Psychiatria Danubina, 2025 Vol. 37, No. 1, pp 46-54 https://doi.org/10.24869/psyd.2025.46 *Original research*

Evaluating machine learning algorithms for prediction of treatment response for sleep disturbances in patients with schizophrenia: A post-hoc analysis from a randomized controlled trial

Archana Mishra, Rituparna Maiti, Monalisa Jena & Anand Srinivasan

All India Institute of Medical Sciences (AIIMS), Bhubaneswar, Odisha, India

received: 28. 4. 2023; revised: 8. 5. 2024; accepted: 13. 11. 2024

Summary

Background: *A post-hoc analysis was planned to create and compare machine learning algorithms to predict treatment responses to sleep disturbances in patients with schizophrenia.*

Subjects and Methods: *This post-hoc analysis was done on a randomized controlled trial (NCT03075657), studying the effect of add-on ramelteon on sleep and circadian rhythm disturbances in 120 patients with schizophrenia. We created models using random forest, k-nearest neighbors, extreme gradient boosting machine, R part Classification and regression trees and logistic regression algorithms. R language with mlbench, caret, MASS, rPART packages were used. Box plot and dot plot were plotted to visualize comparisons among the models.*

Results: *The logistic regression algorithm was found to be the best-fit model with a specificity of 0.93 and sensitivity of 0.45, and ROC 0.78. Predominant symptom domain (positive or negative), urinary melatonin and global PSQI score at baseline were the most important variables when plotted in terms of mean decrease accuracy. These variables contributed significantly to the final model in the logistic regression algorithm, and the accuracy of this algorithm was found to be 90% for prediction.*

Conclusions: *Machine learning models are an emerging trend in clinical research and should be translated into clinical practice. The logistic regression model predicted responders with 90% accuracy.*

Keywords: *Schizophrenia; sleep disorder; machine learning*

* * * * *

INTRODUCTION

Sleep and circadian rhythm disturbances are frequent in patients with schizophrenia in addition to the cardinal features, which are used to diagnose the disease and ascertain its severity (Kaskie et al., 2017). Antipsychotic therapy helps alleviate the positive and negative symptoms, but they are ineffective in improving sleep disturbances. Patients report adverse consequences of these sleep disorders in the form of worsening symptoms that warrant treatment (Chemerinski et al., 2002). Conventional sedatives like benzodiazepines may impact phase shifts in circadian rhythm, thus setting up a vicious cycle in sleep disorders and psychosis symptoms (Hajak et al., 1996). Melatonin is a robust biological marker of circadian rhythmicity (Claustrat & Leston, 2015). Though there are pieces of evidence supporting the detection of schizophrenia with the help of various characteristics including structural MRI and multiple kernel learning identifiers; algorithms for prediction of response to interventions are lacking (Oh et al,. 2020; Chilla et al., 2022). In our previous study, augmentation therapy with melatonin receptor agonist ramelteon improved sleep and circadian rhythm in schizophrenia (Mishra et al., 2020). The data for this study was taken from the above-mentioned randomized controlled trial studying the effect of add-on ramelteon on sleep and circadian rhythm in patients with schizophrenia.

The primary objective of the clinical trial was to evaluate the change in serum melatonin (a robust biomarker of circadian rhythm), and the sample size was calculated accordingly. A regression model was built for the same in the original study. The present study aimed to generate multivariate prediction models for treatment response in terms of improvement in sleep using the Pittsburgh Sleep Quality Index (PSQI) scores in patients with schizophrenia. As a considerable percentage of patients do not respond to antipsychotic drugs and treatment optimization is recommended, personalized therapy to provide optimal augmentation is the need of the hour (Kane & Correll, 2010). Individualized therapy needs to consider the baseline differences in clinical, socio-demographic, comorbid conditions, substance abuse and patient expectations, i.e. selection of predictors for determining patient response before prescribing therapy. A combination of various

supervised learning algorithms of different classification models has been utilized to create the most accurate model, which could determine the most important and influential predictors of response catered to our dataset. Selecting a prescription medication for a particular patient or group with specific characteristics may be made possible by incorporating the patient-specific features in the model to predict the best response. The model thus created incorporating those predictor characteristics may enable deciding on a prescription drug for individual patients or subgroups.

