Using Machine Learning Approaches to Predict a Relapse in Schizophrenia

Vikas Nibhanupudi

Word Count: 1462

John McCrae Secondary School

vikas.nibhanupudi@gmail.com

+1-613-797-2549

68 Keighley Circle, Kanata, ON, K2K 3H8

Abstract

Schizophrenia is a rare, chronic mental illness with symptoms including delusions and hallucinations. There is no cure for schizophrenia, so antipsychotic medications are used to alleviate symptoms despite having numerous side effects. People with schizophrenia often stop taking their medication, resulting in a severe relapse requiring hospitalization and high doses of antipsychotic medication. The aim of this research is to create a machine learning model that uses smartphone data to predict whether someone is at risk of relapse. The dataset used to train and test the models was obtained from the CrossCheck study. Three machine learning classification algorithms were implemented: Logistic Regression, K-Nearest-Neighbors, and Multilayer Perceptron Classifier. Three tests were performed for each model. In the first test, all input features were used to train the model. In the second test, all input features were used and hyperparameter optimization was performed. In the final test, Principle Component Analysis was performed to reduce the number of input features while still preserving trends in the data. In the first set of tests, Multilayer Perceptron Classifier achieved the highest accuracy of 60.37%. In the second set of tests, Logistic Regression achieved the highest accuracy of 64.57%. In the final set of tests, Multilayer Perceptron Classifier achieved the best overall accuracy of 71.16%. These findings suggest that smartphone data is a useful indicator in predicting schizophrenic relapse and initiating preemptive treatment. Additional research can be done on how to improve the accuracy further and implement this model in a clinical setting.

Keywords: Schizophrenia, machine learning, relapse

Introduction

Schizophrenia is a chronic mental illness that occurs in 1% of Canadians and is primarily characterized by delusions, hallucinations, and cognitive disfunction [1, 2]. Although it can occur at any age, schizophrenia is most prevalent in young adults between 18 and 25 years of age [3]. The life expectancy of those with schizophrenia is generally 15 to 25 years shorter than the average person [4]. At this time, there is no cure for schizophrenia, so antipsychotic medications are used to manage symptoms [5]. However, antipsychotic medications have severe side effects such as akathisia (restlessness), tardive dyskinesia (involuntary movements), and rapid weight gain [6]. Furthermore, individuals who are diagnosed late with schizophrenia often develop anosognosia; a condition where their own mind tricks them into believing they are normal even when they are ill [7]. People who develop anosognosia may refuse treatment or stop taking their medication, resulting in a severe relapse that will require hospitalization and high doses of antipsychotic medication [8].

It is imperative that schizophrenia is diagnosed accurately and early to minimize the use of antipsychotic medication and ensure a quick recovery. The aim of this research is to develop a machine learning model that can accurately predict whether someone is at risk of schizophrenic relapse using smartphone sensor data. This will allow people to seek medical attention in the initial stages of relapse. When schizophrenia is diagnosed in its early stages, the efficacy of treatment options such as cognitive behavioural therapy and antipsychotic medications is higher. The ultimate objective of this machine learning model is to initiate preemptive treatment for those at risk of relapse.

Methods

The dataset used to train the model was obtained from a study conducted by Cornell University titled *Crosscheck: Integrating self-report, behavioural sensing, and smartphone use to identify digital indicators of psychotic relapse* [9]. In this study, people with schizophrenia were monitored using a multimodal data collection software called CrossCheck to identify indicators of psychotic relapse. Smartphone data such as the number of incoming and outgoing calls, sleep duration, speech frequency, app usage, and physical activity were collected over a span of 12 months to identify trends associated with relapse [9]. Participants of the study were also required to complete Ecological Momentary Assessments (EMA) every other day about their mood [9]. These questionnaires were scored from -15 to 15, where lower scores indicated a higher risk of relapse [9].

