

Evaluation of OCD-Spectrum Disorders Using Data Analytics and Visualization Techniques

R. Yamini, Drishti De and Chavi Garg

Abstract--- *Mental or psychological disorders are a serious problem to be resolved as they affect emotional stability and security for both the person and the environment. Obsessive Compulsive Disorder (OCD) is a long-lasting psychological disorder in which a person has uncontrollable, reoccurring thoughts (obsessions), and behaviours (compulsions) that he or she feels the urge to repeat over and over. This disorder is also known to branch into its various subtypes called the OCD-spectrum disorders, which include hoarding, body dysmorphic disorder, trichotillomania, etc. Hence, this paper is aimed towards discovering common symptoms and trends as well as the cognitive pathways influenced by various factors that lead to these disorders. This survey investigates recent research efforts and conducts a comprehensive overview of the work on medical big data, especially as related to OCD, its subtypes and its repercussions. It focuses on the full cycles of data processing which includes OCD data pre-processing, data analytics tools and algorithms along with visualization techniques. It attempts to integrate statistical techniques with clinically certified medical diagnosis in order to characterize the several elements that contribute to the mental transformation that a person undergoes. The result of this survey is expected to illustrate the standard ways that could help us better understand what treatment might work for which kind of patient.*

Keywords--- *Mental Disorders, Obsessive Compulsive Disorder (OCD), Obsessions, Compulsions, CYBOCS, Adolescent Mental Health, OCD-Spectrum Disorders*

I. INTRODUCTION

A mental disorder (psychiatric disorder) is a behavioral pattern that inflicts harm on the patient themselves or their surroundings (such as, people they are related to or the objects that are a part and parcel of their everyday life). It is measured by an impairment of higher grade. A mental health professional is usually hired to diagnose a person suspected to suffer from it. Mental disorders are a combination that describe how a patient behaves in a situation, perceives the surroundings, feels or thinks. This may be associated with specific regions or functions of the brain, often in a social context. There are various kinds of mental disorders specific to a person's stimuli and course of action, which include: Anxiety disorders, mood disorders, psychotic disorders, impulse control disorder, addiction disorder, eating disorder and personality disorder to mention a few.

Obsessive-compulsive disorder (OCD) is a disorder in which the patient is affected mentally and feels the need

R. Yamini, Assistant Professor, Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, Tamil Nadu.
E-mail: yaminir@srmist.edu.in

Drishti De, Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, Tamil Nadu.
E-mail: dd9672@srmist.edu.in

Chavi Garg, Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, Tamil Nadu.
E-mail: cg5440@srmist.edu.in

to perform certain activities constantly or compulsions while in some cases has certain thoughts or obsessions such as - washing hands, counting of things, and checking to see if a door is locked repeatedly. It also includes unwanted negative or upsetting thoughts and also special repetitive thoughts designed to counteract some of the other negative thoughts. Some may have difficulty in throwing things out and performing actions based on religiosity with excessive checking and ritualized ways of walking. In addition to these, according to a study by Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, emphasis can also be put on the occurrence of certain situations, places or objects that trigger distressing thoughts and sudden urges to do specific things ritualistically. These come under two types of important associations in OCD with the first type being associations with objects, thoughts or situations that may cause rise in stress or anxiety in people and the second type of association is between carrying out rituals and decreasing stress levels which are mostly short-lived. Thus mostly, people found avoiding such situations, places and objects. Treating such cases can work out well only when such associations are broken, and the therapeutic treatment as such is called “exposure and ritual prevention”.

The onset of OCD is mostly evenly progressive but can also start suddenly in some cases. The symptoms of this disorder may deviate in severity as time elapses and this may relate to the manifestation of certain stressful events in the patient’s lives. As symptoms are usually known to worsen with age, it becomes increasingly difficult for a person to remember the exact initiation of the disorder, but can sometimes recall the first observed incidence of the said symptoms related to the disorder that was disrupting their lives. While the cause of some people acquiring the disorder while others do not is unknown, yet there are specific “thinking mistakes” about harm that occur in OCD. Such thinking mistakes include:

- Thinking about certain actions is the same as doing those or having an urge to do those.
- People should control these type of recurring thoughts.
- If people do not try to prevent any sort of harm, then it is the same as harming others.
- A person is responsible for the harm incurred despite the circumstances they were undergoing.

According to various researches, OCD is caused due to an imbalance of neurotransmitters in the brain and serotonin (a “feel good” chemical) plays an essential role in it. When a person suffers from OCD, the serotonin levels have shown to be undergoing a drastic fall, hence, leading to the formation of repetitive thoughts, which are mostly negative. This imbalance in the level of neurotransmitters is proven to be mostly genetic or hereditary as chances of this disorder passing down in a family is usually higher. Other factors may include a history of consistent child abuse or stress inducing events. People, especially children with OCD have shown unusual dopamine and serotonin activity in various regions of their brains.