Furthermore, we have tried to run simulations for a few clinical trial scenarios assessing response in terms of improvement in sleep with various sample sizes and calculating the power of the study. Clinical scenario evaluation provides a utilitarian framework that can evaluate competing designs and strategies in drug development. Successful submission, if effective, or early termination of the research, if ineffective, are the requirements of the modern drug development era in an efficient and timely manner. Especially in situations where patient recruitment and follow-up have been challenging for many investigators, interim analysis can be done, and simulations may be run for early decision-making on the fate of the trial. The present analysis was done to generate multivariate prediction models for treatment response in patients with schizophrenia.

SUBJECTS AND METHODS

Population

A total of 120 patients with schizophrenia of 18-65 years of age, irrespective of gender, diagnosed as per DSM-V criteria were enrolled as per the study except for patients with schizoaffective disorder or schizophrenia with somatoform disorders, highly agitated patients who needed immediate treatment, patients who were already under treatment, who were substance abuse or history of organicity, patients with a known history of dementia, obstructive sleep apnea syndrome, diabetes mellitus, pregnant and nursing women, and patients with a history of allergy or hypersensitivity to ramelteon were excluded from the study. The study was approved by the Institutional Ethics Committee, AIIMS, Bhubaneswar and the research was completed in accordance with the Helsinki Declaration and the ethical guidelines for biomedical research on human subjects (2017) from the Indian Council of Medical Research (ICMR).

Predictors and outcome

Data collected on the group defined as predominant symptom domain (positive or negative) as per Positive and negative symptom score (PANSS) scale treatment (ramelteon or control), age, gender, height, weight, BMI, smoking history, years of smoking, PANSS scale, urinary melatonin, serum melatonin, global Pittsburgh Sleep Quality Index (PSQI) score at baseline and serum N-acetyltransferase (AANAT) were considered possible predictors (Vrbova, 2018). The outcome or dependent variable was the treatment response, defined as ≥20% improvement in PSQI score. PSQI scores were measured at baseline, and the end of the study, and as questions in this instrument are specific to the past month and the proposed duration of enrolment in our study was one month. The outcome was dichotomized into response and nonresponse to ascertain a clinically significant benefit of the study drugs. A summary of the baseline characteristics of the patients recruited has been presented in Table 1.

Preparation of dataset

Individual cases were presented in rows, and variables were presented in columns. Dataset for classification was prepared by removing all the cases with missing values from predictors and responses. The dataset used for this analysis had approximately 18% missing data, and a generally accepted method of dealing with missing data containing more than 10% of information is deleting all records with missing data without significant loss of statistical power in modelling results (Cismondi et al., 2013). The data was run in R language. The structure and summary of the dataset were analyzed, and the class of the items changed to fit for further analysis. A split of 70:30 was used for the training and test dataset.

Supervised learning for determining predictors

Supervised learning targets an outcome of interest by an investigator, which was the response as defined by improvement in PSQI score. In this study, our aim was to build a model fit for differentiating treatment outcomes from each other based on any combination of predictor variables from the above-mentioned list. Training models were prepared using the most common linear and non-linear algorithms viz. extreme gradient boosting machine (XGBOOST), R part Classification and regression trees (rPART), k-nearest neighbors (KNN), Logistic

Table 1. Baseline characteristics across positive and negative symptom groups

Parameters	Predominant positive symptom group		Predominant negative symptom group	
	Control Group	Test Group	Control Group	Test Group
Number recruited	30	30	30	30
Age(years)	34·0 ± 8·38	38·6 ± 10·68	37·87 ± 13·84	34·97 ± 12·35
Sex (M: F)*	21:9	15:15	14:16	21:9
BMI (kg/m^2)	21·01 ± 5·55	23·29 ± 4·13	22·87 ± 5·05	23·52 ± 4·35
Smoking Status (Non-smoker: Ex-smoker:Smoker)*	18:0:12	23:0:7	22:1:7	22:0:8
Serum melatonin at 2 p·m· (pg/ml)	17·05 ± 3·29	17·49 ± 3·84	16·50 ± 3·70	17·49 ± 3·84
Serum melatonin at 2 a·m· (pg/ml)	51·99± 19·38	55·63± 18·69	55·20 ± 20·80	48·58 ± 16·57
Urinary 6aMTs (pg/ml)	0·88 ± 0·33	0·95 ± 0·32	0·83 ± 0·31	0·78 ± 0·25
Serum AANAT (ng/ml)	19·88 ± 6·40	21·38 ± 6·07	19·17 ± 7·84	16·73 ± 7·16
Total PANSS	91·90± 25·54	98·37± 27·15	86·1 ± 23·98	85·7 ± 22·12
Global PSQI	11·50 ± 3·67	10·97 ± 3·16	11·83 ± 4·52	12·03 ± 3·23

