Automatic Verbal Analysis of Interviews with Schizophrenic Patients

Shihao Xu*, Zixu Yang†, Debsubhra Chakraborty‡, Yasir Tahir‡, Tomasz Maszczyk*, Victoria Yi Han Chua*
Justin Dauwels*, Daniel Thalmann‡, Nadia Magnenat Thalmann‡, Bhing-Leet Tan†,∮, Jimmy Lee Chee Keong†,§
*School of Electrical and Electronic Engineering, Nanyang Technological University
∮Health and Social Sciences, Singapore Institute of Technology
‡Institute for Media Innovation, Nanyang Technological University
§Lee Kong Chian School of Medicine; †Institute of Mental Health
Singapore

***Abstract*—Schizophrenia is a long-term mental disease associated with language impairments that affect about one percent of the population. Traditional assessment of schizophrenic patients is conducted by trained professionals, which requires tremendous resources of time and effort. This study is part of a larger research objective committed to creating automated platforms to aid clinical diagnosis and understanding of schizophrenia. We have analyzed non-verbal cues and movement signals in our previous work. In this study, we explore the feasibility of using automatic transcriptions of interviews to classify patients and predict the observability of negative symptoms in schizophrenic patients. Interview recordings of 50 schizophrenia patients and 25 age-matched healthy controls were automatically transcribed by a speech recognition toolkit. After which, Natural Language Processing techniques were applied to automatically extract the lexical features and document vectors of transcriptions. Using these features, we applied ensemble machine learning algorithm (by leave-one-out cross-validation) to predict the Negative Symptom Assessment subject ratings of schizophrenic patients, and to classify patients from controls, achieving a maximum accuracy of 78.7%. These results indicate that schizophrenic patients exhibit significant differences in lexical usage compared with healthy controls, and the possibility of using these lexical features in the understanding and diagnosis of schizophrenia.**

***Index Terms*—NLP; Schizophrenia; Speech Recognition; Langauge Feature; Machine Learning**

I. INTRODUCTION

Schizophrenia is a chronic mental disorder characterized by positive symptoms such as hallucinations or delusions; negative symptoms such as apathy, blunting of effect, or alogia; and cognitive impairments in attention, memory, and executive functions [1]. Language deficits associated with schizophrenia have been extensively studied since the last century, where these deficits have been observed lexically, sub-lexically, at sentence and discourse levels [2]. As language provides a wealth of interpretations regarding emotion and psychology [3], Natural Language Processing (NLP) is becoming a potential method for future research and clinical applications [4]. Coupled with the speed and ease of data collection brought about by technological advances [5], several data-driven applications using NLP methods have greatly contributed to the diagnosis, interpretation, and understanding of mental illnesses [6], [7]. We aim to keep up with the objective data-driven approach, and in this paper, we explore the potential of using NLP tools to aid automated analysis of verbal content generated by schizophrenic patients.

In earlier work [8], we explored Linguistic Inquiry and Word Count (LIWC) [9] as a tool for NLP. In this work, we explore two additional NLP tools: Diction [10] and Doc2Vec [11]. Tools such as LIWC and Diction allow us to explore changes in linguistic characteristics in schizophrenia patients [12]. Kei Hong et al. applied LIWC and Diction to analyze the linguistic features of autobiographical narratives that differentiate schizophrenia patients and controls [13]. They found distinct differences in the usage of words related to LIWC category *I* and Diction feature *self* between patients and healthy controls, and they found that patients were more likely to talk about the topics *money*, *trouble*, and *family*. Another study by Minor et al. [14] analyzed manually transcribed structured interviews with schizophrenic patients using LIWC and found that the number of *anger* words used in the interviews significantly predicted the severity of the positive symptoms.

Besides analyzing language differences and characteristics using manual lexicons like LIWC and Diction, document embeddings tools could also reflect the semantics of the document without using prior knowledge [11]. Yuan and colleagues applied Doc2Vec and Latent Dirichlet Allocation (LDA) to medical records of patients diagnosed with autism spectrum disorder (ASD), and were able to differentiate between ASD patients and healthy controls at a classification accuracy of 83% with a recall of 91% [15]. In [16], the authors applied Doc2Vec to tweets in the disease-based datasets on Twitter and achieved great classification results between disease and disease-free tweets with a recall of 84.6%.

