A Data-Driven Approach for the Analysis of Behavioral Disorders With a Focus on Classification and Severity Estimation

Rama Krishna Thelagathoti and Hesham H. Ali
College of Information Science and Technology
University of Nebraska Omaha
Omaha, NE 68182, USA
rthelagathoti@unomaha.edu and hali@unomaha.edu

2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) | 978-1-6654-6819-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/BIBM55620.2022.9994876

***Abstract*—It has been well-established that behavioral disorders such as depression and schizophrenia are hard to diagnose. Classifying the type of the disorder is a challenging task because of the diversity of the symptoms. Furthermore, estimation of the severity level of such disorders is another challenging factor in the diagnosis process. Current clinical diagnostic procedures tend to be mostly dependent on the self-reporting of patients or the limited observational evaluations of clinicians. Hence, there is a need for objective analytical methods that leverage the current informatics advancements in studying behavior disorders. In this study, we proposed two major contributions. First, we introduce a data-driven approach that takes advantage of the available public databases to differentiate between multiple behavioral disorders. Second, we developed an index (Behavioral Health Score (BHS)) to assess the severity level. The proposed approach is based on the established correlations between mobility and health by analyzing the mobility data collected from 72 participants suffering from multiple behavioral disorders using wearable sensors. Obtained results demonstrate that the proposed correlation network model can distinguish between different types of disorders and BHS index can be utilized to estimate the severity levels of the disorder. Moreover, the proposed model can add a useful information to healthcare providers to better diagnose behavioral conditions and develop data-driven treatment options.**

***Index Terms*—behavioral disorders, correlation networks, severity assessment, data-driven model**

I. Introduction

The number of people suffering from behavioral disorders including depression and schizophrenia is raising at an alarming rate. According to the World Health Organization (WHO), 13% of the world population is suffering from at least one behavioral health problem [1]. As per the 2019 statistics, 280 million people have suffered from depression and 24 million individuals have experienced schizophrenic symptoms [1]. It is an undeniable fact that behavioral disorders have a significant impact on the overall well-being of an individual. Especially patients experience diverse symptoms such as disturbed mood, withdrawal from social gatherings, feeling of worthlessness, and hopelessness [2]. Because of the heterogeneity in the type of symptoms it is challenging to classify the type of the disorder. Moreover, estimating the severity level of such disorder is another difficult task. Existing diagnostic methods for classification and severity assessment are predominantly subjective. Because overall diagnosis is made based on either self-reporting feedback or observation by a trained physician or both [3] [4]. However, the holistic assessment is driven by human feedback and observation.

To reduce human involvement, the scientific community has proposed numerous guidelines like the Diagnostic and Statistical Manual of Mental Disorders (DSM-V edition), and the International Classification of Diseases (ICD) to classify the type of the disorder [5]. Furthermore, several diagnostic procedures are available to measure the severity level of the disorder. For example, the MADRS score is a symptom rating scale to estimate the severity level of the depression and likewise, the BPRS score for schizophrenia [3] [4]. Albeit, most of these procedures are symptom rating scales that are primarily performed by limited human observation. There are several drawbacks to these kinds of procedures. First, identifying the type of the disorder from symptoms is challenging and, also rating the disorder requires enormous human effort. Second, the decision is solely dependent on human feelings and judgment. Third, to determine the efficacy of antipsychotic treatment or interventional therapy, an evaluation of the severity is required for each visit. This needs patients to visit labs and hospitals frequently. Thus, these methods are not efficient as there is a lack of objective diagnosis. To overcome these limitations, this paper puts forth two contributions that have the potential to assist physicians and experts to evaluate the type of the disorder and estimate its severity.

Recent studies demonstrate that the mobility of people suffering from behavioral disorders is lower than that of healthy individuals [6] [7]. Furthermore, at least one motor sign has been prevalent in these people. Thus, mobility can be considered as an alternative approach to identifying the disorder. The two contributions of this paper are fundamentally built by analyzing the mobility data collected from 72 subjects consisting of healthy participants (control group) as well as individuals suffering from behavioral disorders (condition group). Our first contribution is to build a computational model to discriminate between multiple psychological disorders. Our second contribution is to develop an index that can identify the severity level of the disorder. In this document, we refer motor activity data as mobility data. The rest of the manuscript is organized as follows. Section II describes the related work carried out on the same problem and the methodology is presented in section III.