BMI: Body mass index; 6aMTs: 6-sulphatoxymelatonin; AANAT: arylalkylamine N-acetyltransferase; PANSS: Positive and negative syndrome scale; PSQI: Pittsburgh sleep quality index
All the values are represented in mean ± SD. *Values represented in numbers

Regression and Random Forest (RF) and summarized sensitivity, specificity and precision – value for all models. Package 'mlbench', 'caret', 'MASS' and 'randomForest' was used in R language was used for creating and tuning the models For creating and tuning the models, package 'mlbench', 'caret', 'MASS' and 'randomForest' was used in R language. (Leisch & Dimitriadou, 2021; Kuhn, 2008; Venables & Ripley, 2002; Liaw & Wiener, 2002). A common standard configuration for comparing models was used, which included repeated cross-validation with ten folds and three repeats with the use of the *caret* package. trainControl configuration was used for training models. The results of each model fit were compared with the help of calling resamples on the list of models, box and whisker and dot plots were made, and the significance of the difference between different algorithms was calculated. For the random forest model, each tree was trained by approximately two-thirds of the data, and a random selection of cases was made from the original data with replacement. Two thousand iterations were done, and 5000 trees were grown for each iteration with mtry rule of mtry = √predictors. Model performance in dependency of the number of variables was plotted. Out of all predictor variables, some predictor variables were selected at random, and the best split out of these was used to split the node. Out of the bag (OOB), the error rate was calculated for each tree, and the number of trees and predictors in each tree were optimized to maintain the OOB error rate to a minimum. Variables were plotted in order of importance in terms of mean decrease accuracy.

Variable selection was made using '*varSelRF'* package, which does backward elimination of variables based on the importance of the predictors. The entire process was repeated 50 times with random starting seeds, and the variables that got selected in >50% of runs were chosen for validation in further simulations of our trials.

The accuracy, kappa, sensitivity and specificity of each model were tabulated, and the area under the curve was determined for the testing set of data. Model performance was evaluated by plotting the ROC curve after creating an object for prediction using the prediction function.

Clinical scenario evaluation by model creation and data simulation

'Mediana' package in R language was used to evaluate the estimated power of possible clinical scenarios with different sample sizes (Paux & Dmitrienko, 2019). Data models, analysis models and evaluation models were created. Three case scenarios were simulated, first where treatment was inferior to placebo, second in which there was no difference between them (standard), and third, where treatment was superior to placebo as observed in our original trial was defined (optimistic). Simulation run on different sample sizes was tried (5 to 100). Sim. parameters object was defined with 1000 n.sims, and results were written in a doc file, and chances of success were noted down. A two-sided alpha level of 0.5 was

considered, and the marginal power method was used in the Criterion object to create evaluation models.

RESULTS

After four weeks of therapy with ramelteon, 78.35% had improvement in PSQI scores and hence achieved a response. 93.87% of participants receiving add-on ramelteon and 62.5% of the control group showed a response. A total of 98 cases were included after removing rows of data containing missing data either for columns of variables or responses.