In the above-mentioned studies, all texts analyzed were either written (digitally or handwritten), or manually transcribed from audio recordings. Moreover, we did not find any research which applied Doc2Vec [17] to the speech of schizophrenic patients. To this end, we applied both lexical analysis and document representation to automated transcriptions of interview recordings with schizophrenic patients and healthy controls. This paper is aligned with our overall research aim of developing objective and automated methods to evaluate schizophrenic patients. In this study (as in [8]), we explore the potential of harnessing lexical features and word representations, while in our earlier studies we investigated non-verbal cues [18] and movement [19].

II. Experimental Design

This experiment is held in conjunction with the Institute of Mental Health Singapore (IMH). A total of 75 participants were involved in the study: 50 *Patients* diagnosed with schizophrenia by a trained clinician from IMH, and 25 healthy *Controls*. All participants are recruited by the clinicians in IMH and are matched for age, gender, ethnicity, and education. The demographic information of the participants is displayed in Table I.

TABLE I
Demographic information of Patients and Controls.

		Patient (N = 50)	Controls (N = 25)
Gender	Male	25	11
	Female	25	14
Age	Mean (years)	30.3	29.92
	Range (years)	20-46	19-47
Education	University	7	4
	Diploma/JC/ITE	27	15
	High School	16	6
Ethnicity	Chinese	42	21
	Malay	5	3
	Indian	3	1

In this experiment, a professional psychometrician from IMH conducted a semi-structured interview with every participant. The participants were asked a certain list of predetermined questions, and subsequent follow-up questions were determined by the nature of answers given. Based on their responses, the psychometrician rated each participant on the Negative Symptoms Assessment (NSA-16) [20]. The NSA-16 evaluates patients along the following factors: communication, emotion, social involvement, motivation, and retardation. For each of the 16 items, a score ranging from 1 to 6 can be given to measure the behaviors of schizophrenia (1 denotes no symptoms, and 6 denotes severe symptoms). We did not pre-determine the time limit of the interview. On average, each interview lasted for around 25 minutes. We analyzed the entire length of the interview recordings (about 31 hours).

III. System Overview

In this section, we briefly discuss the main steps in our analysis: language-based features, speaker diarization, speech recognition, and the classification method.

A. Features

In this study, we group the features into two classes: lexical features extracted from the pre-existing dictionary and document embeddings generated through Doc2Vec. For the extraction of lexical features, we utilize two dictionary-based tools: LIWC 2015 (as in our earlier work [8]) and Diction 7.0, which extract lexical categories from the text files. On the other hand, we apply Gensim [21] to extract document embeddings where transformed every document to a vector.

1) LIWC and Diction: LIWC counts the different categories of words that related to the motivation, emotion, social and psychological states of people. For each set of transcribed text, we counted the number of words used that corresponded to 78-word categories outlined by the newest LIWC dictionary. The word count of each category was then normalized by the length of the interview. Diction scores any given passage according to 5 master categories (Activity, Optimism, Certainty, Realism, and Commonality) and 37 pertinent sub-categories (e.g., Centrality, Satisfaction). Scores for the 5 master categories are calculated by combining the z-scores of the sub-categories and then adding 50 to avoid negative values.

2) Doc2Vec: Different from the LIWC and Diction, Doc2Vec is an advanced NLP algorithm which represents a document as a vector and considers not only the distance between different words but also measures the differences between documents. This unsupervised algorithm represents a variance-length paragraph into a fix-length vector, which points the paragraph into a high dimensional space. We load GloVe pre-trained word embedding (GloVe.27B.50d.word2vect) to generate our document vectors [22], where Distributed Bag of Words of Paragraph Vector (PV-DBOW) model [11] was applied to train the Doc2Vec model with the fixed parameters. For each document, both the word vector length and document vector length are 50.