II. Materials and methods

A. The Overview

Fig. 1 shows the pipeline for the proposed computational model. The pipeline begins with acquiring a dataset. After preprocessing the data, relevant features are extracted. The network construction step is central to the overall methodology. In this step, a graph is constructed by considering all the subjects as nodes and their interrelationships as edges. The process of clustering identifies subgroups in the graph that will eventually categorize the subjects according to their behavior health condition. Finally, the BHS index is computed to estimate the illness severity.

B. Dataset description

The dataset used for this study is obtained by merging two independent datasets published by the same research group [8] [9]. The first dataset is the ‘Depresjon’ dataset consisting of the motor activity data of 55 adults 23 of them are suffering from either unipolar or bipolar depressive disorder and 32 of them are healthy individuals [8] . The second dataset is the ‘Psykose’ dataset comprising 22 subjects that are diagnosed with schizophrenia and 32 subjects the healthy counterparts [9]. The healthy participants in both datasets are the same persons. Since the healthy group is the same between the two datasets, we have generated a new dataset by combining the ‘Depresjon’ and ‘Psykose’ datasets. As a result, the new dataset contains 3 groups of subjects: depressive, schizophrenic, and healthy. Besides, depressive, and schizophrenic subjects are considered as the condition group, and healthy participants are considered as the control group as described in Table I.

Data collection procedure is explained elsewhere [10]. Most of the participants have shared their motor activity data for 13 days. However, a few participants have shared less than 13 days. Besides, the dataset contains two kinds of information, general demographic information including age and gender, and clinical information such as the MADRS score [3] and BPRS score [4]. The MADRS score and the BPRS score

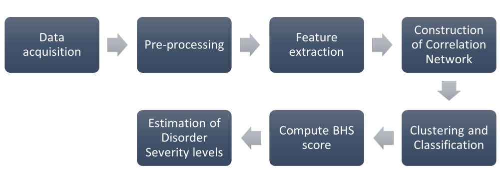


Fig. 1: The pipeline of methodology

TABLE I: Multiple groups in the dataset

Main group	Subgroup	Description	Count
Condition	Depression	Diagnosed with depression	23
Condition	Schizophrenia	Diagnosed with schizophrenia	22
Control	Healthy	Healthy and not diagnosed with any psychological disorder	32
		Total	77

indicate the severity levels of depression and schizophrenia respectively. In general, a higher score represents higher levels of illnesses.

C. Classification of multiple behavioral disorders

feature extraction and construction of correlation network are described in [10]. In the context of this study, the classification task is to find the hidden subgroups that exhibit similar mobility patterns. We hypothesize that mobility patterns are distinguishable between people with behavioral disorders and healthy individuals. In addition to this, the mobility patterns are not only distinguishable between patients with disorder and healthy group but also among the patients suffering from different psychological disorders including depression and schizophrenia. It implies the mobility characteristics of depressed and schizophrenic patients are distinct. In this process, First, we have constructed a correlation network graph (CNG) by utilizing the mobility data. The CNG is a graph with 77 subjects as nodes and the pair-wise correlation as edges. Then, we employed Markov Clustering (MCL) technique to uncover the hidden subgroups from the main group of 77 subjects. A cluster is a subgraph in which all nodes in the cluster are densely connected to the nodes with in the same cluster and sparsely connected to nodes in the other clusters [11]. The MCL algorithm is a popular unsupervised clustering algorithm that is well suitable for extracting clusters in biological networks [12]. MCL works by a random walk property of a graph where all nodes are randomly visited to find the strongly connected components in the graph. Each identified cluster (subgroup) signifies a strongly connected group of subjects with similar mobility profiles. The significance of this methodology is that subgroups are identified without using any supervised label which is already present in the dataset.

