2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Clustering and Feature Analysis of Smartphone Data for Depression Monitoring

Binh Nguyen[1,*], Sharadha Kolappan[2], Venkat Bhat[2,3], and Sridhar Krishnan[1]

[1]Department of Electrical, Computer, and Biomedical Engineering, Ryerson Univerity, Toronto, ON M5B 2K3
[2]Interventional Psychiatry Program, St. Michael's Hospital, Toronto M5B 1W8
[3]Department of Psychiatry, University of Toronto, Toronto, ON M5S 1A1
[*]Corresponding author: Binh Nguyen, binh.nguyen@ryerson.ca

Abstract— **Modern advancements have allowed society to be at the most innovative stages of technology which involves the possibility of multimodal data collection. Dartmouth dataset is a rich dataset collected over 10 weeks from 60 participants. The dataset includes different types of data but this paper focuses on 10 different smartphone sensor data and a Patient Health Questionnaire (PHQ) 9 survey that monitors the severity of depression. This paper extracts key features from smartphone data to identify depression. A multi-view bi-clustering (MVBC) algorithm is applied to categorize homogeneous behaviour subgroups. MVBC takes multiple views of sensing data as input. The algorithm inputs three views: average, trend, and location views. MVBC categorizes the subjects to low, medium and high PHQ-9 scores. Real-world data collection may have fewer sensors, allowing for less features to be extracted. This creates a focus on prioritization of features. In this body of work, minimum redundancy maximum relevance (mRMR) is applied to the sensing features to prioritize the features that better distinguish the different groups. The resulting MVBC are compared to literature to support the categorized clusters. Decision Tree (DT) 10-fold cross validation shows that our method can classify individuals into the correct subgroups using a reduced number of features to achieve an overall accuracy of 94.7±1.62%. Achieving high accuracies with reduced features allows for focus on low power analysis and edge computing applications for long-term mental health monitoring using a smartphone.**

I. Introduction

Depression is a common mental illness that affects more than 264 million people [1]. While, there are many known methods and effective treatment options for mental disorders like depression, approximately 76% and 85% of people in low- and middle-income countries do not receive treatment, respectively [2]. Early detection and treatment of depression, promotes remission, reduces relapse and lowers financial burden of the disease [3]. Having said that, these benefits can be realized thanks to recent technological advancements which have given opportunity to digital phenotyping.

Digital phenotyping is defined as data collection of people's phenotypes using digital devices. The popularity of digital phenotyping is reflected in the ubiquity of modern smartphones, allowing for collection of geolocation, social media, and other behavioral and physiological recording [4]. Moreover, the adoption of technology is supported by recent studies which have proven the use of smartphone data to extract features for complex computation related to depressive mood [5].

The use of smartphone devices has been widely adopted in various applications, including clinical setting (digital psychiatry) [6] and outside of clinical settings (education, employment, financial services) [7]. Modern smartphones have the ability to collect activity data [8], speech [9], circadian [10] and social interactions [11]. Similar works have incorporated smartphone sensors to classify the severity of depression in individuals. In Masud et al. [12], a dataset composed of 33 subjects from Bangladesh was collected with smartphone and PHQ-9 information. A support vector machine (SVM) classifier was trained using 12 features and achieved an accuracy of 86.4% for prediction of depression levels. Moreover, Farhan et al. [13] used wrapper method and SVM for feature extraction and classification, respectively, for analysis on the Studentlife dataset from Dartmouth [14]. Average, trend, and location matrix views were created from the smartphone sensing data and applied to MVBC to create behaviour subgroups. The key features were then used for a 10-fold SVM classifier to classify the subgroups with an overall accuracy of 87.1%. The study by Grunerbl et al. [15] was based on individuals in psychiatric hospital where participants' collected the fusion of accelerometer, sound, phone calls and GPS data for detecting mental state change and achieved an accuracy of 76%.