For this research, three machine learning classification algorithms were implemented: K-Nearest-Neighbors, Logistic Regression, and Multilayer Perceptron Classifier (MLP Classifier). All machine learning was performed using Python version 3.6.9 on Google Colaboratory. The dataset was randomly split into training and testing data in a ratio of 75 to 25. These models used smartphone data as inputs to classify whether someone was at risk of relapse, as measured by their EMA score. In this research, an EMA score greater than 8 indicated that the person was not at risk of relapse (0) while a score less than or equal to 8 indicated that the person was at risk of relapse (1).

Three sets of tests were performed for each model. In the first set of tests, all 138 input features were used and hyperparameter optimization was not performed. In the second set of tests, all input features were used and hyperparameter optimization was performed through stratified tenfold cross validation with grid search. Once the optimal hyperparameters were found, each model was fit in the training dataset and tested in the testing dataset. In the final set of tests, Principle Component Analysis (PCA), a dimensionality reduction technique, was also implemented. When PCA is used, only features with a high variance are used to train the model. This reduces the likelihood of overfitting since features with a low variance are not used to train the model. The total explained variance ratio was plotted to determine the number of inputs needed for 95% of variance to be explained (Figure 1).

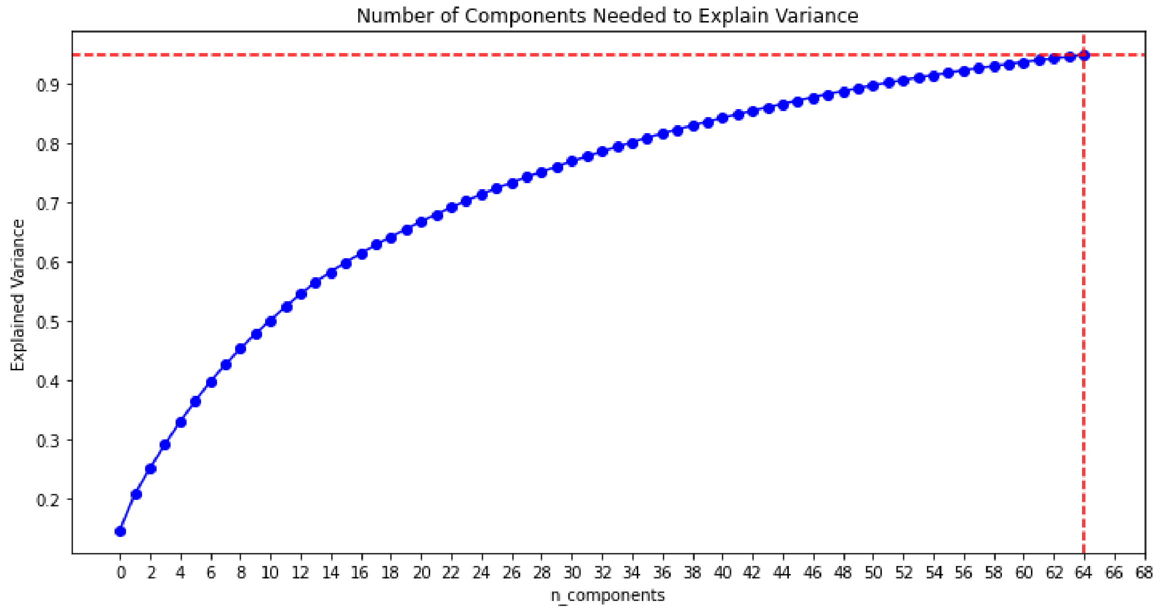

Figure 1: Number of Components Needed to Explain 95% of Variance

As shown in Figure 1, 64 inputs features were needed to explain 95% of variance. All of the models were trained using the 64 selected inputs and were tested on the testing data.