D. Estimation of Disorder Severity

Assessment of disorder severity is one of the critical factors in holistic diagnosis. Unfortunately, there is no clinical test that can measure the severity of mental disorders. Current methods for severity estimation are merely dependent on self-reporting symptoms and trained physician observation [3] [4] [13] [14]. Moreover, estimating the disorder severity using computational techniques is always not approachable. In this manuscript, we are introducing a quantitative metric BHS that can possibly reveal the severity of the disorder. Although the formalization of the BHS is still in progress, our interim results proves that there is enough evidence for the strong relationship between measured BHS and clinician rating. The constructed CNG in the previous step is a graph where 77 subjects are represented as nodes and any two nodes are connected by an edge if they have a strong interrelationship with respect to their mobility features. We hypothesize that the constructed CNG has hidden information that can plausibly reveal the severity of the disorder. The BHS is computed as follows.

1) Preliminaries:

- Let G = (V,E) be a CNG where V = { P1, P2, ..., p77 } which means each node in CNG is a participant from the dataset.
- Let E(Pj,Pm) is an edge between vertices Pj and Pm where (Pj,Pm) $\leq$ 77.
- A clustering C = { C_1,...,C_k } of G is a partition of the node set V into non-empty subsets and k is the total number of clusters.
  Where C_i, i $\leq$ k, is a cluster. In other words, C_i is a sub graph of G.
- assume h is the total number of nodes in C_i
- Now compute sum of inter-cluster edges and sum of intra-cluster edges for each vertex in C_i
- S_{ij} is the sum of the inter-cluster edges of cluster C_i and vertex Pj. S_{ij} can be computed as follows.

$$S_{ij} = \sum_{1}^{h} E(Pj, Pm) \tag{1}$$

  h is the total number of nodes in C_i, where E(Pj,Pm) is an edge between Pj and Pm , Pj $\in C_i$, and Pm $\notin C_i$. S_{ij} is computed for all the nodes in C_i, i $\leq$ k .
- Similarly, $S^{'}_{ij}$ is the sum of the intra-cluster edges of cluster C_i and vertex Pj. $S^{'}_{ij}$ can be computed as follows.

$$S^{'}_{ij} = \sum_{1}^{h} E(Pj, Pm) \tag{2}$$

  h is the total number of nodes in C_i, where E(Pj,Pm) is an edge between Pj and Pm , Pj $\in C_i$, and Pm $\in C_i$. $S^{'}_{ij}$ is computed for all the nodes in C_i, i $\leq$ k .
- Community nodes - A node is said to be the ‘Community node’ of a cluster if all its edges are connected to the nodes that belong to the same cluster.
- Bridge nodes - A node is said to be the ‘Bridge node’ of a cluster if its edges are connected to nodes that belong to the same cluster and also to other clusters.

2) calculation of BHS index:

$$BHS(Pj) = \frac{S_{ij}}{S^{'}_{ij}} \tag{3}$$

where S_{ij} is sum of inter-cluster edges, $S^{'}_{ij}$ is sum of intra-cluster edges, and Pj is an jth particpant (j $\leq$ 77)

3) Interpretation of BHS index: the value of the BHS index is a spectrum that ranges between 0 and 1. However, we use either 0 or 1 as a binary categorization of the nodes in each cluster. Having said that, BHS index can be tweaked. We assume that the experts those have sufficient knowledge on behavioral disorders can be able to tweak and alter the results. In this current study, we provide a binary categorization of the disorder severity. It implies, If BHS(Pj) = 0 then node Pj is a community node of the cluster, and all its edges are connected to the nodes that belong to the same cluster. Whereas if BHS(Pj) = 1 then Pj is the bridge node of the cluster, and all its edges are connected not only to the nodes that do not belong to the same cluster but also to the nodes in adjacent clusters.