Digital technology has given a new perspective on understanding the relationship between behavior and depression. Although, mental health affects people differently and may fluctuate across time, a method to champion this concern is to use cluster analysis to identify the behavioural subgroups and feature selection for efficient continuous long-term monitoring. This paper aims to identify user subgroups of depression while using reduced features. To validate our work, machine learning (ML) is applied to the data to efficiently and accurately classify the users into the correct subgroups. The works by Farhan et al. [13] and Masud et al. [12] inspired our hypothesis of optimally classifying behavioral subgroups. Our proposed paper replicated [13] methodology with some challenges but achieved an overall higher accuracy with a smaller feature set. MVBC is applied to identify the homogenous clusters of individuals [16], which represent subgroups of user behavior indicative of depression, specifically the PHQ-9 score. Followed by the MVBC algorithm, we use mRMR [17] to create an optimal feature. mRMR is a feature selection tool based on the criterion that combines minimal-redundancy and maximal-relevance [18]. The reduced feature set is then applied with a DT classifier due to its lightweight characteristics and effectiveness on small datasets. The effectiveness of the algorithm is evaluated using a publicly available dataset collected from the StudentLife project at the Dartmouth College [14]. The motivation of this work is to reproduce and validate [13], while achieving a better accuracy and improving the effectiveness and efficiency of the ML model. By improving the efficiency, we can promote the application of edge computing for long-term monitoring of mental health using a smartphone. Furthermore, the novelty aspect of this work is the use of a lightweight model that suits real-world needs.

II. Dataset and Methods

The dataset used was the Dartmouth Studentlife data [14]. The dataset involves data collection of 60 college students for over 10 weeks. Students would use their smartphones to collect passive and automatic sensing data from built in sensors, self-report surveys and academic performance. This paper focuses only on the 10 sensing data and the PHQ-9 survey. The sensing data include physical activity, audio inferences, conversation inferences, Bluetooth scan, light sensor, global positioning system (GPS), phone charge, phone lock, WiFi, WiFi location. Moreover, PHQ-9 is a questionnaire involving 9 different questions to evaluate the severity of the participant's depression. The PHQ-9 scale has a 0-27 range with cut-offs values for the diagnosis; from no depression to severe depression. During the experiment, students were required to take an entry and exit questionnaire. Additionally, the sensing features were calculated continuously and automatically throughout the experiment. Each sensor and data had its respective sampling frequency. More information can be found on the Dartmouth StudentLife website [14].

The methodology used is composed of five steps: data pre-processing, feature extraction, clustering, feature ranking and then classification of subjects into the identified user clusters.

1) Pre-processing: During data pre-processing, users that did not complete their second questionnaire questions, also known as the exit survey, would be represented as incomplete data and be removed.

2) Feature Extraction: The feature extraction is composed of three views. These views include the average view, trend view and location view. Each view contains a set of extracted features that can be represented in a matrix.

The average view contains a set of features and is calculated by taking the daily value of the feature and then averaging it over all the days that a participant is enrolled. This view reflects the participant's overall behavior. A complete list of the features and description can be seen in Table I.

TABLE I: Features extracted for Average View

Sensor data	Feature description
Activity	Act_s, Act_w, and Act_r which represent activities for stationary, walking and running, in a day, respectively.
Conversation	$Convo_d$ and $Convo_c$ which represent conversation of total duration and number of conversations, per day, respectively.
Light	$Dark_d$ $Dark_c$ which represent the total duration and number of times when a user is in a dark environment, per day, respectively.
Audio	$Audio_q, Audio_n$, and $Audio_v$ which represent the duration of which user's are classified as quiet, noisy and voice, per day, respectively.
Phone lock	$PhoneLock_d$ and $PhoneLock_c$ which represent the total time for phone lock duration and the number of times phone is locked, per day, respectively.

The trend view is calibrated to determine the variation of several sensors over the period of the study. The trend view includes variation of the following: activity (walking), audio (noise) and conversation duration. To extract the features in trend view, two signal processing techniques were applied to each sensor data. The wavelet transform filter was used to remove noise and maintain the overall trends of the signals. Specifically the Haar wavelet in the transform was used to preserve the peak and trends of a time series curve. Secondly, the filtered signal was fitted to the least squares problem in Eq. 1 where $\bar{y}_d$ is the denoised daily-average on the d_{th} day and $\mathbf{c}$ is the parameters c1,...,c4. The features extracted included amplitude (c1), period (c2), phase (c3), and intercept (c4) from each individual's sensor [19].

$$\begin{aligned} &\min_c \sum_d (f(\mathbf{c}, d) - \bar{y}_d)^2 \\ &\text{subject to} f(\mathbf{c}, x) = c_1 \sin(\frac{2\pi}{c_2} x + c_3) + c_4 \end{aligned} \tag{1}$$

The location view was inspired by previous works which showed a significant correlation between location and depressive mood disorder [5], [20]. The location view involves various perspectives of location using the GPS data. The features in location view is composed of location variance, time in top 3 cluster locations (t_{c1}, t_{c2}, t_{c3}), entropy of participant location, normalized entropy, percentage of time a participant is spent at their home, percentage of time a participant is spent moving, and total distance covered. After extracting the features, the three views should represent three matrices.