Results

In the first set of tests, with all input features and no hyperparameter optimization, MLP Classifier achieved the highest accuracy of 60.37% while Logistic Regression and K-Nearest-Neighbors achieved accuracies of 57.53% and 58.28%, respectively. Confusion matrices were plotted for each model to determine how many false diagnoses were false negatives (Figures 2-4). A false negative occurs when the predicted diagnosis is not at risk of relapse (0) while the actual diagnosis is at risk of relapse (1). Although MLP Classifier achieved the highest accuracy, it had 297 false negatives, which is the highest of all the models (Figure 2). Meanwhile, Logistic Regression had the lowest number of false negatives at 217 and K-Nearest-Neighbors had the median number of false negatives at 290 (Figures 3-4).

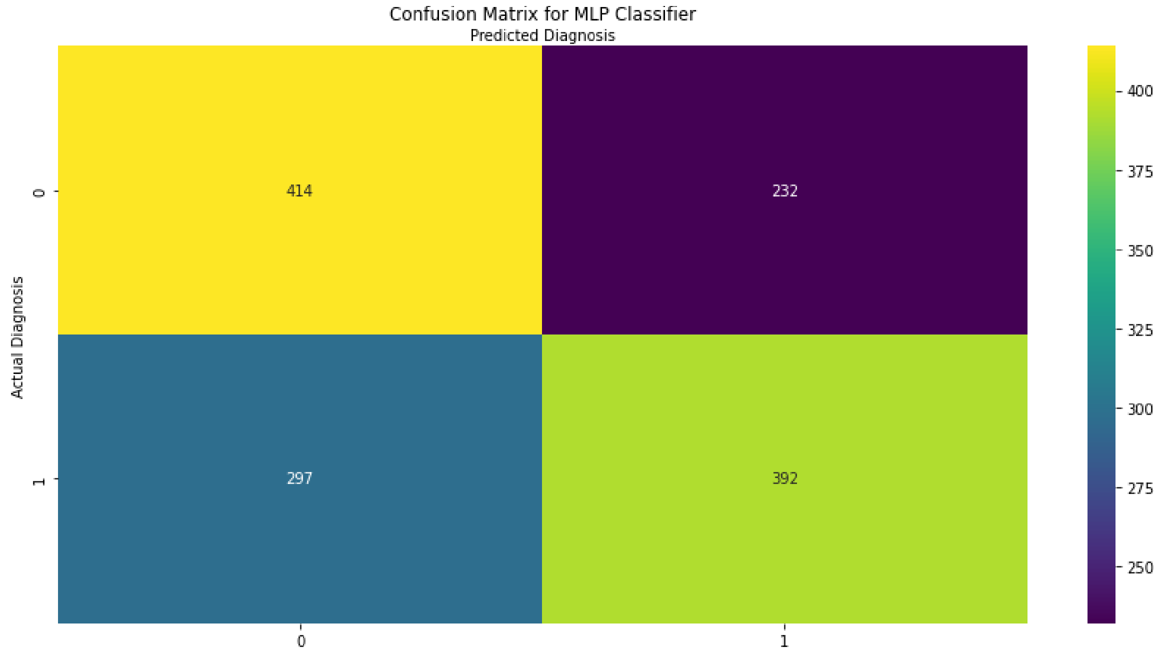

Figure 2: Confusion matrix for MLP Classifier without hyperparameter optimization

