test group, and determining what member of that group suffer from ASD. From there, a naive Bayes algorithm can be applied to the data and used to identify patterns indicative of ASD. If the algorithm can successfully predict ASD in the targeted age group, it can be deployed on a larger scale to collect more data to improve the algorithm and it can be modified to support various target age groups. With enough success, the program could be further modified for use as not only a diagnostic tool, but as a treatment for patients with ASD.

Deployment to a larger scale will involve extensive testing and polishing of the existing program to come up with a complete suite of tests measuring many different aspects of ASD in distinct ways. Once the program reaches this point, and with university approval, the program can be exported to a web browser so it can be taken online and collect data online. It can then be modified to also give a suggestion of a diagnosis of ASD so it can be used by real people online. If the online program proves successful, it may be possible to create more diagnostic tools for the various mental illnesses that exist.

IX. Conclusion

The successful implementation of a machine learning algorithm to diagnose ASD will provide parents and doctors with a more effective means of diagnosing and helping those suffering from ASD. It also has the potential to vastly increase our understanding of the symptoms of ASD and perhaps provide clues to its causes and increase our understanding of social learning. The use of a program as the primary platform for the test will increase patient interaction with the application, leading to a greater quality and quantity of the data collected.

References

[1] B. J. Tonge, A. V. Brereton, K. M. Gray, and S. L. Einfeld, "Behavioural and emotional disturbance in high-functioning autism and asperger syndrome," *Autism*, vol. 3, no. 2, pp. 117–130, 1999.

[2] B. G. Haskins and J. A. Silva, "Asperger's disorder and criminal behavior: forensic-psychiatric considerations," *Journal of the American Academy of Psychiatry and the Law Online*, vol. 34, no. 3, pp. 374–384, 2006.

[3] O. Teitelbaum, T. Benton, P. K. Shah, A. Prince, J. L. Kelly, and P. Teitelbaum, "Eshkol–wachman movement notation in diagnosis: The early detection of asperger's syndrome," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 32, pp. 11 909–11 914, 2004.

[4] S. Srivastava, M. Heller, J. Srivastava, M. Kurt Roots, and J. Schumann, "Tangible games for diagnosing adhd–clinical trial results."

[5] L. Wing, "Asperger's syndrome: a clinical account." *Psychological medicine*, 1981.

[6] P. Howlin and A. Asgharian, "The diagnosis of autism and asperger syndrome: findings from a survey of 770 families," *Developmental Medicine & Child Neurology*, vol. 41, no. 12, pp. 834–839, 1999.

[7] D. V. Bishop, "Autism, asperger's syndrome and semantic-pragmatic disorder: Where are the boundaries?" *International Journal of Language & Communication Disorders*, vol. 24, no. 2, pp. 107–121, 1989.

[8] A. Klin and F. R. Volkmar, "Asperger's syndrome," *Handbook of autism and pervasive developmental disorders*, vol. 2, pp. 88–125, 1997.

[9] M. G. Winner, *Thinking About You Thinking About Me*. Think Social Pub, 2007.

[10] I. Dziobek, K. Rogers, S. Fleck, M. Bahnemann, H. R. Heekeren, O. T. Wolf, and A. Convit, "Dissociation of cognitive and emotional empathy in adults with asperger syndrome using the multifaceted empathy test (met)," *Journal of autism and developmental disorders*, vol. 38, no. 3, pp. 464–473, 2008.

[11] C. S. R. D. L. S. Frederick L. Coolidge, Peter D. Marle and P. Monaghan, "Psychometric properties of a new measure to assess autism spectrum disorder in dsm-5," *American Journal of Orthopsychiatry*, vol. 83, no. 1, p. 126–130, 2013.

[12] C. Koning and J. Magill-Evans, "Social and language skills in adolescent boys with asperger syndrome," *Autism*, vol. 5, no. 1, pp. 23–36, 2001.