3) Clustering: Following the three views, the MVBC algorithm is applied to identify homogeneous behavior groups [16]. The objective of the MVBC algorithm is to find subgroups of user behavior, which involves finding the same row clusters from all the views. The algorithm solves an objective functions that returns the approximation error with a Frobenius norm of the matrix difference. The details and objective function is further elaborated in [16]. The objective function returns a binary vector and the row clusters that are the same across the different views can be identified as the nonzero entries. This resulted in the identification of three clusters.

4) Feature ranking: Followed by the clustering of the subgroups, the mRMR algorithm was used to select a subset of the feature that are most correlated with the classes. mRMR was selected of the 19 feature ranking techniques [17] tested as it extracted the best features to represent the subgroups of behavior. Each feature ranking techniques was tested using the same classifier and the mRMR provided the best performance.

The mRMR is a feature selection algorithm is based on the the minimal (min)-redundancy and maximal (max)-relevance criterion [18]. The goal was to maximize the distance Φ between the max-Dependency and min-Redundancy as in equation (2). Maximum dependency is computationally expensive, thus, maximum relevance was introduced as a simpler approximation. Maximum relevance (D) between the subset of features $x_i \in S$ and the target class c was obtained as in equation (3). Redundancy among features was estimated by the MI values between two features. Minimum redundancy R calculation is provided in equation (4).

$$\max \Phi(D, R), \Phi = D - R \tag{2}$$

$$\max D(S, c), D = \frac{1}{|S|} \sum_{x_i \in S} I(x_i; c)) \tag{3}$$

$$\min R(S), R = \frac{1}{|s|^2} \sum_{x_i, x_j \in S} I(x_i; x_j) \tag{4}$$

5) Classification: Classification involved the separability of the subgroups that were identified from the MVBC algorithm. To assess the separability of the subgroups, a 10-fold cross validation DT classifier was constructed. The classifier was chosen due to its simplicity and effectiveness on small datasets. DT is a tree-like model of decisions and analyzes all the possible consequences of a decision. Additionally, the DT classifier was tested on the complete and reduced feature set to verify that the selected features can represent the clusters reasonably well.

III. Results

After the pre-processing, 38 users from the dataset were used. We use the MVBC algorithm from [16]. The parameters used were given by [13], where the hyperparameters s_{v^1}, s_{v^2} and s_{v^3} values were set to 5. Furthermore s_ω was set to 9 and 7 for cluster 1 and cluster 2, respectively. After parameter setting, the MVBC algorithm is applied and three clusters were identified, which had 9, 7 and 23 participants, respectively. To see if the clusters are discriminative of the PHQ-9 scores, we visualize the bar plot that present the relative mean values (RMV) of the PHQ-9 scores (average of pre and post) and the features for each of the three clusters. The RMV of the features is the difference between the mean samples in the cluster and mean samples in the entire samples, divided by the standard deviation (STD)

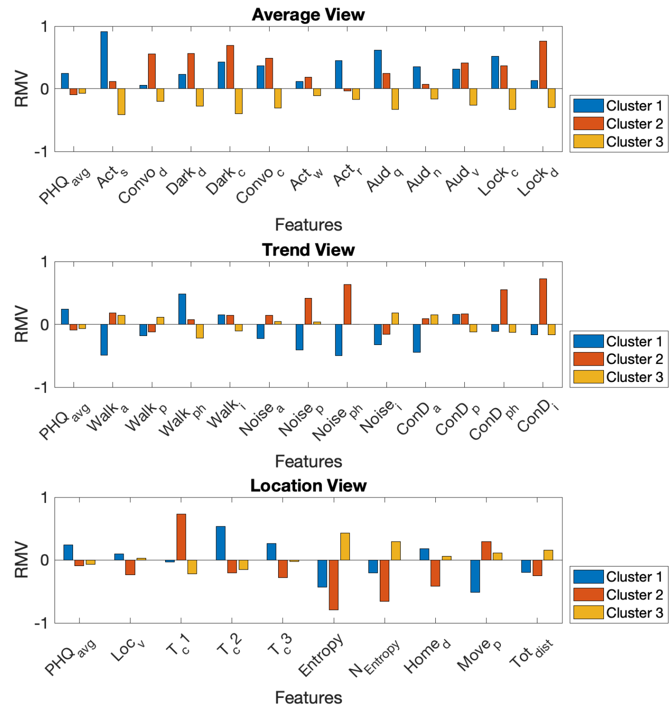

Fig. 1: The characteristics of the three clusters in the average view, trend view, location view