II. OCD IN CHILDREN AND ADOLESCENTS

According to a recent survey by Dr Owen Kelly, PhD, “Obsessive Compulsive Disorder (OCD) is generally referred to as a childhood-onset if the obsessions and compulsions happen to occur prior puberty”. The average age of the onset is estimated to be around 10 years, even though children of 5 or 6 years of age may also develop the illness. An interesting discovery is the effect of OCD being greater in boys as a childhood-onset while

its effects females more after the onset of puberty. The content of children's obsessions differs from that of adults, with specific obsessions related to the death of certain family members like parents. Thus, the rituals or compulsions of children may also likely tend to involve or be centered around themselves or mostly their family members than in case of adults. Unlike children, adolescents may experience greater incidence of sexual-oriented obsessions due to undergoing puberty [12]. According to the American Academy of Child and Adolescent Psychiatry (AACAP), OCD commonly initiates in children and young adults as seen in approximately 0.005% of young patients. The obsessive thoughts may vary with age of the child concerned and tends to change over time. An older child or a teenager with OCD may have fears related to contracting ailments through contaminated objects and foods thus are more prone to give way to certain rituals or compulsions. These children OCD cases may or may not have a previous family history, but children who frequently observe their parents having obsessions and compulsions are more inclined to acquiring the disorder. Due to societal issues or the fear of being mistreated by peers, these children and adolescents are often victims of shame and embarrassment relating to their OCD and hence become hesitant to talk about their thoughts and behaviors [14]. Therefore, good communication with parents, guardians or trusted adults can increase the efficiency of understanding the matter and thus the children and adolescents can be provided with proper care and support.

III. EXISTING SYSTEM

Previous researches in OCD have seen the use of global field and global area synchronization methods along with support vector machine method and cross validation method for the classification of synchronization values of OCD in different patients [17]. Besides work based on classification features of mental disorders were developed using Fuzzy Logic with a triangular fuzzy membership function [3]. These researches with their fairly advanced methodologies have only targeted the obsessions and compulsions rather than also specifying the equal weightage of the OCD-spectrum disorders.

In addition to the above, survey of all these research papers have also indicated the use of either fewer databases or the lack in analysis of available databases based on multi-factor approach and practical scenarios. Thus, the need for detailed OCD and OCD-Spectrum disorders data reports should be resolved and emphasis needs to be put on using advanced testing techniques for filtering out optimized solutions by enhancing testing data to make it consistent with the training data.

Models based on classification of the various categories of OCD using supervised machine learning techniques like SVM, Decision Tree Classifier, K Nearest Neighbors, etc., will pave the way for better diagnosis and future outcomes.

IV. PROPOSED SYSTEM

To diagnose OCD and OCD-Spectrum diseases using standard clinical assessment questionnaires (for example, YBOCS and extended versions of YBOCS scale) through a user-interface and check for the severity of the mental disorders of each of the patients and to apply statistical and predictive analysis on the sample data collected through the diagnosis of OCD-Spectrum diseases using standard clinician-certified assessment questionnaires to predict the

OCD severity for newly registered patients.

Using the certified extended clinical assessment scale, a system is to be developed which will input values marked by the patients related to the symptoms experienced by them in order to calculate the extent of each individual disorder under OCD-spectrum. Thus, this proposed system shall be able to differentiate and put forward the OCD subtype with maximum severity and render individual yet effective treatment to each of the patients.

In Fig. 1., the illustration of basic architecture diagram represents the concerned system for proposed work consisting of namely two main components. They are: User Interface, which can be further divided into patient database and the components of the patient database that are needed to diagnose the disorder, suggest effective treatment and predict future OCD trends based on different parameters. The second main component is the prediction system. The prediction system uses both classification and regression features to classify different levels of severity accurately and then come up with individual patient based effective treatment by assessing all the statistical values individually to understand the intensity of the treatment required. This system hence uses a model consisting of both statistical analysis and predictive analysis of the values received from the various clinical questionnaires being used. The statistical values are stored in different patient disorder tables with the total score for each spectrum disorder to make the analysis easier.

Algorithm

Different algorithms have been used to analyze the already obtained dataset based on CYBOCS self-report questionnaire. The algorithms used are Naïve Bayes Classifier, Random Forest, Decision Tree and K-Nearest Neighbors. Table 1 shows the parameters used to find the levels of OCD.

Age	Gender	CYBOCS-Obsessions	CYBOCS- Compulsions	OCD Severity
<8	M/F	0-4	0-4	0-4
8-12	M/F	0-4	0-4	0-4
13-17	M/F	0-4	0-4	0-4

Table 1: Attributes Determining OCD Severity

A comparative analysis of previously utilized algorithms like Naïve Bayes (accuracy: 100%), Random Forest (accuracy: 63%) and Decision Tree (accuracy: 72%) classifiers have proven to give an overall superficial overview of the data being used for assessing OCD in children and early adolescents. Naïve Bayes Classifier algorithm/technique is based on Bayes' Theorem and is one of the most popular and powerful machine learning tool used for predictive analysis, while on the contrary, Decision Tree and Random Forest (advanced level of Decision Tree Algorithm as more trees with decision rules increases the robustness of the model) Algorithms can be used for creating both classification and regression models by learning various decision rules, thus proving to be stronger than the former algorithm.

