SCHIZOPHRENIA DETECTION FROM SOCIAL MEDIA

Mr N Sandeep Chaitanya, Computer Science & Engineering, VNRVJIET, Hyderabad, India,
sandeepchaitanya_n@vnrvjiet.in

Ch Sai Sri Harsha Computer Science & Engineering VNRVJIET, Hyderabad, India,
harshachunduri27@gmail.com

P Srikanth, Computer Science & Engineering, VNRVJIET, Hyderabad, India,
srikanthpanditi1729@gmail.com

T Prashanth, Computer Science & Engineering, VNRVJIET, Hyderabad,
*India*prashanth.art2001@gmail.com

L Nikhitha, Computer Science & Engineering, VNRVJIET, Hyderabad, India,
nikhitha.labishetti24@gmail.com

Abstract–
Schizophrenia is a serious intellectual disease that is one of the main reasons for disability in the world. The detection of this kind of a mental disorder is paramount for the well being of human beings; schizophrenia gets usually identified in later stages. Where it becomes a lot difficult for the person to get treated hence we need another way to identify schizophrenia beforehand. So that, it will not go out of control and can be treated smoothly one of the ways currently to detect schizophrenia is using mri scans of the brain which can only detect schizophrenia in later stages the world we live in is a digital world and almost all of the people including all age groups use the internet and use social media as a platform to express their opinions and feelings thus by using the social media profile of a person we would be able to detect schizophrenia this paper mainly deals with various existing methodologies and drawbacks related to the above problems like detecting schizophrenia in early stages and detection of schizophrenia through different types of social media posts.

Keywords: Deep Learning, MRI scans, NLP, Schizophrenia, social media, Diagnosis

INTRODUCTION

Schizophrenia is a serious mental disorder due to which the person gets both positive and negative symptoms. The positive symptoms are the ones that patients experience that healthy people don't experience. These include delusions, disorganized speech, and hallucinations. The negative symptoms are the ones that ordinary people don't experience. These include lack of motivation, emotional expression. Almost 10% of the people in this world are suffering with schizophrenia and the number will only grow in this modern eraSchizophrenia is such a mental disorder which if not diagnosed and treated the health of the person gets more and more deteriorated. Hence we need to identify it early which will not worsen the health. Early detection of schizophrenia is very challenging. One of the current ways to detect schizophrenia is through using MRI scans. Studies showed that schizophrenia alters the structure of the brain which will be detected through these MRI scans. But these can only be useful once the disorder has gotten into latter stages which can be harmful for the patient. The digital world we are living, has almost all of the people addicted to the internet. Almost all people irrespective of the age groups use social media. Use of social media is only ever growing. People tend to share their feelings and emotions online using applications like instagram, twitter, reddit, facebook etc. The contents of these posts can be used to identify schizophrenia in the people. The posts are of the format text, images, audio and videos. These data can provide information about their mental health which will help us in detecting schizophrenia. In our model we use Machine learning and deep learning models to detect schizophrenia.

RELATED WORKS

Jina Kim, Jieon Lee, Eunil Park & Jinyoung Han [1] In this paper they used XGBoost and CNN models for detecting mental illness. The CNN model used word2vec API in python. CNN has multiple layers including the embedding layer and finally sigmoid function for classification. CNN has a very high