of the entire samples. The results for the three views are summarized in Fig. 1

An observation is that clusters 1, 2, and 3, have high, low and medium PHQ-9 scores, respectively. This observation indicates that MVBC was able to separate the subgroups to be discriminative of PHQ-9 scores. Cluster 1 represent high PHQ-9 score, which indicate more depressive mood disorders, while cluster 2 represent low PHQ-9 score, which do not represent depressive mood disorders. Furthermore, the feature relationship to cluster 1 and 2 are compared. It can be seen in the Fig. 1, that high PHQ-9 scores are associated with high Act_s, Aud_q, $Lock_d$, $Cond_i$, $Entropy$ and low $ConD_d$, $Walk_p$, $ConD_p$, $ConD_i$, $Walk_i$, $Move_{percentage}$. The opposite relationship of PHQ-9 scores and respective features are also valid. These feature relationships match the observations reported in [13].

Furthermore, mRMR was applied to rank the key features. The complete feature set have 33 features while the reduced features have 16 features. The number of features was determined to be 16 as it was the optimal number of features without the trade-off of classification accuracy. The reduced feature set includes $Noise_i$, Act_w, $Walk_a$, $Walk_{ph}$, T_c2, $Noise_a$, $ConD_i$, Act_r, $Walk_p$, $ConD_p$, Aud_n, Tot_{dist}, T_c3, $ConD_a$, Act_s, and $Convo_d$.

After clustering using MVBC algortihm, the clusters are labelled accordingly and the reduced feature set was used alongside a DT classifier to validate the separability. A 10-fold cross validation was used to achieve an overall 94.7±1.62% classification accuracy. Additionally, the confusion matrix of the results can be seen in Table II.

TABLE II: Confusion Matrix where C1, C2, and C3, represent Cluster 1, Cluster 2 and Cluster 3, respectively

Actual \ Predicted	C1	C2	C3
C1	9	0	0
C2	0	6	1
C3	0	1	21

IV. Discussion and Conclusion

This paper has presented an approach to identify homogeneous behavioral groups from a publicly available dataset. The MVBC algorithm was applied to the sensing data from the smartphone followed by mRMR feature ranking was used to obtain the rank of the key features. The reduced feature set had a total of 16 features, which is less than half of the complete feature list. Of the reduced feature set, trend view seems to be the most significant as 8 of the 16 features pertain to this view. These features include $Noise_i$, $Walk_a$, $Walk_{ph}$, $Noise_a$, $ConD_i$, $Walk_p$, $ConD_p$, and $ConD_a$. To emphasize this point, users with low PHQ score correspond to high $Walk_a$ and $ConD_i$ and low $Walk_p$. Having a high $Walk_a$ and $ConD_i$ are indicative of more walking and conversation as the intercept represents the approximate mean value. $Walk_p$ corresponds to a low frequency curve, representing a more stable pattern. Lastly, a DT 10-fold cross validation was used to identify the three clusters that are discriminative of PHQ-9 scores. DT was chosen for its ability to handle small datasets and its efficiency. Cross validation was used to create a generalized analysis. The classifier model achieved an overall accuracy of 94.7±1.62%.

The motivation of this work is to replicate and further develop the work in [13]. The novelty aspect includes the effective feature ranking, and emphasis of the trend view to create an efficient and effective ML model.

Due to the high accuracy with the use of a reduced feature set, detection of depression may be done locally on smartphones with future opportunities for low power secured connected healthcare and continuous human activity monitoring applications [21], [22].

Acknowledgment

The authors would like to thank NSERC and Mitacs for funding the project and Ryerson University and St. Michael's Hospital for the support.

References

[1] S. L. James, D. Abate, K. H. Abate, S. M. Abay, C. Abbafati, N. Abbasi, and et al., "Global, regional, and national incidence, prevalence, and years lived with disability for 354 Diseases and Injuries for 195 countries and territories, 1990-2017: A systematic analysis for the Global Burden of Disease Study 2017," *The Lancet*, vol. 392, pp. 1789–1858, nov 2018.

[2] P. S. Wang, S. Aguilar-Gaxiola, J. Alonso, M. C. Angermeyer, G. Borges, E. J. Bromet, and et al., "Use of MH services for anxiety, mood, and substance disorders in 17 countries in the WHO world MH surveys," *Lancet*, vol. 370, pp. 841–850, sep 2007.

[3] "Depression: the importance of diagnosis and prevention - ProQuest."