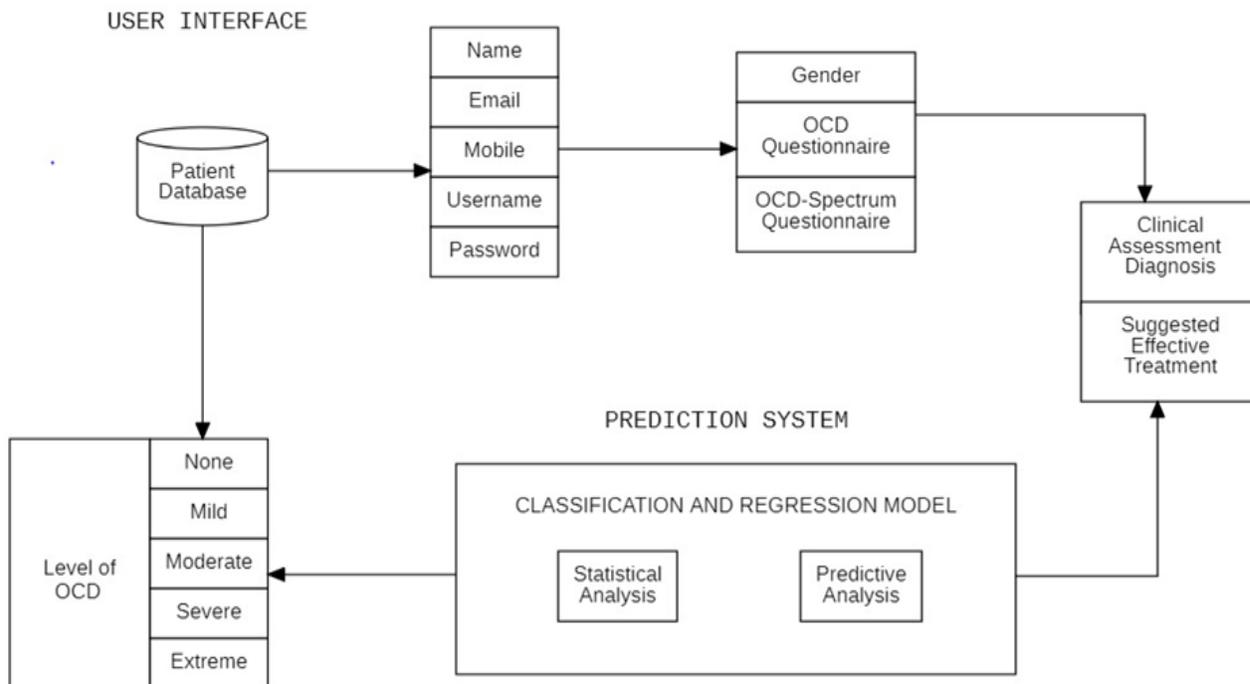


Fig. 1: Architecture Diagram for the Proposed Model

Another supervised machine learning technique that has not yet been utilized for this type of analysis is K-Nearest Neighbors (KNN) classifying algorithm (accuracy: 93%). K-Nearest Neighbors (KNN) is a non-parametric algorithm with the purpose to use a dataset in which the data points are classified into various groups in order to predict the classification of a new data point. In this algorithm, an object or data point is classified by a majority vote of all its neighbors and the object is assigned to the class that is found to be the most common among its 'k' nearest neighbors.

Algorithm	Accuracy(%)
Naïve Bayes	100%
Random Forest	63%
Decision Tree	72%
K-Nearest Neighbors	93%

Table 2: Comparative Analysis of Various Algorithms

Several visualization techniques have been used to further comprehend the data and have a clear vision of future prediction inclinations. Some of the visualization techniques used are:

- Heat Map (used for representing common user behaviour),
- Scatter Plot (represents relation between two variables),
- Histogram (accurate measure of the distribution of numerical data),
- Box Plot (quartiles divide the numerical data into groups which are illustrated graphically) and
- Faceting (visualizes distribution of one variable with multiple variables and the relationship among them)

V. EXPERIMENTS

Children's Yale Brown Obsessive Compulsive Scales (CYBOCS) is the children's version of the YBOCS scale for assessing the level of obsessions and compulsions in adults, specially modified according to the requirements of the children and their intellectual ability.

It is a clinical-report questionnaire that was designed to analyze the symptoms of obsessive-compulsive disorder from childhood to early adolescence stage. Goodman et al. developed it to convert the subjective answers of patients [9] into numerical data in order to measure the OCD severity. The CYBOCS-SR is a Children's Yale Brown Obsessive Compulsive Scale self-report questionnaire that aims to evaluate the severity of Obsessive-compulsive Disorder (OCD) symptoms in Children and Adolescents. This scale includes 10 five-pointer questions that need to be assessed in children, which are then summed to a score ranging from 0-40 with a score of 20 or more pointing towards significant clinical severity. The scores obtained from this self-report questionnaire are satisfactorily valid and reliable. Scahill et al. checked their reliability [10] and proved it time and again in multiple researches along with its accuracy which has been much above the satisfactory level in most of them.