BHS(Pj) $= 0 \longrightarrow$ community node
BHS(Pj) $\geq 0 \longrightarrow$ bridge node

The BHS index is derived from the property of homogeneity and separation of the clustering process. In general, a good clustering algorithm produces highly homogeneous and well-separable clusters [12]. It means that clusters must have a high volume of intra-class edges and a low volume of inter-cluster edges. The same concept is adapted to build the BHS index function. We divide all the nodes in the same cluster into two groups based on their BHS index. Therefore, we propose that the community nodes are tightly coupled to their cluster and so their disorder severity is High whereas boundary nodes are loosely connected and their disorder severity is Low compared to the community nodes.In the case of healthy subjects, the interpretation is different. Since healthy people do not have any illness, we categorize them based on their mobility profile as shown in Table II.

III. Results

A. Classification of multiple behavioral disorders

As mentioned earlier in Table I, the dataset contains 3 subgroups and they are categorized into 2 main groups. The aim of the classification task is to process the mobility data pertaining to 72 subjects and differentiate different subgroups hidden in the main group. The novelty of our approach is that the classification task is totally driven by data and does not include any supervised label. Hence, our approach is completely data-driven rather than label-driven. As described in the previous section we first constructed the CNG and then employed the MCL clustering algorithm to identify subgroups. Then, our methodology was able to detect three major subgroups as depicted in Fig. 2(a). It is evident from Fig. 2(a) that the CNG has three major clusters and also there is a fair separation between three clusters. Although, the method itself can not reveal the name of the group, from the dataset we can understand that these groups are depressed, schizophrenia, and healthy.

B. Estimation of disorder severity

The classification task identifies three major subgroups from the main group. Now, the BHS index is computed for each subject in the cluster using equation 3. By using Table II, all the subjects in each subgroup are divided into 2 categories: community nodes and boundary nodes. Subjects who fall under the community nodes are the people with high severity of the disorder whereas the subjects under the boundary nodes with lower severity. In the case of a healthy group, severity does not make any sense. So, the community nodes and boundary nodes are categorized according to their mobility composition. To understand the figure better, we have used a different color codes, as explained in Table II. Different colors have been used for each category for fine-grained representation in Fig. 2(b). The next section elaborates on the validation of the results concerning the clinical data available in the dataset.

IV. Validation

To validate our results, we take the advantage of the symptom rating scores present in the dataset. The dataset contains symptom rating scores for both depression and schizophrenia. The MADRS score indicates the seriousness of the depressive disorder while the BPRS score represents the severity of the schizophrenia. Since healthy people do not have any disorder, they are excluded from validation. Besides, subjects without clinical scores are also excluded from the validation.

A. Depression group

The MADRS score is the symptom rating score measured by expert clinicians [3]. In Fig. 2(b). subjects belonging to community nodes are colored in red while boundary nodes are colored in purple. Their respective MADRS scores are plotted with the same colors in Fig. 3(a). It is clear from Fig. 3(a) that most of the subjects who are categorized as High severity by our methodology are also having high MADRS scores. This demonstrates that there is a strong correlation between the measured BHS index and the MADRS depression rating score.

B. Schizophrenia group

The dataset contains the BPRS score that represents the severity of the schizophrenia group. However, we can not find a strong relationship between measured BHS index and BPRS score is as shown in Fig. 3(b). There could be several reasons behind these results. Even though depression and schizophrenia are classified as behavioral disorders, there are many differences in terms of diagnosis and estimation of the severity. According to DSM-5 guidelines, there is no perfect test that can reveal the seriousness of the disorder [5]. Furthermore, for an objective conclusion, a history of six months along with functional impairment has to be assessed [15]. Besides, the dataset tells that schizophrenic patients are using multiple medications such as clozapine, and mood stabilizers. Though we do not contradict the BPRS rating, we believe that further investigation by expert clinicians is required.

V. Discussion

The acquired dataset contains two main groups out of which one main group consists of individuals suffering from multiple psychological disorders and the second main group is healthy counterparts. The major focus of this study is two folds. First, classify the main group and identify subgroups by analyzing their mobility data. Second, estimate the severity level. The dataset contains the supervised label for each which indicates the subgroup that the subject belongs to. However, We have used labels only for validation. Although the interpretation of BHS index is well defined for condition, the healthy group has a different interpretation.Despite the fact that this group is healthy, many of them are connected to either depression or schizophrenia group. many studies prove that reduced mobility is one of the early signs of illness [10]. Having additional clinical history such as their medical history would help in further investigation of these subjects. However, people who are categorized as community nodes are having high mobility and they seem to be healthier than boundary nodes. Similarly, boundary nodes in the depression group have edges incident to the nodes in the schizophrenia group and vice versa. This might be the indication of the presence of multiple behavioral disorders [5] [15].