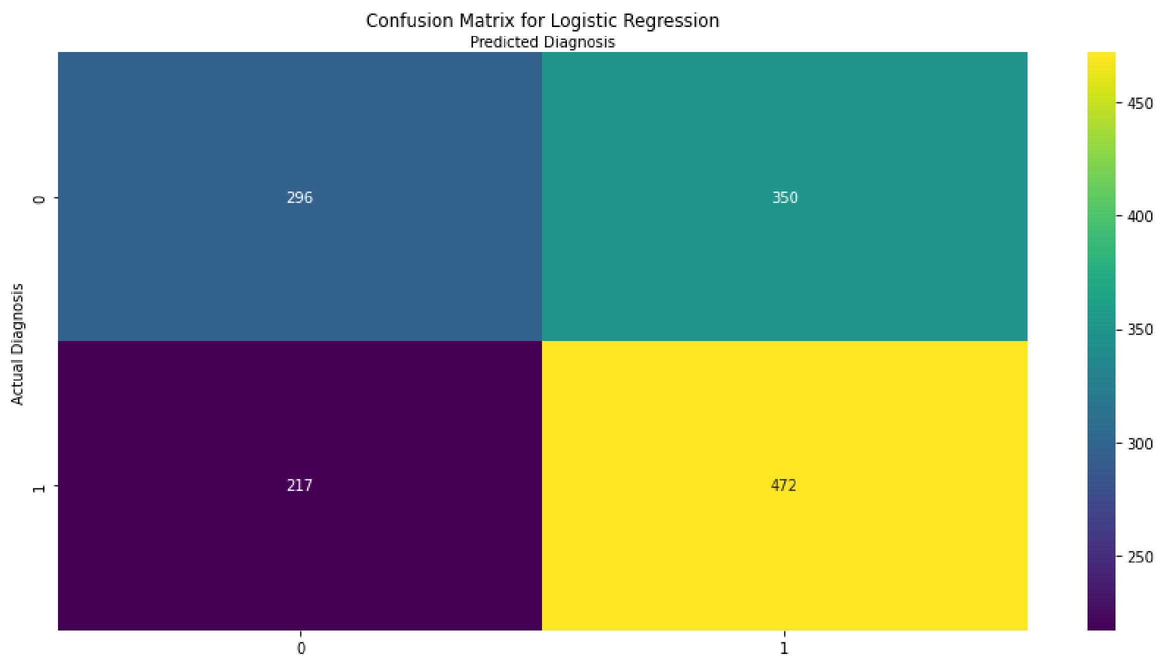

Figure 3: Confusion matrix for Logistic Regression without hyperparameter optimization