Clinical Scenario Evaluation

For response, when defined as improvement in PSQI in an optimistic scenario (test drug is better than the comparator i.e. scenario in our study), a sample size of 10 would provide a power of 58.21%, while 15 shows a power of 79% (underpowered). When analyzing data with a sample size of 20, a power of 86.70% and a power of 90% and above could be achieved with a sample size of 30 and more. Whereas, in the inferior (test drug is inferior to comparator) and standard (test drug equivalent to comparator) scenario, the sample sizes of less than 50 per group achieved power as high as 8-10%. The trend shows that in these two scenarios, large sample sizes (probably

Table 2. Optimal ROC, sensitivity and specificity in training for each algorithm

Algorithm	ROC	Sensitivity	Specificity
Random Forest	0.7133333	0.25	0.9833333
XGB	0.7116667	0.05	1
KNN	0.7358333	0	1
Logistic Regression	0.7866667	0.45	0.9266667
rPART	0.6666667	0.35	0.9833333

ROC: Receiver operating characteristics; XGB: extreme gradient boosting; KNN: k-nearest number; rPART: Recursive Partitioning and Regression Trees

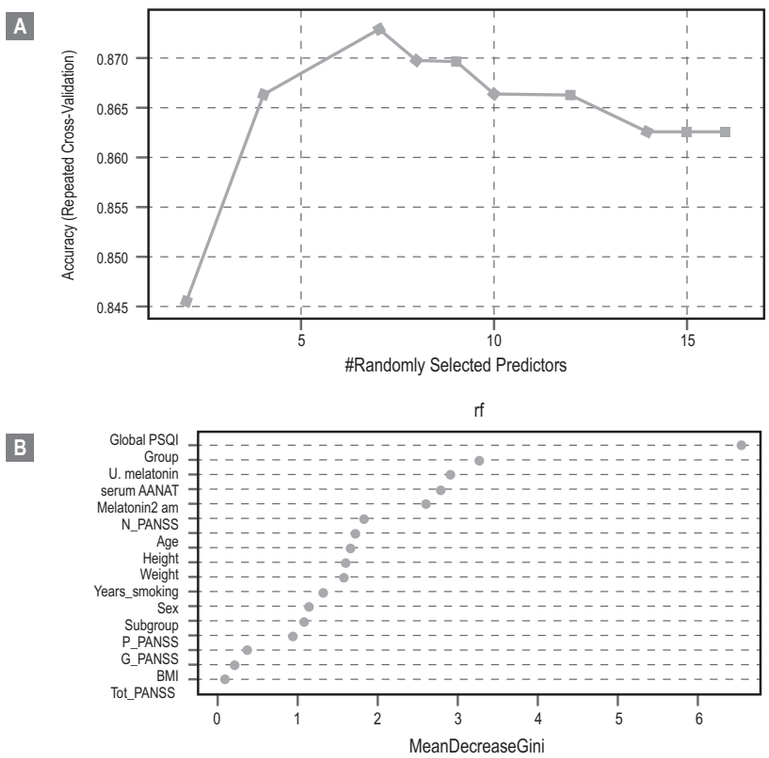

Figure 1. (A) Box plot (B) Dot plot on resamples with 'ROC' as metric (C) ROC curve drawn on test datasets for each machine learning algorithm.

Table 3. Metrics of various machine learning algorithms in the testing dataset

	Sensitivity	Specificity	Precision	Recall	F1	AUC
Random Forest	1.0000000	0.0000000	0.7931034	1.0000000	0.8846154	0.8478261
XGB	1.0000000	0.1666667	0.8214286	1.0000000	0.9019608	0.8115942
KNN	1.0000000	0.0000000	0.7931034	1.0000000	0.8846154	0.7463768
Logistic Regression	0.9565217	0.6666667	0.9166667	0.9565217	0.9361702	0.8623188
rPART	1.0000000	0.3333333	0.8518519	1.0000000	0.9200000	0.6666667

F1 = weighted average of precision and recall; AUC = Area under the curve; XGB: extreme gradient boosting; KNN: k-nearest number; rPART: Recursive Partitioning and Regression Trees

Table 4. Significance of differences between the metric distribution (ROC) of machine learning algorithms. P value adjustments were done using Bonferroni correction. Lower diagonal: p-value for H0 (smaller is better); Upper diagonal: estimates of the difference

	Random Forest	XGB	KNN	Logistic Regression	rPART_DT
Random Forest		0.033333	0.009167	-0.041667	0.078333
XGB	1		-0.024167	-0.075000	0.045000
KNN	1	1		-0.050833	0.069167
Logistic Regression	1	1	1		0.120000
rPART	1	1	1	1	

XGB: extreme gradient boosting; KNN: k-nearest number; rPART: Recursive Partitioning and Regression Trees

in multiples of thousands) would have been required to achieve statistical significance for the comparison between the test drug and the comparator successful submission. When simulating an improvement of 16% more than the control group, a sample size of 100 could have achieved marginal power of 82.6%.