B. Speaker Diarization and Speech Recognition

In this experiment, the interview was recorded by a two-channel H4N recorder through two lapel microphones attached to the psychometrician and participant respectively. For the participant channel, the interference from psychometrician channel makes it occasionally hard to recognize the speech from the participant. Hence, we applied automated speaker diarization to segment out the participant's voice from the participant channel. We then preprocessed the audio to clear the patients' voices and filter out the interference as much as possible. The filtered audios were then subjected to Kaldi toolkit [23] to automatically transcribe them into text files. The pre-trained ASpIRE Chain Model [1] was used to off-line process the audios, and the official word error rate of ASpIRE Chain Model is 15.6% for generic conversational English. To test whether there is bias in our automatically extracted features, we manually transcribed 10 random recordings. We compared the total word counts of each LIWC category between these manual transcriptions and corresponding Kaldi's transcriptions. The relative words frequency histogram is presented in Fig.1, where we only present random 20 categories due to the limited space. It shows that the relative words frequency of Kaldi's transcriptions is almost in line with the actual distribution. The average of transcribed words per recording is 939.

C. Classification Method

In this experiment, the features that we extracted from text files were used to classify the schizophrenia patients and

[1] http://kaldi-asr.org/models.html

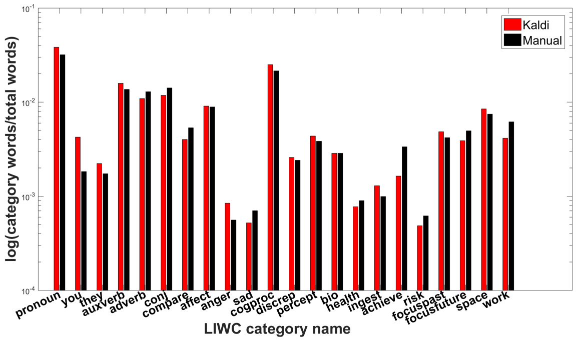

Fig. 1. Frequency histograms of word counts of LIWC categories of manual and Kaldi transcriptions.

healthy controls and predict the NSA-16 scores. We used the Scikit-learn toolkit [24] to perform leave-one-out cross-validation for classification and prediction tasks. In each cross-validation loop, one sample was held as the testing set, and the rest of the samples made up the training set. For classification of lexical features, we applied 10-fold cross-validation grid-search on the training set to select significant features (sorted by ANOVA F-value) and to optimize the parameters of 5 classifiers (Support Vector Machine (SVM), Logistic Regression (LR), GradientBoost, AdaBoost, and RandomForest). Additionally, in the 10-fold cross-validation grid-search, SMOTE over-sampling technique [25] and standardization were applied to balance and scale the training set. Finally, we used these 5 optimized classifiers to predict the testing data and soft vote the prediction scores to provide the final prediction of the testing set. For classification of document vector, we applied LR classifier and SMOTE over-sampling technique with leave-one-out cross-validation. The initializations and processes in the Doc2Vec network are randomized, rendering a different end-state document vector during each iteration of the program. Thus, we trained the model multiple times in each cross-validation fold and averaged the continuous output probabilities of LR to provide the final prediction of the testing set.

IV. ANALYSIS AND RESULTS

In this section, we present the results of our analysis. First, we present the correlation matrix between the lexical features and the NSA-16 scores. Next, we show the prediction results of the NSA-16 scale based on the features from LIWC and Diction. Finally, we present the classification results of schizophrenic patients and controls.

A. Correlation

We conducted the Pearson correlation test for the lexical features extracted from LIWC and Diction and the subjective NSA-16 scores. We only calculated the correlation coefficients for *Patients*, as the NSA questionnaire was not designed for *Controls*. Due to space limitations, only matrix plots of features with a correlation coefficient of at least ± 0.3 are presented in Fig. 2. Most of the LIWC features negatively correlate with NSA-16 scores, while most of Diction's features positively correlate with NSA-16 scores. Specifically, NSA2 (Restricted Speech Quantity), NSA6 (Reduced modulation of intensity), and NSA15 (Reduced expressive gestures) have the strongest negative correlations with lexical features used by patients during the interview.

B. NSA-16 Items Prediction

As we observed high correlations between several NSA-16 items and lexical features, we used lexical features to predict the scores of subject NSA-16 items. We categorized the NSA-16 scores into 2 classes: Unobservable (Class Low: for ratings of 1 and 2, which represent the absence of observable symptoms or behaviors related to schizophrenia) and Observable (Class High: for ratings of 3 and above, which represent the presence of observable symptoms or behaviors related to schizophrenia). We used LIWC and Diction features as input and applied the soft voting classification method outlined in Section III-C to predict the 16 NSA items. We filtered out the prediction results of NSA items with unbalanced sample size, and outline the prediction results (with AUC larger than 0.7) in Table III. We find that the NSA2 and NSA6 items could be well predicted by using lexical features.