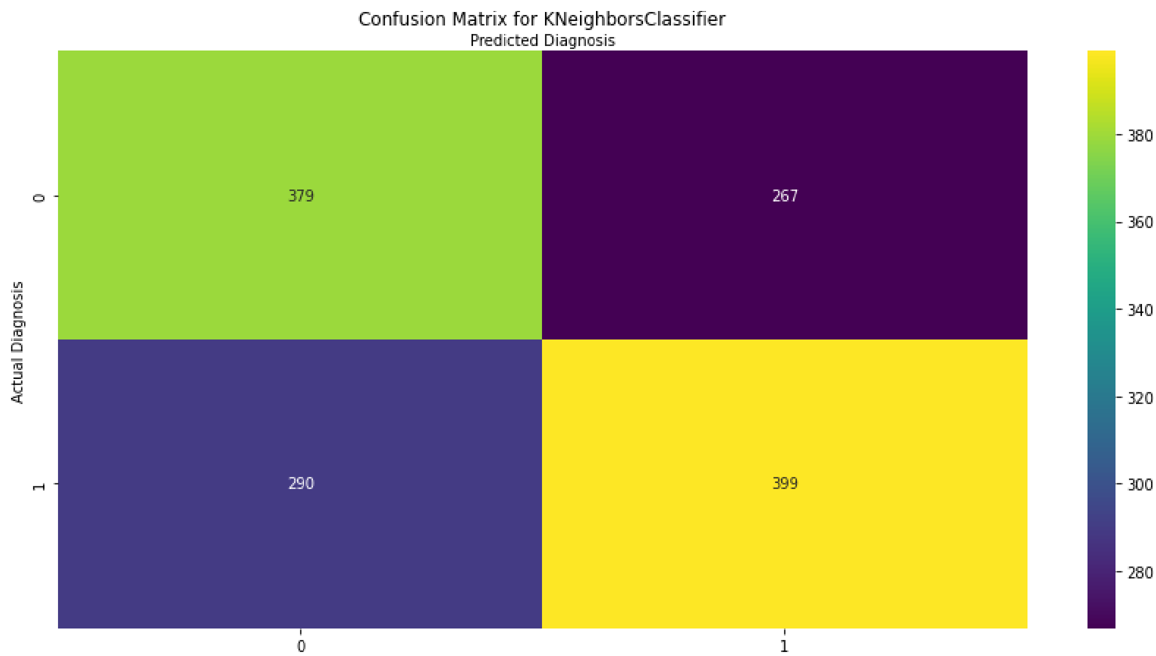

Figure 4: Confusion matrix for K-Nearest-Neighbors without hyperparameter optimization

In the second set of tests, once hyperparameter optimization was performed, the accuracy of each model grew (Table 1). Logistic Regression had the highest accuracy of 64.57% while MLP Classifier and K-Nearest-Neighbors had accuracies of 63.15% and 62.92%, respectively (Table 1).

Table 1: Comparison of the Accuracy of Each Machine Learning Model Before and After Hyperparameter Optimization

Accuracy of Each Machine Learning Model Before and After Hyperparameter Optimization			
Model	Accuracy Before Hyperparameter Optimization (%)	Accuracy After Hyperparameter Optimization (%)	Change (%)
MLP Classifier	60.37	63.15	+2.78
Logistic Regression	57.53	64.57	+7.04
K-Nearest-Neighbors	58.28	62.92	+4.64

In the final set of tests, once PCA was implemented, the accuracy of each model grew further and the number of false negatives decreased. MLP Classifier achieved the highest accuracy of 71.16%, while Logistic Regression and K-Nearest-Neighbors achieved accuracies of 64.87% and 69.36%, respectively. MLP Classifier had 194 false negatives, the lowest of all the models, while Logistic Regression and K-Nearest-Neighbors had 204 and 198 false negatives, respectively (Figures 5-7).