Training and validating models using linear and non-linear algorithms

Models were compared based on the ROC metric, as depicted in Table 2. Box plots and dot plots were made to visualize the comparison between models. (Fig 1(A,B)). For the RF algorithm, the largest value of RoC was used to select the optimal model with a minimum node size of 1. In the XGBOOST algorithm, the tuning parameter was held at a constant value of 1, and the final optimal model used was that with a value of 0.71 for RoC. In the KNN algorithm, k with the highest RoC was determined to be ten, and in the rPART algorithm, the model with a complexity parameter of 0.053 was the optimal model with the highest RoC. The logistic regression model was seen to have the highest RoC of all the five models run in this study. The model with a logistic regression algorithm had a RoC of 0.786.

The significance of the difference between the metric distributions of different machine learning algorithms was calculated and has been depicted in Table 3. Adjustment of p-value was made using Bonferroni correction as this is the simplest and most common approach giving maximal power. When compared with each other, the algorithms show no significant difference (Table 4).

In the random forest model, for response prediction, accuracy increased with the number of variables showing a peak at seven variables and plateaus and declines thereafter (Fig 2(A)). In resampling, accuracy and kappa were 0.872 and 0.484, respectively and the tuning parameter mtry for the above values was seven for an OOB error rate of 0.133. Variables were plotted as per their importance, and Global PSQI score at baseline, urinary melatonin metabolite, serum AANAT enzyme and serum melatonin levels at 0200 hours were found to be most important (Fig 2 (B)).

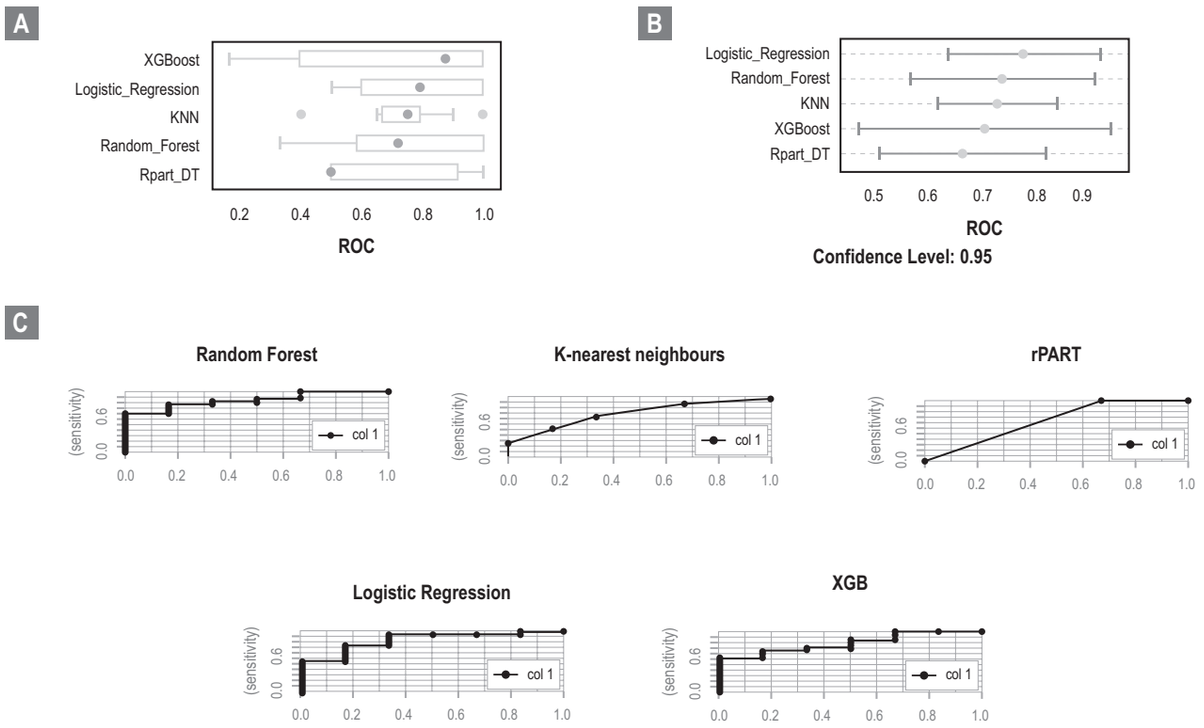

Figure 2. (A) Accuracy versus randomly selected predictor set (Random Forest Model) (B) Variables plotted in order of importance (decreasing gini)