C. Classification of Participants

Finally, we utilized the dictionary-based features and document embeddings to classify *Patients* (Class P, N=50) and *Controls* (Class H, N=25). We tested the accuracy of leave-one-out cross-validation for a total of 9 times, and present the median accuracy and standard deviation in Table II. The baseline was calculated by assuming all the samples belong to the majority class (e.g., assume all sample fall into the Class P). Notably, using LIWC features to classify patients and healthy controls provide the highest classification accuracy of 78.7% with a standard deviation of 3%. Across 9 experiments with the Doc2Vec algorithms, we achieved a median accuracy of 77.3% with a standard deviation of 1%.

TABLE II
PATIENTS V/S CONTROLS CLASSIFICATION RESULTS.

Feature		Confusion Matrix P	Confusion Matrix H	Precision	Recall	F-score	AUC	Accuracy	Baseline	STD
LIWC	P	44	6	0.81	0.88	0.85	0.81	**78.7%**	66.7%	0.03
	H	10	15	0.71	0.60	0.65				
Diction	P	37	13	0.66	0.74	0.70	0.57	57.3%	66.7%	0.02
	H	19	6	0.32	0.24	0.27				
LIWC Diction	P	40	10	0.80	0.80	0.80	0.74	73.3%	66.7%	0.05
	H	10	15	0.60	0.60	0.60				
Doc2Vec	P	40	10	0.85	0.80	0.82	0.84	77.3%	66.7%	0.01
	H	7	18	0.64	0.72	0.68				

V. DISCUSSION

In this section, we discuss the most salient lexical features to explore the differences in the speech produced by patients and controls. We applied the Kruskal-Wallis Test to all lexical features and computed the corresponding p-values. The most salient categories (with p-values smaller than 0.07) and their p-values are shown in Table IV, and these results are related to the questions the psychometrician asked during interviews.

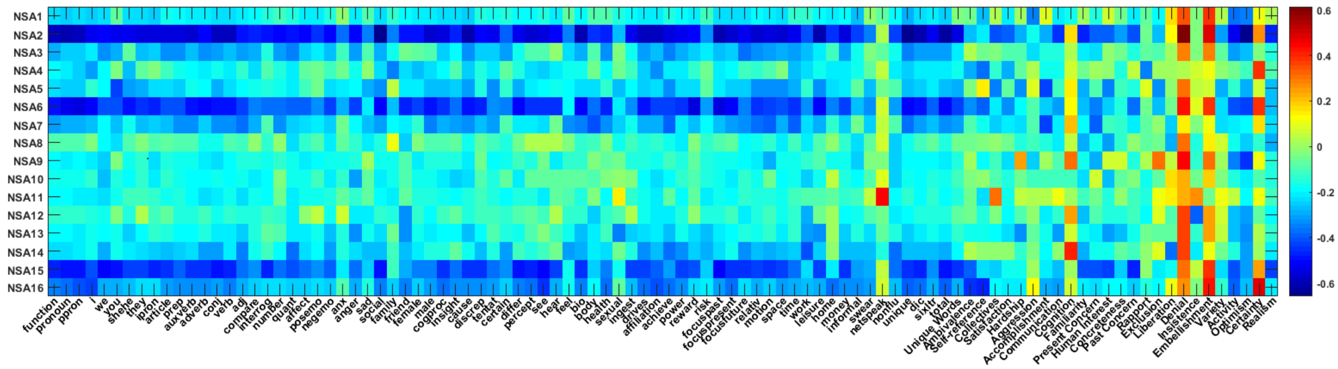

Fig. 2. Correlation coefficients of NSA scores with LIWC and Diction features.

TABLE III
PREDICTION RESULTS FOR NSA-16 ITEMS WITH LEXICAL FEATURES.