accuracy of 86

Jihoon Oh, Baek-Lok Oh, and Kyongsik Yun [3]. MRI images were used as input. They used three dimensional CNN architecture for classification. The input was converted to a video of MRI images. They used the ReLU activation function in the neural network. usdetectschizophrenia. Image processing was done using python openCv library.Image processing was done using python openCv library.Image statistics such as width, height were extracted. HSV, RGB, Haar- Cascade Classifier were used to evaluate the properties used for detection of schizophrenia like saturation, colorfulness, aspect rKatrin Hansel, Inna Lin, Michael Sobolev, Whitney Muscat [4]. Instagram data of users isYi Ji Bae, Midan Shim , Won Hee Lee [5] for the dimensionality reduction they used lda for optimum repeated terms which returns a probability distribution of terms for the challenge of textual information in this tf-idf have become used to degree the maximum essential word it acts like a vectorizer. Binary classification was done with help of logistic regression Mitchell, Kristy Hollingshead, Glen Coppersmith [6] They used liwc which helped infinding a number of times a user writes i or me in a post and also used a algorithcalled brown brown clustering that finds similar words to make clusters and merged to form another cluster these end leaf clusters are the features values of the model they used two different models such as svm and max entropy model for this svm model accuracy is 823 for max entropy model accuracy is 8.19% this model works only for balanced dataset they used google api which takes enter as speech and converts them into indicators and the indicators can be transformed to textual content layout this enables in changing enter from speech to textual content. They used a 2 second pause to become aware of the length of a sentence Vinnarasu A, Deepa V Jose [7] nlp is to extract the features from the speech which has a few valuesmberly McManus, Rachel L. Haynes, Jonathan D. Tatum [8] They used the natural language toolkit nltk for python to carry out tokenization and lemmatization extracted textual features and used for producing the very last numeric feature vectors in this ann and support vector machines have been used they were given special accuracy precision recall and f1 rating for every model svm with pca version helped in identifying the users who're stricken by schizophrenia with an accuracy of 8.93 only for diagnosed individualsDafa Shi, Xiang Yao, [9] In this paper both the MRI and fMRI scans are used to detect schizophrenia . They used a new method called the M3 method to classify schizophrenia patients without using any Global Signal Regression. But we require large sample size data to detect schizophreniaElisa Veronese, Umberto Castellani, [10] It is used to categorize Schizophrenia and Health Control. The MRI datasets are analyzed deeply after which they used sample category techniques to extract the data from the datasets, So they used a few distinct approaches of SVM based total techniques to categorize the MRI records and nucleotide polymorphism records.Joel Weijia Lai , Candice Ke En Ang, [11] In this paper they used various AI and ML techniques to detect schizophrenia. They have taken both MRI scans and also PET scans for detecting it. Both linear and non-linear SVM methods were used to analyze the data. They used DFA to classify the data and the accuracy was 77%Jia Fu, Sen Yang, [12] In this paper they detected schizophrenia primarily based totally on speech instead of textual content via means of the use of CNN approaches. They used the CBAM module to categorize the data into schizophrenic and non-schizophrenic.The accuracy became round ninety one, but it did not find out the severity of the disease.Jina Kim, Jieon Lee , Eunil Park & Jinyoung Han [13] In this paper they used XGBoost and CNN models for detecting mental illness. The CNN model used word2vec API in python. CNN has multiple layers including the embedding layer and finally sigmoid function for classification. CNN has a very high accuracy of 86%.Tianlin Zhang, Annika M. Schoene, Shaoxiong Ji and Sophia Ananiadou[14]. Main objective of this paper was to compare how machine learning and deep learning models can be useful in natural language processing for detection of schizophrenia. Deep learning models were much more efficient and easier to use.Jihoon Oh, Baek-Lok Oh, and Kyongsik Yun [15]. MRI images were used as input. They used three dimensional CNN architecture for classification. The input was converted to a video of MRI images. They used ReLU activation function in the neural network.Katrin Hansel, Inna Lin, Michael Sobolev, Whitney Muscat [16]. Instagram data of users is used to detect schizophrenia. Image processing was done using python openCv library.Image processing was done using python openCv library.Image statistics such as width, height were extracted. HSV, RGB, Haar- Cascade Classifiers were used to evaluate the properties used for detection of