One of the datasets used for this research was obtained by conducting a survey on 87 adolescent participants for OCD using CYBOCS questionnaire. The dataset contains data for 41 (47.1%) males and 46 (52.9%) females, whose age ranged from 13 to 17. The breakdown of the participants' racial categorization or ethnicities was: - Asian (n=10), Black (n=1), Biracial (n=2), Caucasian (n=72) and Hispanic (n=2). The dataset also includes scores ranging from 0 to 4 for the 10 questions in the self-report questionnaire, with the first 5 scores implying 'obsessions' [2] and the rest implying 'compulsions'. These scores are then summed up to give the total obsession score and total compulsion score respectively. In order to derive the OCD severity in each patient, these total scores are fused to give the ultimate OCD score (out of 40). The severity data is mined from the previously calculated ultimate OCD score.

The other dataset consists of all the statistical values based on the usage of various clinician rated questionnaires for the OCD-spectrum disorders, which are: Hoarding, Kleptomania, Trichotillomania, Excoriation (Skin-picking disorder), Body Dysmorphic Disorder, Hypochondriasis, etc. These questionnaires consist of around 5-12 multiple choice questions specific to the different spectrum disorders and the answers are mostly rated from 0-4 where they denote values like none, mild, moderate, severe and extreme. The questions in these questionnaires mainly target the time occupied by the them (both the thoughts and the compulsive activity) and the amount of resistance that a patient needs to put up with in order to prevent from yielding to the disorders. Most of these questionnaires are the extended versions of the YBOCS questionnaire. These statistical values derived from each of the OCD-spectrum disorder questionnaires are then stored into a patient database which consists of different tables for the different OCD subtypes. The scores obtained from each of the answers to every question is then summed up to give the total score for the subtype disorders. This aids in the process of calculating the severity of the disorders and thus acts as the main element in order to classify the various severity levels. These total scores are then connected together in a different table in the database corresponding to the concerned patient username it is related to. In the case of suggesting fairly accurate and efficient treatment based on severity, the patient database table values need to be assessed individually to check which disorder has the most severity and put it as a primary focus of the treatment

suggestions provided. Thus, understanding the need of our project and the requirement of accurate dataset representation along with just-in-time prediction for new input values by comparing it with the previous values in the patient tables, we opt the use of KNN algorithm due to its versatility with both classification and regression techniques and the features that are a must in our system. The other main focus of our project, besides the diagnosis and treatment of the spectrum disorders, is to analyze the current trends in OCD based on different influencing parameters using suitable visualizations. These visualizations will help gain insight on the complexity of the OCD-spectrum disorders and the list of parameters that play a major role in the sure diagnosis of OCD and its subtypes.

VI. RESULTS

The user web-interface developed using all the clinical questionnaires finally displays an output of the severity of the disorders based on the total disorder scores. This helps the users to understand the level of OCD they might be suffering from and the extent with which OCD might tend to its subtypes. A figure showcasing the final display pages for different users is shown as:

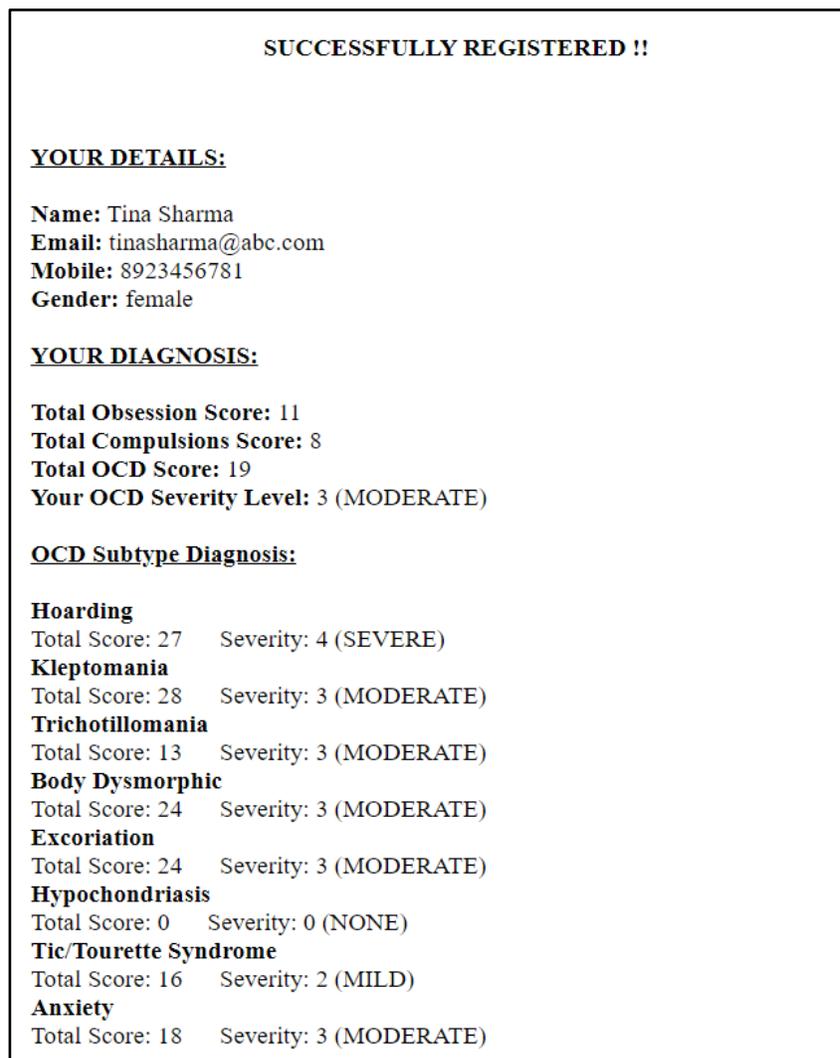


Fig. 2A: Diagnostic Result for User-1

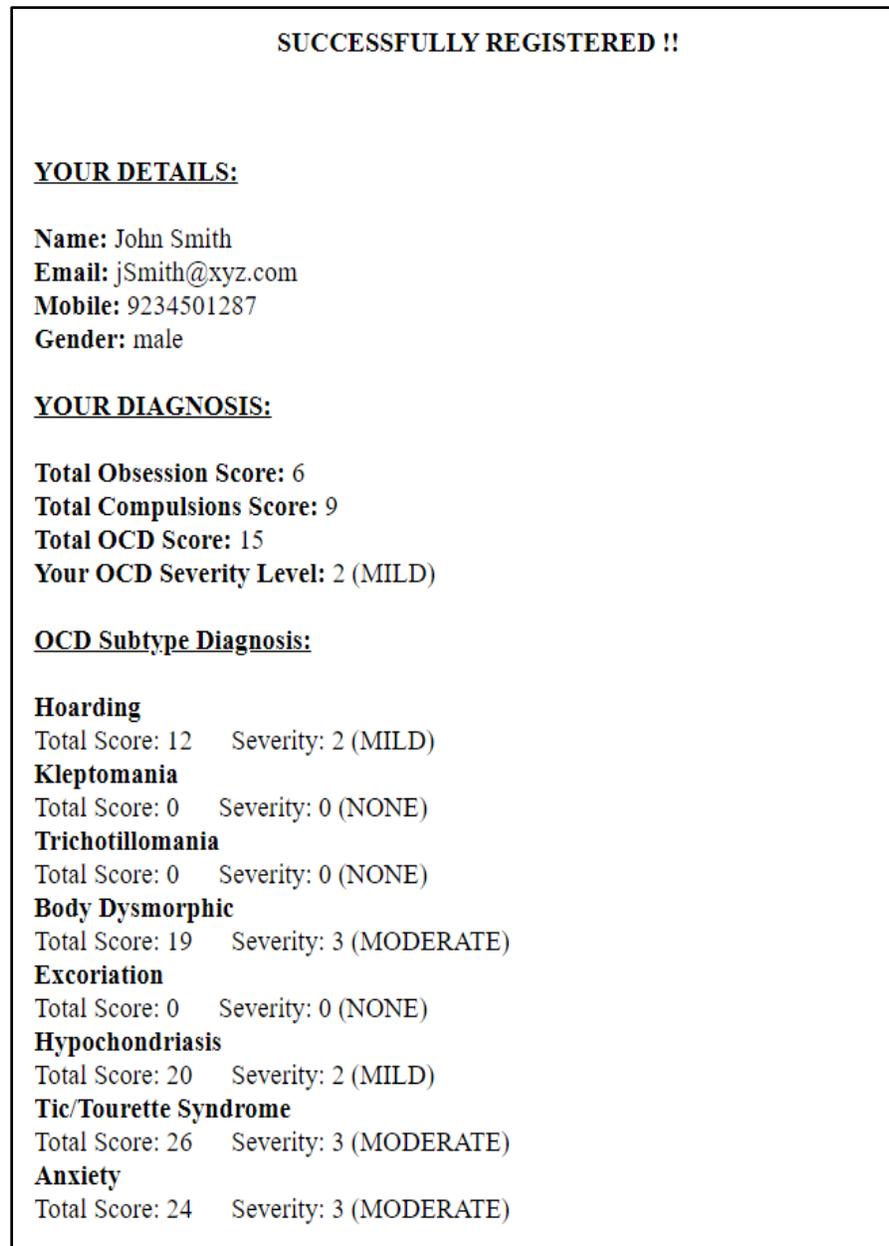


Fig. 2B: Diagnostic Result for User-2

On the other hand, results of the visualization techniques applied on the dataset used for our research on OCD in children and young adolescents can be interpreted as – OCD level or severity in children and adolescents is mostly dependent on the parameter of CYBOCS_total present in the dataset (based on analysis of Heatmap visualization) and is independent of the parameter representing the age of the participants/patients. Hence, a patient can suffer any level of OCD irrespective of their age (by analyzing Scatter plot visualization tool). This is consistent with the study conducted by Thomsen on children of Denmark [14]. Comprehending the graphical result obtained using histogram plot (Fig. 3), the OCD severity/level of 4 has been found to be the most frequent ($f \geq 30$) among the 87 participants used for obtaining our data. Thus, an OCD score ranging from 26-30 (OCD level: 5) is being observed as the score

with the highest frequency ($f = 15$), while on the other hand, a score ranging from 35-38 (OCD level: 6) is found to be the least common among the participants. Besides that, the most affected among the age groups of 87 different participants are the 15-year olds whereas the least affected are those with 13 years of age.