Table 5. Coefficients and significance of their contribution in the regression model

	Estimate	Std Error	Z value	Pr (>|z|)
Intercept	-26.50279	36.52351	-0.726	0.46806
Group	3.20493	1.46672	2.185	0.02888
Subgroup	3.26517	4.77426	0.684	0.49403
Age	0.02743	0.05244	0.523	0.60098
Sex	2.39213	1.67366	1.429	0.15292
Weight	-0.06917	0.26764	-0.258	0.79606
Height	0.98875	22.24188	0.044	0.96454
Years of smoking	0.04275	0.19137	0.223	0.82323
MT_A	0.09683	0.12105	0.800	0.42374
MT_N	0.01127	0.56995	0.020	0.98422
AANAT	-0.14271	0.13515	-1.056	0.29102
U_MT	6.71441	34.90497	0.192	0.04874
P_PAN	0.13765	0.16821	0.818	0.41317
N_PAN	-0.19454	0.14984	-1.298	0.19417
G_PAN	0.13030	0.11043	1.180	0.23804
Gl_PSQI	1.16260	0.41644	2.792	0.00524

MT_A: serum melatonin at 2 pm in afternoon; MT_N: serum melatonin at 2 am in night; AANAT: arylalkylamine N-acetyltransferase; U_MT: Urinary melatonin in first void early morning sample; P_PAN: positive domain score in PANSS; N_PAN: negative domain score in PANSS; G_PAN: general domain score in PANSS scale; GI_PSQI: Global impression in PSQI scale; PANSS: Positive and negative syndrome scale; PSQI: Pittsburgh sleep quality index

Predictions on test datasets

The sensitivity, specificity, precision, F1 (weighted average of precision and recall), recall and AUC for all the models were calculated and have been represented in Table 3. The table clearly depicts that the logistic regression model best predicts the response in the given dataset. Four-fold plots were built to confirm and compare the accuracy of the models, and ROC curves were plotted (Fig 1(C)). A logistic regression model with an overall accuracy of 90% is reconfirmed to be the best model with four-dot plots. The accuracy of RF model was 0.793 and kappa of zero, for XGB an accuracy of 0.827 and kappa 0.24, for KNN accuracy was 0.793 and kappa of zero, for the logistic regression accuracy 0.896 and kappa 0.66 and for rPART, the accuracy was 0.862 and kappa 0.44 while testing our prediction models on the test dataset.

Predictions with logistic regression algorithm

Beta coefficients of the variables and the significance of their contribution to the model fit have been depicted in Table 5. The group (positive or negative symptom domain), global PSQI score at baseline and urinary melatonin metabolite were the predictors with significance values <0.05. The AIC value for the model built with these three variables was 57.3, and residual deviance (49.3) was lower than null deviance (99.3). The final model was created with coefficients (standard error) for the symptoms domain, global PSQI and Urinary melatonin being 2.59 (1.46), 0.82(0.46) and 3.13 (10.34), respectively. The intercept value was -10.6557 (35.78). This model can predict responders with 90% accuracy.

DISCUSSION

The success of clinical trial programs is based on statistical planning and execution at present times. But these analyses are based on parts of a trial program in isolation. These may or may not include considerations on elements of sensitivity analysis, robustness and efficiency of a drug development program. Modern drug development programs should focus on efficient modelling techniques for clinical trial optimization. The clinical scenario evaluation framework evaluates a range of assumptions that could be either standard (or real) or optimistic, or pessimistic. While evaluating clinical case scenarios for this study, we reverse-sequenced the steps involved in the analysis of the endpoint of response in PSQI, such that information on robustness and sensitivity for assessment of the efficacy of the drug could be made to ensure integrated and efficient conduct of the trial and satisfaction status for the information received. Though PSQI evaluations were amongst secondary objectives, simulations run for clinical scenario evaluation showed that the sample size used in our study was adequate according to the marginal power estimation criterion.