schizophrenia like saturation, colorfulness, aspect ratio.Wei Yen Chong,Bhawani Selvaretnam, Lay-Ki Soon [17] This paper used sentiment classification. In this they used subjectivity classification and semantic association. In the subjectivity classification method the system scans the tweets phrase via way of means of phrase and reveals out the phrase that incorporates sentiment if the phrase withinside the tweet includes fine or terrible sentiment weightage the tweet might be categorized as object or subject also additionally the neutral. They used SVM, NB, maximum entropy in classifying sentiment. Nisha Pawar, Zainab Shaikh, Poonam Shinde, Prof. Y.P. Warke [18] In this paper they used Optical Character Recognition (OCR) for image to text. The Tesseract OCR engine makes use of LSTM. LSTM is part of RNN. it is likewise appropriate at spotting large part of textual content statistics rather than single characters tesseract ocr engine substantially reduces mistakes created withinside the technique of person reputation tesseract assumes that the enter photo is a binary photo and processing takes region step-by-step the first step is to apprehend related components outlines are nested into blobs these blobs are prepared into textual content lines text lines are damaged in step with the pitching if there's a hard and fast pitch among the characters then reputation of textual content takes region thats a two-pass techniqueSu Myat Mon, Hla Myo Tun,[19] In this paper they detected schizophrenia based on the audio data, they applied HMM method to train and test the audio data. Then the data is pre-processed and the sample data is extracted to the feature vectors.It achieved an accuracy of 87.6%. But it didn't predict well for larger sample dataNeha Gaur, Dr. Neetu Sharma (HOD), [20] In this paper they used NLP techniques to capture the sentiments of the text. They used Naive Bayes classifier to classify the data according to the emotions it predicted. Pre-processing is done after converting the text into lowercase letters.

Table 1: Table showing different methodologies, pros, cons and the results obtained in this literature survey.

S.No	Title	Methodology	Pros/Cons	Year
1	A deep learning model for detecting mental illness from user content on social media.	They used XGBoost and CNN models for detecting mental illness. The CNN model used word2vec API in python. CNN has multiple layers including the embedding layer and finally sigmoid function for classification.	CNN has a very high accuracy of 86% It did not consider several factors including socio-demographic and regional differences.	2022
2	NLP applied to mental illness detection: a narrative review	Main objective of this paper was to compare how machine learning and deep learning models can be useful in natural language processing for detection of schizophrenia. Machine learning models requires feature engineering but deep learning models learnt features by themselves. Deep learning models were much more efficient and easier to use.	It showed how Deep learyis better than ML for NLP. Deep Learning had very high accuracy and was easier to train and use. The result was unstable for extremely long and extremely short samples of text.	2022

3	Identifying schizophrenia using structural MRI with a deep learning algorithm	MRI images were used as input. They used three dimensional CNN architecture for classification. The input was converted to a video of MRI images. They used ReLU activation function in the neural network.	This algorithm could classify MRI images with different scanning parameters Degraded performance for younger patients with shorter duration of illness.	2022
4	Utilizing Instagram Data to Identify Usage Patterns Associated With Schizophrenia Spectrum Disorders	Instagram data of users is used to detect schizophrenia. Image processing was done using python openCv library.Image processing was done using python openCv library.Image statistics such as width, height were extracted. HSV, RGB, Haar- Cascade Classifier were used to evaluate the properties used for detection of schizophrenia like saturation, colorfulness, aspect ratio.	Difference in images uploaded by the people with SSD and people without SSD was clearly identified Sample size was small so it limits the generalisability	2022
5	Schizophrenia Detection Using Machine Learning Approach from Social Media Content	For the dimensionality reduction they used lda for optimum repeated terms which returns a probability distribution of terms for the challenge of textual information in this tf-idf have become used to degree the maximum essential word it acts like a vectorizer. Binary classification was done with help of logistic regression	The accuracy is 89%. The model identified the most repeated words in each topic of schizophrenia and non-schizophrenia groups. This model is only used for analysis of reddit data which were posted in the reddit platform. we cannot give our own textual data as input to this, because it does not have web application	2021
6	Quantifying the Language of Schizophrenia in Social Media	They used liwc which helped in finding a number of times a user writes i or me in a post and also used a algorithm called brown brown clustering that finds similar words to make clusters and merged to form another cluster these end leaf clusters are the features values of the model they used two different models such as svm and max entropy model for this svm model accuracy is 823 for max entropy model accuracy is 8.19% this model works only for balanced dataset	For SVM Model Accuracy is 82.3% For Max entropy model Accuracy is 81.9%. This works for only balanced dataset It does not focus on more linguistics features of the post.	2015