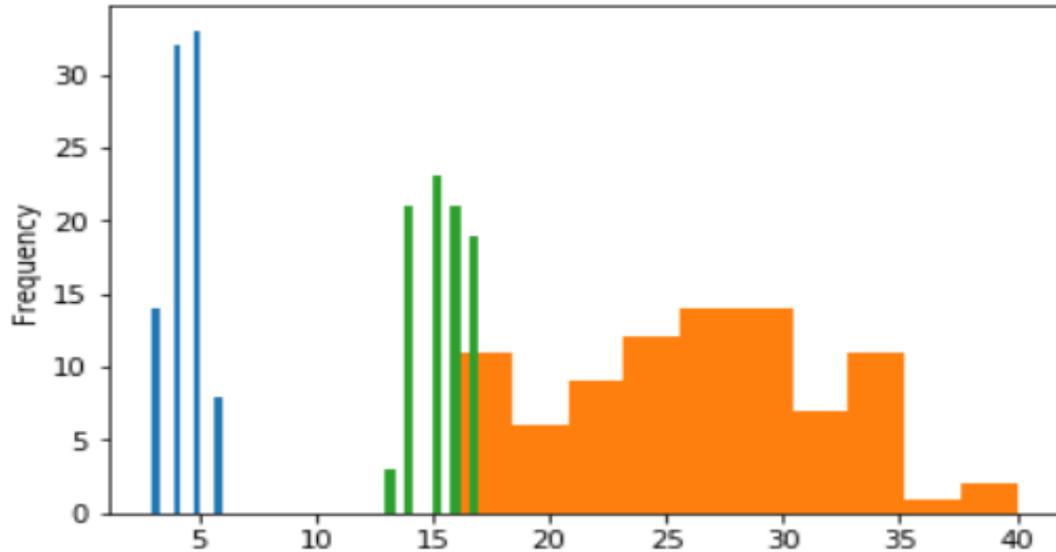


Fig. 3: Histogram Depicting Frequency of OCD According to OCD Level, Age and CYBOCS Score Respectively