One of the limitations of this study is that the proposed BHS metric is a binary value: High or Low. However, our future plan is to further divide the subgroup and identify multiple categories such as mild, moderate, severe, and very severe. This requires exploring several other graph properties including graph centrality measures. Current results may contain some artifacts where BHS might not reflect the actual clinical score for some individuals. This is one of the limitations of the study. Albeit, it is always not possible to secure perfect results in the biological field, . Another limitation of our study is that certain nodes are isolated and disconnected from the graph. They are formed as singleton clusters. It is mainly because their mobility is distinct and not correlated with any other node.

VI. Conclusion

Behavioral disorders including depression and schizophrenia are difficult to diagnose. Because of the variations in the symptom categories, it is extremely complicated to classify the type of the disorder and estimation of the disorder severity. In this manuscript, we have introduced a data-driven approach that can classify multiple behavioral disorders and estimate their severity level. We processed the mobility data and built a correlation network graph by establishing pairwise correlation. Then, we employed the MCL clustering technique to identify the strongly connected clusters. Our results demonstrate that the constructed graph is able to differentiate between multiple psychological disorders as well as a healthy group. Furthermore, we introduced an index (BHS) that can help in estimating the severity of the disorder. We believe that the proposed approach would facilitate healthcare providers to better diagnose behavioral conditions and treat with computational outcomes.

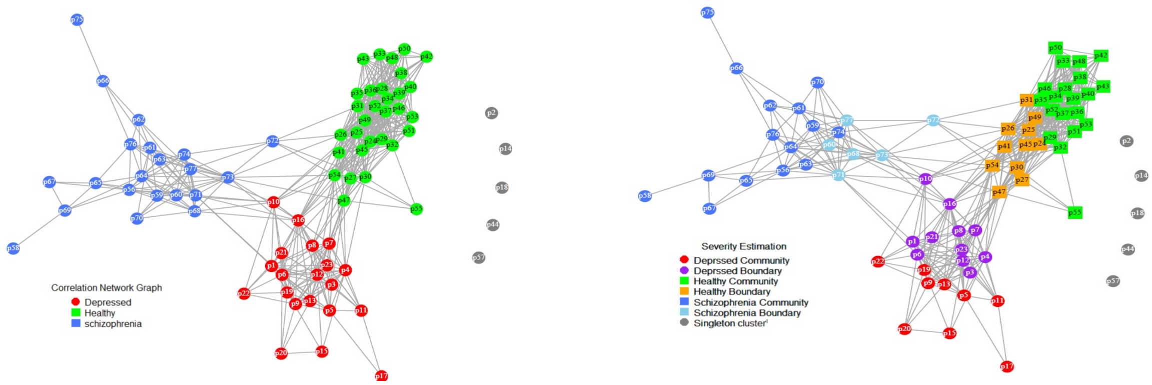
References

[1] Mental disorders, “Mental disorders,” 2022, [Online; accessed 17-August-2022]. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/mental-disorders

[2] The National Institute of Mental Health (NIMH), “Depression,” 2021, [Online; accessed 17-August-2022]. [Online]. Available: https://www.nimh.nih.gov/health/topics/depression

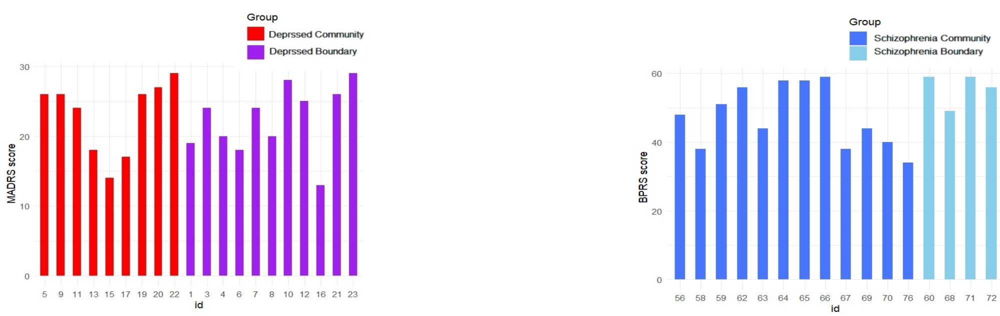
TABLE II: Results of Severity estimation