7	Speech to text conversion and summarization for effective understanding and documentation	nlp is to extract the features from the speech which has a few values they used google api which takes enter as speech and converts them into indicators and the indicators can be transformed to textual content layout this enables in changing enter from speech to textual content. They used a 2 second pause to become aware of the length of a sentence	They used 2 seconds as a pause to identify it as a sentence. It has the ability to identify the questions It appends punctuations like ('?' , '.', and ',') to the textual data It takes more time in conversion of very big speech limited to the fashion domain.	2019
8	Mining Twitter Data to Improve Detection of Schizophrenia	They used the natural language toolkit nltk for python to carry out tokenization and lemmatization extracted textual features and used for producing the very last numeric feature vectors in this ann and support vector machines have been used they were given special accuracy precision recall and f1 rating for every model svm with pca version helped in identifying the users who're stricken by schizophrenia with an accuracy of 8.93 only for diagnosed individuals	SVM+PCA model helped in identifying the users who are suffering from schizophrenia with an accuracy of 89.3% It only identifies diagnosed individuals. It does not identify undiagnosed accounts.	2015
9	Machine Learning of Schizophrenia Detection with Structural and Functional Neuroimaging.	In this paper both the MRI and fMRI scans are used to detect schizophrenia . They used a new method called the M3 method (multimodal imaging and multilevel characterization with multiclassifier) to classify schizophrenia patients without using any Global Signal Regression.	M3 method was successfully detected the SZ patients with an accuracy of 83.49%. It can be used for early detection of SZ. larger sample size should be recruited in the future to replicate and enrich our findings.	2021
10	Machine Learning Approaches: From Theory to Application in Schizophrenia	It is used to classify both Schizophrenia and Health Control. The MRI images are analyzed deeply and then they used pattern classification methods to extract the information from the images, So they used some different types of SVM (Support vector machine) based methods to classify the MRI data and nucleotide polymorphism data.	It classified the healthy controls and patients affected by schizophrenia with an accuracy of 87%. SVM analysis requires feature vectors should be of same length	2013

11	Schizophrenia: A Survey of Artificial Intelligence Techniques Applied to Detection and Classification	In this paper they used various AI and ML techniques to detect schizophrenia. They have taken both MRI scans and also PET scans for detecting it. Both linear and non-linear SVM methods were used to analyze the data. They used DFA(discriminant function analysis) to classify the data and the accuracy was 77%.	They used other Neurological Scans like PET scans and EEG signals to classify the data. Accuracy is high when they used SVM methods. It didn't worked properly on small sample sizes.	2021
12	Sch-net: a deep learning architecture for automatic detection of schizophrenia	In this paper they detected schizophrenia based on speech rather than text by using CNN approaches. They used the CBAM module to categorize the information into schizophrenic and non-schizophrenic. The accuracy was around ninety one, however it didn't discover the severity of the disease.	It is a speech based approach to detecting schizophrenia patients. This model can only achieve the classification of patients and healthy controls, but not severity. The generalization of the proposed model should be further verified.	2021
13	Machine Learning in Detecting Schizophrenia: An Overview	They used SVM (suppport vector machine) to classify the data into schizophrenia and healthy control groups. It applied 10-fold cross validation to confirm the results. Different ML models are presented for detecting schizophrenia patients among other cases based on the significant regions of the brain using MRI data.	It is used different types of ML techniques to choose the best among them. It achieved accuracy of 70-90%. This model can work only for small data and early prediction is not possible. if it is combined with the big data then it will detect in early stages of the disorder.	2020
14	machine learning classification of schizophrenia patients and healthy controls using diverse neuroanatomical markers and Ensemble methods	They used the ensemble methods to classify schizophrenia and healthy control cohorts using a diverse set of neuroanatomical measures (cortical and subcortical volumes, cortical areas and thickness, cortical mean curvature). Also they correlated such neuroanatomical features with Quality Of Life(QOF).	It is MRI and Quality of life based approach of detecting schizophrenia patients. It achieved accuracy of 83%. First, test the cross sectional dataset, second assess the utility of this classification system within a longitudinal dataset. Time consuming	2022