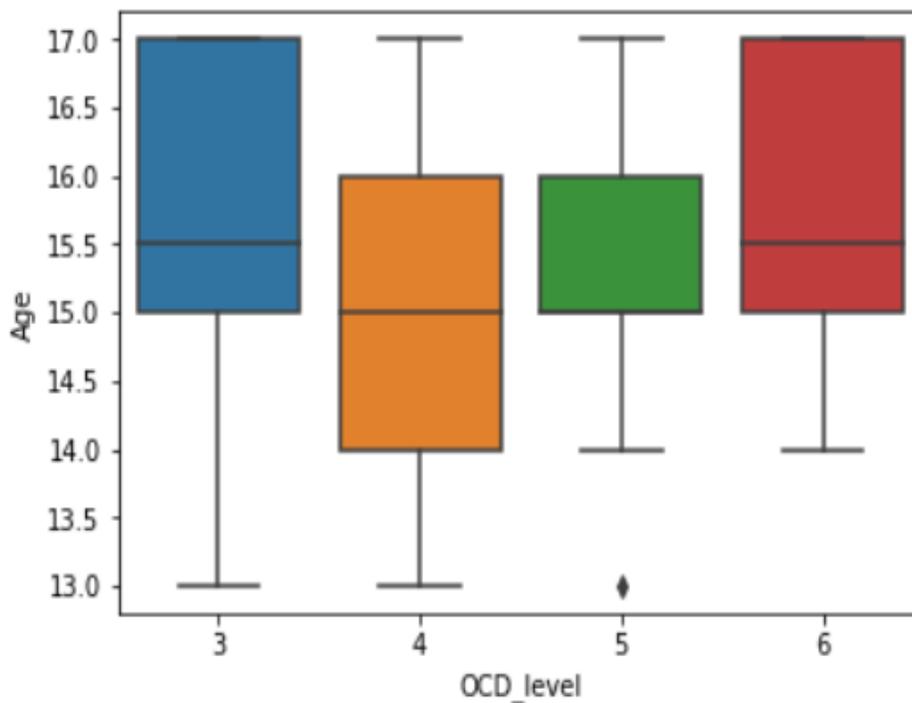


Fig. 4: Variation of OCD Level Distribution According to Age

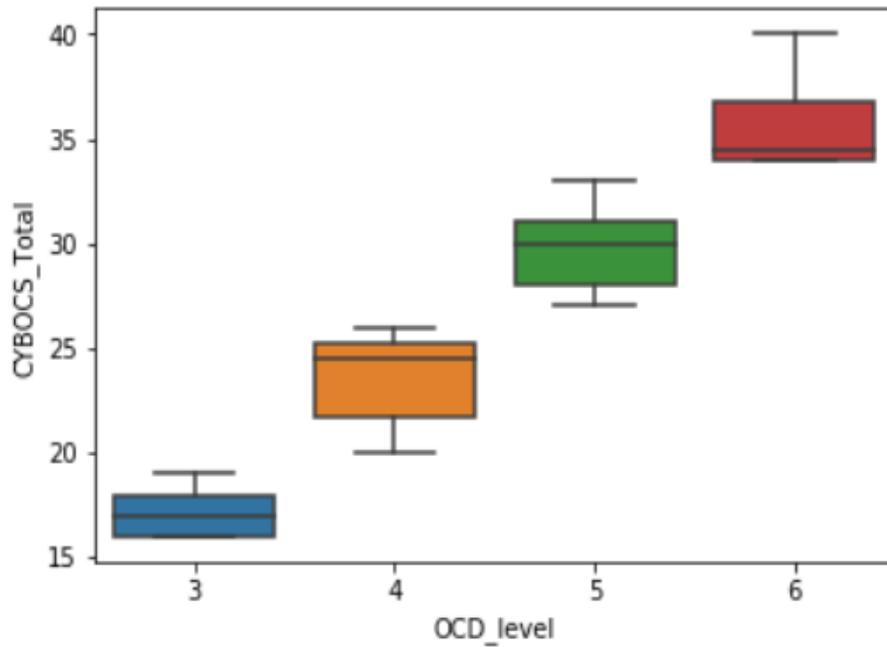


Fig. 5: Variation of OCD Level Distribution According to CYBOCS Score

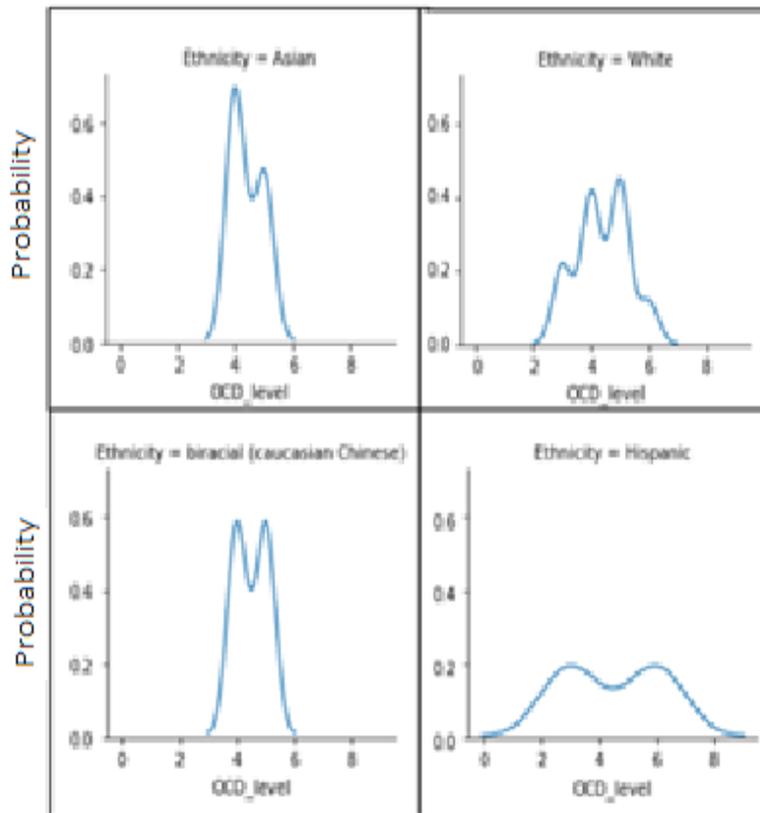


Fig. 6: OCD Level Distribution Among Ethnicities

Box plot (Fig. 4, 5) analysis interprets that the patients with minimum age of 13 were observed to have an OCD level of 3 and 4, average age of 15 with a level of 4 and 5, whereas levels 3 and 6 ranged from ages 15 to 16. Faceting visualization technique (Fig. 6) has been used to provide more insight on the probabilistic role of ethnicity in acquiring OCD in children, where, 60% Asians were observed suffering from an OCD level of 4 while the rest 40-50% with a level of 5. Coming to the effects of OCD on the people having white ethnicity, 40% of them have been diagnosed with a level 4 in OCD, at least 50% were seen tending towards the OCD level of 5, while less than 20% were found to have an OCD level of 6. 60% of the people that hailed from a Chinese ethnicity were shown having OCD level of both 4 and 5, while 20% of the Hispanic patients were observed with an OCD level of 3 and 6 and less than 20% with OCD level 5. According to the dataset obtained by the self-reports of 87 children participants, the patients with ethnicity as black showed negligible amount of OCD.