Main Group	Subgroup	BHS index	Color Code	Node category	Interpretation
Condition	Depression	BHS(Pj) = 0 BHS(Pj) > 0	Red Purple	Community Node Boundary Node	Depression Severity High Depression Severity Low
Condition	Schizophrenia	BHS(Pj) = 0 BHS(Pj) > 0	Blue Skyblue	Community Node Boundary Node	Schizophrenia Severity High Schizophrenia Severity Low
Control	Healthy	BHS(Pj) = 0 BHS(Pj) > 0	Green Orange	Community Node Boundary Node	Healthy and high mobility Healthy but decreased mobility



(a) Classification (b) Severity Estimation

Fig. 2: Classification of multiple Psychological disorders



(a) Depression group (b) schizophrenia group

Fig. 3: Validation

[3] S. A. Montgomery and M. Åsberg, "A new depression scale designed to be sensitive to change," *The British journal of psychiatry*, vol. 134, no. 4, pp. 382–389, 1979.

[4] J. E. Overall and D. R. Gorham, "The brief psychiatric rating scale," *Psychological reports*, vol. 10, no. 3, pp. 799–812, 1962.

[5] F. Edition *et al.*, "Diagnostic and statistical manual of mental disorders," *Am Psychiatric Assoc*, vol. 21, no. 21, pp. 591–643, 2013.

[6] D. Todder, S. Caliskan, and B. T. Baune, "Longitudinal changes of day-time and night-time gross motor activity in clinical responders and non-responders of major depression," *The World Journal of Biological Psychiatry*, vol. 10, no. 4, pp. 276–284, 2009.

[7] S. Walther and W. Strik, "Motor symptoms and schizophrenia," *Neuropsychobiology*, vol. 66, no. 2, pp. 77–92, 2012.

[8] E. Garcia-Ceja, M. Riegler, P. Jakobsen, J. Tørresen, T. Nordgreen, K. J. Oedegaard, and O. B. Fasmer, "Depresjon: a motor activity database of depression episodes in unipolar and bipolar patients," in *Proceedings of the 9th ACM multimedia systems conference*, 2018, pp. 472–477.

[9] P. Jakobsen, E. Garcia-Ceja, L. A. Stabell, K. J. Oedegaard, J. O. Berle, V. Thambawita, S. A. Hicks, P. Halvorsen, O. B. Fasmer, and M. A. Riegler, "Psykose: A motor activity database of patients with schizophrenia," in *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE, 2020, pp. 303–308.

[10] R. Thelagathoti and H. Ali, "A population analysis approach using mobility data and correlation networks for depression episodes detection," *Ann Depress Anxiety*, vol. 9, no. 1, p. 1112, 2022.

[11] S. Emmons, S. Kobourov, M. Gallant, and K. Börner, "Analysis of network clustering algorithms and cluster quality metrics at scale," *PloS one*, vol. 11, no. 7, p. e0159161, 2016.

[12] S. Van Dongen, "Performance criteria for graph clustering and markov cluster experiments," in *National research institute for mathematics and computer science in the*. Citeseer, 2000.

[13] N. Bayley, *Bayley-III: Bayley Scales of infant and toddler development*. Giunti OS Florence, Italy, 2009.

[14] Y. R. Burns, R. M. Ensbey, and M. A. Norrie, "The neuro-sensory motor developmental assessment part 1: development and administration of the test," *Australian Journal of Physiotherapy*, vol. 35, no. 3, pp. 141–149, 1989.

[15] F. F. Ferri, *Ferri's clinical advisor 2013: 5 Books in 1*. Elsevier Health Sciences, 2012.