15	Assessing Schizophrenia Patients Through Linguistic and Acoustic Features Using Deep Learning Techniques	They used BERT model to detect schizophrenia based on the textual and acoustic speech between occupational therapists or psychiatric nurses and schizophrenia patients to predict the level of their thought disorder. They firstly record the conversation between the patient and therapist, then divides into Acoustic features, semantic textual features and syntactic textual features to redirect. first record the conversation between the patient and therapist , then divide it	It has different features such as semantic, syntactic and acoustic features. works for conversations between schizophrenia patients and interviewers improved by detecting only based on schizophrenia patients	2022
16	An overview of artificial intelligence techniques for diagnosis of Schizophrenia based on magnetic resonance imaging modalities: Methods, challenges, and future works	Various artificial intelligence (AI) techniques have been employed with advanced image processing methods to accurately diagnose SZ Introduces the most important conventional machine learning (ML) and deep learning (DL) techniques in the diagnosis of diagnosing SZ Future works in diagnosing SZ using AI techniques and MRI modalities	A comprehensive comparison is also made between ML and DL The DL networks require a lot of data for training, and the lack of free and available datasets are the main reason for the main challenge in the automated diagnosis of SZ accurately.	2022
17	Natural Language Processing for Sentiment Analysis	This paper used sentiment classification. In this they used subjectivity classification and semantic association. In the subjectivity classification method the system scans the tweets phrase by way of means of phrase and reveals out the phrase that incorporates sentiment if the phrase with inside the tweet includes fine or terrible sentiment weightage the tweet might be categorized as object or subject also additionally the neutral. They used SVM, NB, maximum entropy in classifying sentiment.	This sentiment analysis gave high accuracy for twitter data It cannot perform well for complex text information	2014

18	Image to Text Conversion Using Tesseract	In this paper they used Optical Character Recognition (OCR) for image to text. The Tesseract OCR engine makes use of LSTM. LSTM is part of RNN. Tesseract ocr engine substantially reduces mistakes it assumes that the enter photo is a binary photo and processing takes region step-by-step the first step is to apprehend related components outlines are nested into blobs these are prepared as textual content lines text lines are damaged in step with the pitching if there's a hard and fast pitch among the characters then reputation of textual content takes region thats a two-pass technique.	It can extract text from the colored text images This system has an ability to detect and extract english and chinese text from images It does not extract text for complex captchas Blurred text is another problem in this system	2019
19	Speech-To-Text Conversion (STT) System Using Hidden Markov Model (HMM)	In this paper they detected schizophrenia based on the audio data, they applied HMM method to train and test the audio data. Then the data is pre-processed and the sample data is extracted to the feature vectors. It achieved an accuracy of 87.6%. But it didn't predict well for larger sample data.	It can convert low pitch speech to text. But it cannot identify the multiple voices in an audio clip.	2015
20	Sentiment Analysis in Natural Language Processing	In this paper they used NLP techniques to capture the sentiments of the text. They used Naive Bayes classifier to classify the data according to the emotions it predicted. Pre-processing is done after converting the text into lowercase letters.	It is the best approach to detect the polarity of the text.	2017

CONCLUSION

With the increasing ratio of schizophrenia patients it is becoming difficult to find who is suffering from schizophrenia. We performed a literature survey on the related works that address these challenges. From the above literature survey, we can see that schizophrenia detection is implemented using MRI scans which can be used after a patient is known that they are suffering from schizophrenia and there is no proper medication if it goes into dangerous stages. So it is necessary to find out schizophrenia in early stages. Our project aims to detect schizophrenia from social media contents. It detects schizophrenia from the text. It can also work on the audio where audio is converted into text then the model is applied on the text it can also detect schizophrenia from the video where we firstly retrieve audio from the video then audio is converted into text and model is applied on the text. It also works on MRI images.