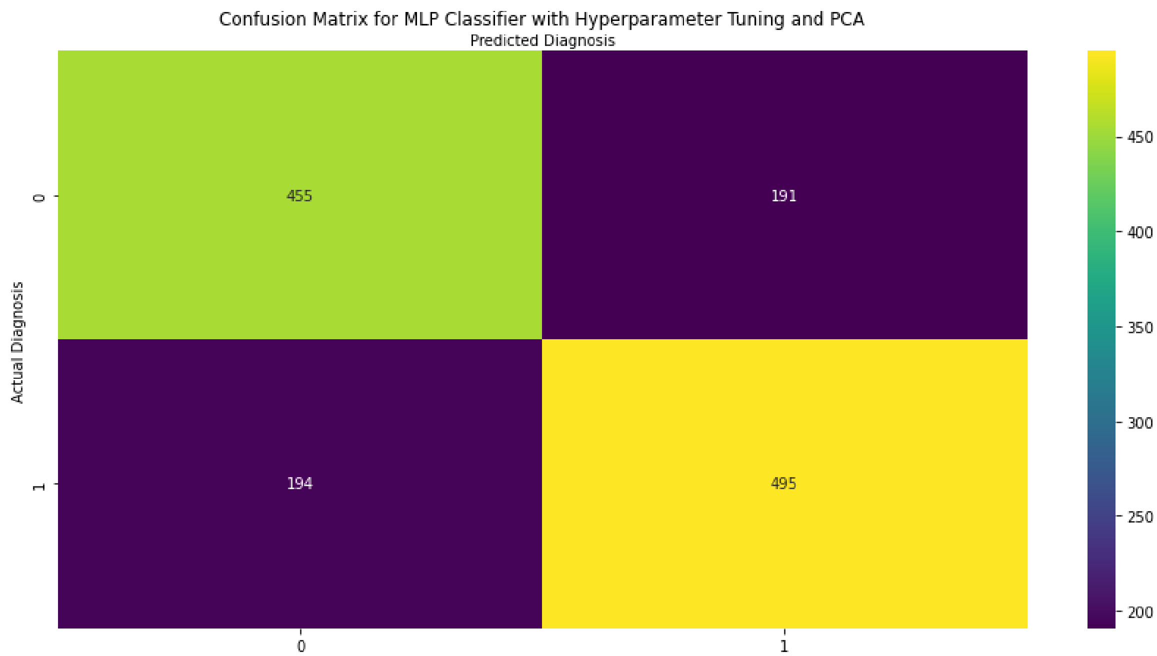

Figure 5: Confusion matrix for MLP Classifier with hyperparameter optimization and PCA

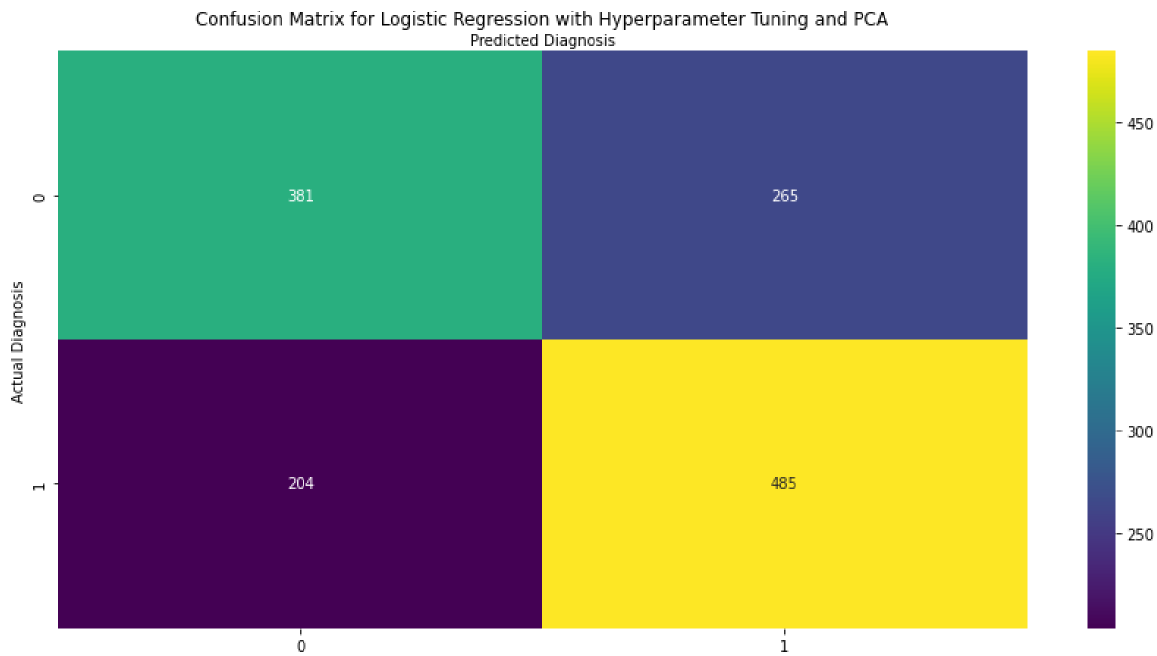

Figure 6: Confusion matrix for Logistic Regression with hyperparameter optimization and PCA