We created machine learning models with the most popular algorithms predicting treatment response in terms of PSQI score and compared the models with respect to ROC, sensitivity and specificity. Logistic regression followed by the Random Forest algorithm was the algorithm that had the highest ROC values along with optimal specificity and moderate sensitivity. Global PSQI score at baseline, symptom group and urinary melatonin levels were the most important variables as per the decision tree analysis in a random forest, and these variables contributed significantly to the logistic regression model. The findings on the training dataset were reconfirmed with accuracy and kappa statistics of the models on test data (Table 2 & 3). The Kappa value reveals the agreement of the outcome of classification when compared with the model created. Kappa statistic for the logistic regression algorithm was moderate to strong (kappa = 0.66) and can account for the reliability of 64-81% of data (McHugh 2012). Whereas random forest and KNN models showed null agreement (kappa = 0) with the outcome of classification. Additionally, the logistic regression model is simplistic to interpret as there are fewer assumptions while creating the model. And, for the present data set, the accuracy of the prediction of responders is high with the logistic regression model.

Schizophrenia is a multifactorial and complex disease and machine learning model for diagnosis and optimizing and informing treatment decisions amongst the available alternatives (Tai et al., 2019). Identifying the predictors of response and relapse in patients may reduce the emotional stress and health care costs and improve the quality of life. Patients with schizophrenia tend to be the highest healthcare utilizers due to the chronic and severe nature of the disease.(Kadakia et al,. 2022) This burden of morbidity on patients, caregivers and healthcare system could be saved by predicting response early in the disease course and individualizing therapy based on response predictor-based treatment algorithms. So far, synthesizing machine learning algorithms and translating them to the prediction of response for therapeutics is lagging behind, though it has been evidenced that the classification of schizophrenia patients is a possibility using machine learning methods using network properties

(Jo et al., 2020). Machine learning models have helped to identify schizophrenia based on neuroanatomical markers or more interestingly linguistic characteristics of social media content (Bae et al., 2021; Santos Febles, 2022).

The models which we built in this study had limitations of small sample size and a limited number of variables. Further models can be done using large databases with extensive and holistic documentation of patient characteristics. Concerns about the reproducibility of machine learning approaches in computational psychiatry are not naïve. Thus, prospective studies accounting for all possible covariates affecting treatment path and response may be done using machine learning algorithms in decision-making at all levels.

CONCLUSION

In conclusion, machine learning models may be used to predict responses in cases of complex diseases like schizophrenia. The fewer assumptions made in creating and training the model, the better the accuracy of its application. A logistic regression model may help in prediction with high accuracy in the present dataset for the administration of ramelteon in sleep and circadian rhythm disturbances in patients with schizophrenia. However, data from large naturalistic studies may help validate the findings of this study in better prediction of response and implication in clinical practice.

Funding source: None
Conflict of interest: None to declare
Data availability statement: Data supporting the findings of this study are available from the corresponding author on request.
Authors' contribution: AM: Concept, study design, statistical analysis, manuscript writing, final approval of the manuscript. RM: Study design, literature search, interpretation of the analysis, critical review of the manuscript, final approval of the manuscript. MJ: Study design, literature search, statistical analysis, manuscript writing, final approval of the manuscript. AS: Concept, study design, statistical analysis, critical review of the manuscript,final approval of the manuscript.

References

Bae, Y. J., Shim, M., & Lee, W. H. (2021). Schizophrenia detection using machine learning approach from social media content. *Sensors (Basel), 21*(17), 5924. https://doi.org/10.3390/s21175924

Chemerinski, E., Ho, B. C., Flaum, M., Arndt, S., Flemenbaum, A., & Andreasen, N. C. (2002). Insomnia as a predictor for symptom worsening following antipsychotic withdrawal in schizophrenia. *Comprehensive Psychiatry, 43*(3), 393-396.