NSA Item	Feature	Confusion Matrix			Precision	Recall	F-score	AUC	Accuracy	Baseline	STD
			Low	High							
NSA2: Restricted speech quantity	**LIWC Diction**	**low**	14	7	0.78	0.67	0.72	0.82	78.0%	58.0%	0.03
		high	4	25	0.78	0.86	0.82	0.82			
NSA6: Reduced modulation of intensity	**LIWC Diction**	**low**	18	7	0.67	0.72	0.69	0.74	68.0%	50.0%	0.03
		high	9	16	0.70	0.64	0.67	0.74			
NSA15: Reduced expressive gestures	**LIWC**	**low**	10	8	0.53	0.56	0.54	0.70	66.0%	64.0%	0.04
		high	9	23	0.74	0.72	0.73	0.70			

When looking at individual LIWC categories, we observed that patients were less likely to use *assentive* words (e.g., yes, yeah, ok), *tentative* words (e.g., maybe, perhaps) and *informal* words (contain netspeak and assent categories, e.g., okay, cuz, oh). These results were similarly observed in citeShihao. Moreover, the differences in *unique* and *sixlter* categories indicate that healthy controls used more unique words and the words longer than six letters in the same length of time. On the other hand, we observed that schizophrenic patients in our study utilized more words in the LIWC category of *religion*, but fewer words in the LIWC categories of*work* and *achieve* during their interviews. Of note, we observed a similar trend in the comparable Diction's variable accomplishment (e.g., establish, finish, influence, proceed), as the schizophrenic patients in our study scored lower in this variable. Similar results have been found in other studies where the speech of healthy controls contained a higher count of the word working [13] and writings of schizophrenic patients feature more words related to religion compared to that of healthy controls [26].

TABLE IV
WORD CATEGORIES SHOWING THE DIFFERENT OF PATIENTS AND HEALTHY CONTROLS.

Feature name	Dictionary	p-Value	Feature name	Dictionary	p-Value
assent	LIWC	0.0008	affiliation	LIWC	0.0261
unique	LIWC	0.0039	compare	LIWC	0.0356
work	LIWC	0.0051	ipron	LIWC	0.0376
Accomplish	Diction	0.0075	relig	LIWC	0.0426
tentat	LIWC	0.0080	Collectives	Diction	0.0477
informal	LIWC	0.0107	Familiarity	Diction	0.0492
netspeak	LIWC	0.0123	drives	LIWC	0.0546
we	LIWC	0.0161	prep	LIWC	0.0561
achieve	LIWC	0.0200	you	LIWC	0.0653
sixltr	LIWC	0.0253	nonflu	LIWC	0.0686

VI. CONCLUSIONS

In this paper, we employed speech recognition and NLP to automatically analyze the interview recordings of schizophrenic patients and healthy control subjects. We extended our earlier NLP approach based on LIWC [8] to Diction and Doc2Vec. The schizophrenia patients show significant language differences compared with healthy controls. These results are promising and present an important step towards our overall goal of creating automated systems to aid clinical diagnosis and understanding of schizophrenia. However, we recognize that the results are limited by the accuracy of the automatic speech recognition system, as it is harder to recognize Singapore English than general British or American English. Moreover, since our study only includes 50 patients, further research is required to confirm the results of this study.

ACKNOWLEDGMENT

This study was funded by the NMRC Center Grant awarded to IMH Singapore (NMRC/CG/004/2013), by the NITHM grant (M4081187.E30), and the RRIS Rehabilitation Research Grant (RRG2/16009). Being Together Centre also supported this research. The Being Together Centre is supported by the National Research Foundation, Prime Minister's Office, Singapore under its International Research Centres in Singapore Funding Initiative.

REFERENCES

[1] Demily, C. and Franck, N., 2008. Cognitive remediation: a promising tool for the treatment of schizophrenia. Expert Review of Neurotherapeutics, 8(7), pp.1029-1036.

[2] Covington, M.A., He, C., Brown, C., Nai, L., McClain, J.T., Fjordbak, B.S., Semple, J. and Brown, J., 2005. Schizophrenia and the structure of language: the linguist's view. Schizophrenia research, 77(1), pp.85-98.

[3] Beattie, G. and Ellis, A.W., 2017. The psychology of language and communication. Taylor & Francis.

[4] Jackson, R.G., Patel, R., Jayatilleke, N., Kolliakou, A., Ball, M., Gorrell, G., Roberts, A., Dobson, R.J. and Stewart, R., 2017. Natural language processing to extract symptoms of severe mental illness from clinical text: the Clinical Record Interactive Search Comprehensive Data Extraction (CRIS-CODE) project. BMJ open, 7(1), p.e012012.

[5] Miotto, R., Li, L., Kidd, B.A. and Dudley, J.T., 2016. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. Scientific reports, 6, p.26094.