[4] P. I. Chow, K. Fua, Y. Huang, W. Bonelli, H. Xiong, L. E. Barnes, and B. A. Teachman, "Using mobile sensing to test clinical models of depression, social anxiety, state affect, and social isolation among college students," *Journal of Medical Internet Research*, vol. 19, p. e62, mar 2017.

[5] L. Canzian and M. Musolesi, "Trajectories of depression: Unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis," in *UbiComp 2015 - Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, (New York, New York, USA), pp. 1293–1304, Association for Computing Machinery, Inc, sep 2015.

[6] A. Ray, S. Sharma, and B. Sadasivam, "Exploring digital psychiatry through a pilot study: Pencil-and-paper versus digital neuropsychological testing," oct 2020.

[7] C. Burr, J. Morley, M. Taddeo, and L. Floridi, "Digital Psychiatry: Risks and Opportunities for Public Health and Wellbeing," *IEEE Transactions on Technology and Society*, vol. 1, pp. 21–33, jul 2020.

[8] E. Garcia-Ceja, V. Osmani, and O. Mayora, "Automatic Stress Detection in Working Environments from Smartphones' Accelerometer Data: A First Step," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, pp. 1053–1060, jul 2016.

[9] R. Wang, A. Dasilva, J. F. Huckins, W. M. Kelley, T. F. Heatherton, A. T. Campbell, W. Wang, . Todd, and F. Heatherton, "Tracking Depression Dynamics in College Students Using Mobile Phone and Wearable Sensing," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 2, 1, Article*, vol. 43, no. 1, p. 43, 2018.

[10] L. L. Craft and F. M. Perna, "The Benefits of Exercise for the Clinically Depressed," *The Primary Care Companion to The Journal of Clinical Psychiatry*, vol. 06, pp. 104–111, jun 2004.

[11] J. Y. Kim, N. Liu, H. X. Tan, and C. H. Chu, "Unobtrusive Monitoring to Detect Depression for Elderly with Chronic Illnesses," *IEEE Sensors Journal*, vol. 17, pp. 5694–5704, sep 2017.

[12] M. T. Masud, M. A. Mamun, K. Thapa, D. H. Lee, M. D. Griffiths, and S. H. Yang, "Unobtrusive monitoring of behavior and movement patterns to detect clinical depression severity level via smartphone," *Journal of Biomedical Informatics*, vol. 103, mar 2020.

[13] A. A. Farhan, J. Lu, J. Bi, A. Russell, B. Wang, and A. Bamis, "Multi-view Bi-clustering to Identify Smartphone Sensing Features Indicative of Depression," in *Proceedings - 2016 IEEE 1st International Conference on Connected Health: Applications, Systems and Engineering Technologies, CHASE 2016*, pp. 264–273, Institute of Electrical and Electronics Engineers Inc., aug 2016.

[14] R. Wang, F. Chen, Z. Chen, T. Li, G. Harari, S. Tignor, X. Zhou, D. Ben-Zeev, and A. T. Campbell, "Studentlife: Assessing MH, academic performance and behavioral trends of college students using smartphones," *UbiComp 2014 - Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 3–14, 2014.

[15] A. Grünerbl, A. Muaremi, V. Osmani, G. Bahle, S. Öhler, G. Tröster, O. Mayora, C. Haring, and P. Lukowicz, "Smartphone-based recognition of states and state changes in BPD patients," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, pp. 140–148, jan 2015.

[16] J. Sun, J. Lu, T. Xu, and J. Bi, "MVSCC via proximal alternating linearized minimization," *32nd International Conference on Machine Learning, ICML 2015*, vol. 1, pp. 757–766, 2015.

[17] G. Roffo, "Feature Selection Library (MATLAB Toolbox)," 2016.

[18] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 1226–1238, aug 2005.

[19] K. Polat and S. Güneş, "Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform," *Applied Mathematics and Computation*, vol. 187, pp. 1017–1026, apr 2007.

[20] S. Saeb, M. Zhang, C. J. Karr, S. M. Schueller, M. E. Corden, K. P. Kording, and D. C. Mohr, "Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: An exploratory study," *Journal of Medical Internet Research*, vol. 17, p. e175, jul 2015.

[21] B. Nguyen, Y. Coelho, T. Bastos, and S. Krishnan, "Trends in HAR with focus on machine learning and power requirements," *Machine Learning with Applications*, vol. 5, p. 100072, sep 2021.

[22] S. Krishnan, "Biomedical signal analysis for connected healthcare," *Elsevier*.