VII. CONCLUSION

With the current implementation work mentioned above, a system shall be developed with the input of sample data based on all the of the tables under patient database. This system then shall be finally tested with the help of real time patient database to check its integrity, validity, reliability and efficiency. Future enhancements in the above-mentioned system may include the use of vernacular assessments based on the culture the patient hails from, as most of the patients in remote and rural areas of our country have very limited knowledge of English language.

REFERENCES

- [1] Matthew Lewis and Lola Canamero, "Robot Models of Mental Disorders"; *Seventh International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*; pp. 193 – 200; 2017.
- [2] Payton J. Jones, Patrick Mair, Bradley C. Reimann, Beth L. Mugno and Richard J. McCally, "A Network Perspective on Comorbid Depression in Adolescents with Obsessive-compulsive Disorder"; *Journal of Anxiety Disorders*; pp. 2 – 28; 2017.
- [3] Meza Silvana, Ricky Akbar, Derisma, Mia Audina and Firdaus, "Development of Classification Features of Mental Disorder Using the Fuzzy Logic Mamdani Method"; *International Conference on Information Technology Systems and Innovation (ICITSI)*; pp.410 - 414; 2018.
- [4] Amir Hossein Yazdavar, Mohammad Saied Mahdavejad, Goonmeet Bajaj, Krishnaprasad Thirunarayan, Jyotishman Pathak, Amit Sheth, "Mental Health Analysis via Social Media Data"; *IEEE International Conference on Healthcare Informatics*; pp. 459 – 460; 2018.
- [5] Luciano Comin Nunes, Plácido Rogério Pinheiro, Mirian Calíope Dantas Pinheiro, Marum Simão Filho, Rafael Espíndola Comin Nunes, Pedro Gabriel Calíope Dantas Pinheiro, "Automatic Detection and Diagnosis of Neurologic Diseases"; *IEEE Access*; pp. 29924 – 29941; 2019.
- [6] John Piacentini, R.lindsey Bergman, Melody Keller and James McCracken, "Functional Impairment in Children and Adolescents with Obsessive-Compulsive Disorder"; *Journal of Child and Adolescent Psychopharmacology*, Vol. 13, Supplement 1; pp. S61 – S69; 2003.
- [7] Isobel Heyman, Eric Fombonne, Helen Simmons, Tamsin Ford, Howard Meltzer and Robert Goodman, "Prevalence of Obsessive-Compulsive Disorder in the British nationwide survey of child mental health"; *British Journal of Psychiatry*; pp. 324 – 329; 2001.
- [8] Y. C. Janardhan Reddy, S. Srinath, H. M. Prakash, S. C. Girimaji, S. P. Sheshadri, S. Khanna and D. K. Subbakrishna, "A follow-up study of juvenile obsessive-compulsive disorder from India"; *Acta Psychiatr Scand*; pp. 457 – 464; 2002.
- [9] Wayne K. Goodman, Lawrence Scahill, Lawrence H. Price, Steven A. Rasmussen, Mark A. Riddle and Judith L. Rapoport, "Children's Yale – Brown Obsessive Compulsive Scale (CY-BOCS)"; *Arch Gen Psychiatry* 46; pp. 1006 – 1016; 1989.

- [10] Lawrence Scahill, Mark A. Riddle, Maureen McSwiggin-Hardin, Sharon I. Ort, Robert A. King, Wayne K. Goodman, Domenic Cicchetti and James F. Leckman, “Children’s Yale-Brown Obsessive Compulsive Scale: Reliability and Validity”; *J Am Acad Child Adolesc Psychiatry*, 36(6); pp. 844 – 852; 1997.
- [11] Bolton D, Luckie M, Steinberg D, “Long – term course of obsessive-compulsive disorder treated in adolescence”; *J Am Acad Child Adolesc Psychiatry*, 34; pp. 1441 – 1450; 1995.
- [12] Catherine K. Kraper, Timothy W. Soto and Alice S. Carter, “Obsessions and Compulsions: The Development and Familial Context”; *Handbook of Development and Psychopathology*; pp. 561 – 579; 2014.
- [13] Daniel A. Geller, “Obsessive-compulsive and spectrum disorders in children and adolescents”; *Psychiatric Clinics of North America*, 29; pp. 353 – 370; 2006.
- [14] Thomsen PH, “Obsessive-compulsive disorder in children and adolescents: Self-reported obsessive-compulsive behaviour in pupils in Denmark”; *Acta Psychiatr Scand*, 88; pp. 212 – 217; 1993.
- [15] Luciano C. Nunes, Placido R. Pinheiro, Mirian C. D. Pinheiro, Marum S. Filho, Rafael E. C. Nunes and Pedro G. C. D. Pinheiro, “Automatic Detection and Diagnosis of Neurologic Diseases”; *IEEE Access*, Vol. 7; pp. 29924 – 29941; 2019.
- [16] Dan J. Stein, “Obsessive-compulsive disorder”; *The Lancet*, Vol. 360; pp. 397 – 405; 2002.
- [17] Oguz Tan, Mehmet Akif Ozcoban, Serap Aydin, “Classification of EEG Synchronization Values of Obsessive Compulsive Disorders Patients using Support Vector Machine method”; *Medical Technologies National Congress*; pp. ; 2016.