[6] Thinking big in mental health. Nature Medicine. 2018 Jan 09; 24(1):1. Nature Publishing Group.

[7] Demner-Fushman, D., Chapman, W.W. and McDonald, C.J., 2009. What can natural language processing do for clinical decision support?. Journal of biomedical informatics, 42(5), pp.760-772.

[8] Shihao, S., Zixu, Y., Debsubhra, C.,Yasir, T., Tomasz, M., Victoria, Y., Dauwels, J., Thalmann, N., Thalmann, D., Tan, B.L. and Lee, J., 2018, 5th International Workshop on Spoken Dialog Systems (IWSDS).

[9] Pennebaker, J.W., Boyd, R.L., Jordan, K. and Blackburn, K., 2015. The development and psychometric properties of LIWC2015.

[10] Loughran, Tim, and Bill McDonald. "The use of word lists in textual analysis." Journal of Behavioral Finance 16.1 (2015): 1-11.

[11] Hirschberg, J. and Manning, C.D., 2015. Advances in natural language processing. Science, 349(6245), pp.261-266.

[12] Tausczik, Y.R. and Pennebaker, J.W., 2010. The psychological meaning of words: LIWC and computerized text analysis methods. Journal of language and social psychology, 29(1), pp.24-54.

[13] Hong, K., Nenkova, A., March, M.E., Parker, A.P., Verma, R. and Kohler, C.G., 2015. Lexical use in emotional autobiographical narratives of persons with schizophrenia and healthy controls. Psychiatry research, 225(1), pp.40-49.

[14] Minor, K.S., Bonfils, K.A., Luther, L., Firmin, R.L., Kukla, M., MacLain, V.R., Buck, B., Lysaker, P.H. and Salyers, M.P., 2015. Lexical analysis in schizophrenia: how emotion and social word use informs our understanding of clinical presentation. Journal of psychiatric research, 64, pp.74-78.

[15] Yuan, J., Holtz, C., Smith, T. and Luo, J., 2016. Autism Spectrum disorder detection from semi-structured and unstructured medical data. EURASIP Journal on Bioinformatics and Systems Biology, p.3.

[16] Magumba, M.A. and Nabende, P., 2016. Ontology Driven Disease Incidence Detection on Twitter. arXiv preprint arXiv:1611.06671.

[17] Le, Q. and Mikolov, T., 2014, January. Distributed representations of sentences and documents. In International Conference on Machine Learning (pp. 1188-1196).

[18] Tahir, Y., Chakraborty, D., Dauwels, J., Thalmann, N., Thalmann, D. and Lee, J., 2016, March. Non-verbal speech analysis of interviews with schizophrenic patients. In Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on (pp. 5810-5814). IEEE.

[19] Chakraborty, D., Tahir, Y., Yang, Z., Maszczyk, T., Dauwels, J., Thalmann, D., Thalmann, N.M., Tan, B.L. and Lee, J., 2017, October. Assessment and prediction of negative symptoms of schizophrenia from RGB+ D movement signals. In Multimedia Signal Processing (MMSP), 2017 IEEE 19th International Workshop on (pp. 1-6). IEEE.

[20] Andreasen, N.C., 1982. Negative symptoms in schizophrenia. Archives of general psychiatry, 39(784-788), p.564.

[21] Rehurek, R. and Sojka, P., 2010. Software framework for topic modelling with large corpora. In In Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks.

[22] Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP) (pp. 1532-1543).

[23] Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hannemann, M., Motlicek, P., Qian, Y., Schwarz, P. and Silovsky, J., 2011. The Kaldi speech recognition toolkit. In IEEE 2011 workshop on automatic speech recognition and understanding (No. EPFL-CONF-192584). IEEE Signal Processing Society.

[24] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. and Vanderplas, J., 2011. Scikit-learn: Machine learning in Python. Journal of machine learning research, 12(Oct), pp.2825-2830.

[25] Chawla, N.V., Bowyer, K.W., Hall, L.O. and Kegelmeyer, W.P., 2002. SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research, 16, pp.321-357.

[26] Kayi, E.S., Diab, M., Pauselli, L., Compton, M. and Coppersmith, G., 2017. Predictive Linguistic Features of Schizophrenia. In Proceedings of the 6th Joint Conference on Lexical and Computational Semantics (* SEM 2017) (pp. 241-250).