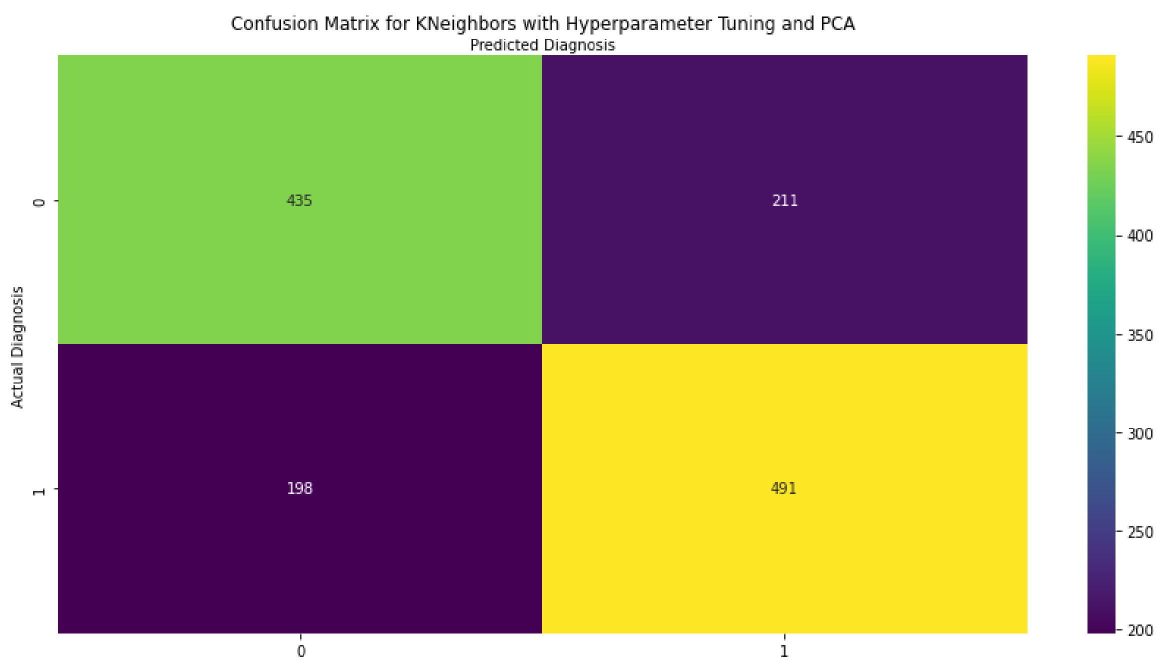

Figure 7: Confusion matrix for K-Nearest-Neighbors with hyperparameter optimization and PCA

Discussion

In this study, smartphone data indicators were used to predict if someone was at risk of schizophrenic relapse. After implementing hyperparameter optimization and PCA, the MLP Classifier model performed the best, achieving the highest accuracy of 71.16% and 194 false negatives, the lowest of all the models.

For this particular application of machine learning, the number of false negatives is an important metric to be considered when judging the model's performance. The objective of this model is to predict signs of schizophrenic relapse to initiate preemptive treatment. Therefore, the number of false negatives should be as low as possible to minimize the number of undetected cases of schizophrenic relapse. If the model predicts more false positives, it is not necessarily an issue because patients can consult their psychiatrist as a precautionary measure. The model is only a tool that helps with the early recognition of symptoms associated with relapse. Ultimately, the patient's psychiatrist makes the decision about whether to start treatment.

The primary limitation of this model is not accounting for the chronological progression of relapse indicators when making predictions. Since the data was collected over a span of 12 months, the model should take into account the progression of relapse indicators before making a prediction. Since the training and testing data was randomly split, the model was unable to use information about the progression of relapse indicators when making a prediction about whether the person was at risk of relapse.

This research has shown that a patient's smartphone data can be used to predict if they are at risk of relapse. Other studies have reported similar findings. According to a study done by Adler et al., participants' smartphone data indicated that there was a 108% increase in behavioural anomalies 30 days before a relapse [10]. Another study conducted by Henson et al. found that there was a 71% increase in behavioural anomalies 14 days before a relapse [11]. These findings suggest that smartphone data is a useful indicator in predicting schizophrenic relapse. Moving forward, machine learning algorithms with more complex architectures such as Artificial Neural Networks (ANN) and multidimensional Support Vector Machines (SVM) can be tested to see if they produce better results. Furthermore, more efficient hyperparameter optimization techniques such Bayesian optimization can be implemented to find a better a set of hyperparameters.