Chilla, G. S., Yeow, L. Y., Chew, Q. H., Sim, K., & Prakash, K. N. B. (2022). Machine learning classification of schizophrenia patients and healthy controls using diverse neuroanatomical markers and ensemble methods. *Scientific Reports, 12*, 2755. https://doi.org/10.1038/s41598-022-06651-4

Cismondi, F., Fialho, A. S., Vieira, S. M., Reti, S. R., Sousa, J. M., & Finkelstein, S. (2013). Missing data in medical databases: Impute, delete or classify? *Artificial Intelligence in Medicine, 58*(1), 63-72.

Claustrat, B., & Leston, J. (2015). Physiological effects in humans. *Neurochirurgie, 61*(1), 77-84.

Hajak, G., Rodenbeck, A., Bandelow, B., Friedrichs, S., Huether, G., & Ruther, E. (1996). Nocturnal plasma melatonin levels after flunitrazepam administration in healthy subjects. *European Neuropsychopharmacology, 6*(2), 149-153.

Jo, Y. T., Joo, S. W., Shon, S. H., Kim, H., Kim, Y., & Lee, J. (2020). Diagnosing schizophrenia with network analysis and a machine learning method. *International Journal of Methods in Psychiatric Research, 29*(1), e1818. https://doi.org/10.1002/mpr.1818

Kadakia, A., Catillon, M., Fan, Q., Williams, G. R., Marden, J. R., Anderson, A., Kirson, N., & Dembek, C. (2022). The economic burden of schizophrenia in the United States. *Journal of Clinical Psychiatry*.

Kane, J. M., & Correll, C. U. (2010). Past and present progress in the pharmacologic treatment of schizophrenia. *Journal of Clinical Psychiatry, 71*(5), 1115-1124.

Kaskie, R. E., Graziano, B., & Ferrarelli, F. (2017). Schizophrenia and sleep disorders: Links, risks, and management challenges. *Nature and Science of Sleep, 9*, 227-239.

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software, 28*(5), 1-26. https://doi.org/10.18637/jss.v028.i05

Leisch, F., & Dimitriadou, E. (2021). Machine learning benchmark problems. *R package version 2.1-3*.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News, 2*(3), 18-22.

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica (Zagreb), 22*(3), 276-282.

Mishra, A., Maiti, R., Mishra, B. R., Jena, M., Nath, S., & Sahu, P. (2020). Effect of add-on ramelteon therapy on sleep and circadian rhythm disruption in patients with schizophrenia: A randomized controlled trial. *European Neuropsychopharmacology, 31*(1), 109-118.

Oh, J., Oh, B. L., Lee, K. U., Chae, J. H., & Yun, K. (2020). Identifying schizophrenia using structural MRI with a deep learning algorithm. *Frontiers in Psychiatry, 11*, 16. https://doi.org/10.3389/fpsyt.2020.00016

Paux, A., & Dimitriadou, E. (2019). Clinical trial simulations. *R package version 1.0.8*.

Santos Febles, E., Ontivero Ortega, M., Valdes Sosa, M., & Sahli, H. (2022). Machine learning techniques for the diagnosis of schizophrenia based on event-related potentials. *Frontiers in Neuroinformatics, 16*, 893788. https://doi.org/10.3389/fninf.2022.893788

Tai, A. M. Y., Albuquerque, A., Carmona, N. E., Subramaniam, M., Cha, D. S., Shek, M., et al. (2019). Machine learning and big data: Implications for disease modeling and therapeutic discovery in psychiatry. *Artificial Intelligence in Medicine, 99*, 101704. https://doi.org/10.1016/j.artmed.2019.101704

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). Springer.

Vrbova, K., Prasko, J., Holubova, M., Slepecky, M., & Ociskova, M. (2018). Positive and negative symptoms in schizophrenia and their relation to depression, anxiety, hope, self-stigma, and personality traits – a cross-sectional study. *Neuroendocrinology Letters, 39*(1), 9-18.

Correspondence:
Rituparna Maiti, MD
Professor, Department of Pharmacology
All India Institute of Medical Sciences (AIIMS),
Bhubaneswar, Odisha, India
pharm_rituparna@aiimsbhubaneswar.edu.in