REFERENCES

Chen, Z., Yan, T., Wang, E., Jiang, H., Tang, Y., Yu, X., Zhang, J. and Liu, C., 2020. Computational

intelligence and neuroscience, 2020.Detecting abnormal brain regions in schizophrenia using structural MRI via machine learning..

Birnbaum, M. L., Ernala, S. K., Rizvi, A. F., De Choudhury, M., & Kane, J. M. (2017).Journal of medical Internet research, 19(8), e7956. A collaborative approach to identifying social media markers of schizophrenia by employing machine learning and clinical appraisals.

Buettner, Ricardo, et al. 2019 IEEE International Conference on E-Health Networking, Application & Services (HealthCom).IEEE,2019."High-performance exclusion of schizophrenia using a novel machine learning method on EEG data."

Chu, L., Qiu, R., Liu, H., Ling, Z., Zhang, T. and Wang, J., 2017..Individual recognition in schizophrenia using deep learning methods with random forest and voting classifiers

de Filippis, Renato, et al. Neuropsychiatric disease and treatment 15 (2019): 1605. "Machine learning techniques in a structural and functional MRI diagnostic approach in schizophrenia: a systematic review."

Schnack, H.G., Van Haren, N.E., Nieuwenhuis, M., Hulshoff Pol, H.E., Cahn, W. and Kahn, R.S., 2016. Accelerated brain aging in schizophrenia a longitudinal pattern recognition study.

Corsico, Paolo. International Journal of Law and Psychiatry 66 (2019): 101479. "The risks of risk. Regulating the use of machine learning for psychosis prediction."

Nikhil Chandran, A., Karthik Sreekumar, and D. P. Subha. Advances in Machine Learning and Computational Intelligence. Springer, Singapore, 2021. 229-236. "EEG-based automated detection of schizophrenia using long short-term memory (LSTM) network."

Zhang, Lei. 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2019. "EEG signals classification using machine learning for the identification and diagnosis of schizophrenia."

Honnorat,Nicolas,"Neuroanatomical heterogeneity of schizophrenia revealed by semi-supervised machine learning methods."Schizophrenia research 214 (2019)

Veronese, E., Castellani, U., Peruzzo, D., Bellani, M., & Brambilla, P. (2013). from theory to application in schizophrenia. Computational and mathematical methods in medicine, 2013.

Filippis, Renato, et al . Neuropsychiatric disease and treatment 15 (2019): 1605 "Machine learning techniques in a structural and functional MRI diagnostic approach in schizophrenia: a systematic review."

Chand, Ganesh B., Dominic B. Dwyer, Guray Erus, Aristeidis Sotiras, Erdem Varol, Dhivya Srinivasan, Jimit Doshi et al. Brain 143, no. 3 (2020): 1027-1038. "Two distinct neuroanatomical subtypes of schizophrenia revealed using machine learning."

Schnack, Hugo G. "Improving individual predictions: machine learning approaches for detecting and attacking heterogeneity in schizophrenia (and other psychiatric diseases). Schizophrenia research 214 (2019): 34-42. "

Bae, Yi Ji, Midan Shim, and Won Hee Lee. Sensors 21.17 (2021): 5924"Schizophrenia Detection Using Machine Learning Approach from Social Media Content."

Tandon, N., & Tandon, R. (2019).Realizing the promise and avoiding the hype. Schizophrenia Research, 214, 70-75 Using machine learning to explain the heterogeneity of schizophrenia.

Buettner, Ricardo, et al. 2019 IEEE International Conference on E-Health Networking, Application & Services (HealthCom)."High-performance exclusion of schizophrenia using a novel machine learning method on EEG data."

Zhang, Lei. 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2019. "EEG signals classification using machine learning for the identification and diagnosis of schizophrenia."

Shim, Miseon, et al. Schizophrenia research (2016): 314-319. "Machine-learning-based diagnosis of schizophrenia using combined sensor-level and source-level EEG features."

Hettige, Nuwan C., et al. A machine learning approach." General hospital psychiatry 47 (2017): 20-28."Classification of suicide attempters in schizophrenia using sociocultural and clinical features.