Conclusion

Smartphone data is a useful indicator for predicting schizophrenic relapse. The best machine learning model created in this research was 71.16% accurate in predicting schizophrenic relapse using smartphone data indicators. Future research is needed to examine how the accuracy of the model can be improved and how it can be integrated in clinical settings.

Reference List

1. Canadian Mental Health Association British Columbia Division [Internet]. Vancouver: CMHA; Schizophrenia; 2014 [cited 2021 Oct 9].
Available from: https://cmha.bc.ca/documents/schizophrenia-3/

2. Gogtay N, Vyas NS, Testa R, Wood SJ, Pantelis C. Age of Onset of Schizophrenia: Perspectives From Structural Neuroimaging Studies. Schizophr Bull Open [Internet]. 2011 May [cited 2021 Oct 9]; 37(3): 504–513.
Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3080674/
DOI: 10.1093/schbul/sbr030

3. Treatment Advocacy Center [Internet]. Arlington: Treatment Advocacy Center; Schizophrenia – Fact Sheet; 2008 Aug 13 [cited 2021 Oct 9];
Available from: https://www.treatmentadvocacycenter.org/evidence-and-research/learn-more-about/25-schizophrenia-fact-sheet

4. Wildgust HJ, Hodgson R, Beary M. The paradox of premature mortality in schizophrenia: new research questions. J Psychopharmacol [Internet]. 2010 Nov [cited 2021 Oct 9]; 24(4_supplement): 9–15.
Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2951588/
DOI: 10.1177/1359786810382149

5. Patel KR, Cherian J, Gohil K, Atkinson D. Schizophrenia: Overview and Treatment Options. P T [Internet]. 2014 Sep [cited 2021 Oct 9]; 39(9): 638–645. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4159061/

6. Muench J, Hamer AM. Adverse Effects of Antipsychotic Medications. Am Fam Physician [Internet]. 2010 Mar [cited 2021 Oct 9]; 81(5): 617–622.
Available from: https://www.aafp.org/afp/2010/0301/p617.html

7. National Alliance on Mental Illness [Internet]. Arlington: NAMI; Anosognosia; date unknown [cited 2021 Oct 9]; Available from: https://www.nami.org/About-Mental-Illness/Common-with-Mental-Illness/Anosognosia

8. Treatment Advocacy Center [Internet]. Arlington: Treatment Advocacy Center; Serious Mental Illness and Anosognosia. 2016 June [cited 2021 Oct 9]; Available from: https://www.treatmentadvocacycenter.org/key-issues/anosognosia/3628-serious-mental-illness-and-anosognosia

9. Ben-Zeev D, Brian R, Wang R, Wang W, Campbell AT, Aung MS, et al. CrossCheck: Integrating self-report, behavioural sensing, and smartphone use to identify digital indicators of psychotic relapse. Psychiatr Rehabil J [Internet]. 2017 Sep [cited 2021 Oct 9]. 40(3): 266–275.
Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5593755/
DOI: 10.1037/prj0000243

10. Adler DA, Ben-Zeev D, Tseng VW, Kane J, Brian R, Campbell A et al. Predicting Early Warning Signs of Psychotic Relapse From Passive Sensing Data: An Approach Using Encoder-Decoder Neural Networks. JMIR Mhealth Uhealth [Internet]. 2020 Aug [cited 2021 Oct 9]. 8(8): e19962.
Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7490673/
DOI: 10.2196/19962

11. Henson P, D'Mello R, Vaidyam A, Keshavan M, Torous J. Anomaly detection to predict relapse risk in schizophrenia. Transl Psychiatry [Internet]. 2021 Jan [cited 2021 Oct 9]. 11(28). Available from https://www.nature.com/articles/s41398-020-01123-7citeas
DOI: 10.1038/s41398-020-01123-7