Neuropsychiatric Disease and Treatment

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

Machine Learning and Interpretability Study for Predicting 30-Day Unplanned Readmission Risk of Schizophrenia: A Retrospective Study

Yuting Tan[1,2,*], Guiling Chen[1,3,*], Shuge Wang[1], Xingxin Zhan[4], Rong Cheng[1,2], Linru Qiao[1], Zhixia Zhang[1], Yaping Liu[1]

[1]Department of Nursing, Tianyou Hospital Affiliated to Wuhan University of Science and Technology, Wuhan, Hubei, People's Republic of China; [2]Institute of Nursing Research, Hubei Province Key Laboratory of Occupational Hazard Identification and Control, School of Medicine,Wuhan University of Science and Technology, Wuhan, Hubei, People's Republic of China; [3]Department of Psychiatry, Tianyou Hospital Affiliated to Wuhan University of Science and Technology, Wuhan, Hubei, People's Republic of China; [4]School of Public Health, Xinyu University, Xinyu, Jiangxi, People's Republic of China

*These authors contributed equally to this work

Correspondence: Zhixia Zhang; Yaping Liu, Department of Nursing, Tianyou Hospital Affiliated to Wuhan University of Science and Technology, Wuhan, Hubei, People's Republic of China, Tel +8613476805827, Email zzx13476805827@126.com; bzmulyp@126.com

Purpose: To build a 30-day unplanned readmission (UPR) risk prediction model based on machine learning (ML) and SHapley Additive exPlanation (SHAP) with data obtained from the electronic medical records (EMRs) of patients with schizophrenia, so as to provide support for early intervention in clinical treatment.

Patients and Methods: This retrospective study selected 1,123 patients with schizophrenia who were hospitalized at least once from January 1, 2021 to June 30, 2024 according to their EMRs. Models were constructed after screening variables using the multiple linear regression and feature importance methods. The model was constructed using five ML algorithms: logistic regression (LR), decision tree (DT), random forest (RF), support vector machine (SVM), and extreme gradient boosting (XGB). The area under the receiver operating characteristic curve (AUC) and SHAP were applied to verify the predictive power and interpretability, respectively, of the five models.

Results: The 30-day UPR rate was 30.54% (343/1,123). The important risk factors were number of somatic comorbid diseases, duration of the disease course, length of the latest hospital stay, drug withdrawal history, and sex. The AUC values of the LR, DT, RF, XGB, and SVM models for predicting the 30-day UPR in the testing set were 0.794, 0.717, 0.823, 0.830, and 0.810, respectively.

Conclusion: An XGB risk prediction model can accurately evaluate the 30-day UPR of patients with schizophrenia. Combined with SHAP, it can provide patients with personalized risk predictions, thereby assisting medical staff in achieving early discharge plans and transitional care.

Keywords: schizophrenia, unplanned readmission, prediction, machine learning, interpretability

Introduction

Schizophrenia is an important challenge in the mental health field, affecting about 24 million people worldwide;[1] its incidence rate is one in every 222 adults (0.45%).[2] The average life expectancy of patients with schizophrenia is about 15 years less than that of general patients and they have a 5–10% lifetime risk of suicide.[3] Even after treatment in professional mental health institutions, the relapse rate of patients with schizophrenia remains high. More than 80% of patients relapse within five years and the recurrence rate is as high as 63% within two years.[4] Up to 86% of patients are readmitted after experiencing a relapse during a 2-year follow-up period,[5] increasing the risk of unplanned readmission (UPR).

Received: 25 February 2025
Accepted: 2 July 2025
Published: 28 July 2025

The UPR rate is a focal point for all healthcare systems and decision-makers because it indicates high costs and unfavorable patient outcomes.[6] Mental illness-related readmission rates range from 5% to 43%,[7,8] which exceed those of the general health condition-related readmission rates of 2.8–38%.[9] Moreover, schizophrenia has the highest readmission rate among all mental disorders.[10] Numerous studies have attempted to predict the risk of rehospitalization after discharge for patients with schizophrenia, with time windows of concern including 7, 30, 90, and 365 days.[11–13] Hospitals widely use the 30-day readmission rate as a quantifiable quality indicator.[14] Compared with a 7-day window, patients may be more likely to experience medication adjustment and social and environmental changes with a 30-day window, resulting in a better opportunity to capture trends in potential relapse risks. It can also reduce the follow-up time and use of medical resources compared with longer time windows.

The "deinstitutionalization" strategy has limited effectiveness in alleviating the 30-day UPR of Chinese patients with schizophrenia,[15] highlighting the urgency of predicting its risk factors. One systematic evaluation identified a low education level, unemployment status, past psychiatric hospitalization history, and hospital length of more than 7 days as key predictors.[16] A reduction in hospital days and the absence of community care after discharge were found to significantly increase the probability of readmission.[17] Specific demographic characteristics (eg, age, sex, education level, and marital status) and clinical characteristics (eg, initial diagnosis, medical comorbidity index, age at first onset, length of the last hospital stay, and number of previous hospitalizations) were also closely associated with 30-day UPR.[16,18–20] However, few studies have constructed clinically realistic and highly interpretable prediction models based on proven predictive factors to accurately predict the 30-day UPR risk of patients with schizophrenia.

Given the remarkable heterogeneity of schizophrenia and its unique disease development trajectory,[21] predicted demand for the disease tends toward personalization rather than simple group-level modeling. Risk prediction models are important for promoting the effective allocation of resources for high-risk patients and for preventing readmission. Previous studies have typically used readmission risk assessment models with traditional logistic regression (LR) or Cox regression,[4,22] which are not ideal for processing large-scale, high-dimensional patient data. Machine learning (ML) algorithms have already adapted to handle nonlinear relationships among complex and interrelated sets of variables.[23] For instance, ML has surpassed traditional statistical models for predicting suicide attempts in patients with major depressive disorder, highlighting its superiority.[24]

Nowadays, ML has been widely used to predict the readmission and prognostic risks of various clinical diseases.[25–27] Through analysis of heterogeneous big data such as electronic medical records (EMRs), ML can find hidden disease rules and risk factors, providing patients with more accurate and personalized risk assessment and intervention suggestions. However, ML models are often called black boxes with internal agnostic properties. In recent years, the SHapley Additive exPlanation (SHAP) method has been used in model interpretation,[28] allowing users to compare the correlations between variables and features by visually presenting the influences and relationships of each input variable on the model output. Some studies have evaluated all-cause readmissions after the discharge of patients with schizophrenia,[10,29] but no research has utilized ML models to predict their 30-day UPR. Meanwhile, the SHAP values used to analyze the interpretability of the models remain poorly understood among such patients.

This study identified the key risk factors leading to 30-day UPR by reviewing standardized EMRs data in the Hospital Information System. It subsequently explored the feasibility of using five ML algorithms to predict the 30-day UPR after the discharge of patients with schizophrenia, and the SHAP method was introduced to improve the interpretability of the model to provide a reference for early clinical identification and intervention.

Material and Methods

Study Design and Participants

This retrospective study extracted the EMRs of patients with schizophrenia hospitalized in a university-affiliated tertiary hospital in Wuhan, Hubei province, China between January 1, 2021, and June 30, 2024, from the Hospital Information System. The "index admission" referred to the initial hospital records of patients diagnosed with schizophrenia who underwent at least one admission and subsequent discharge between January 1, 2021, and June 30, 2024. On this basis, 1,131 inpatients were initially identified, covering their 2,564 admission records. Patients were deemed eligible based on

the following inclusion criteria: they met the International Classification of Diseases-10 diagnosis criteria for schizophrenia and were aged ≥18 years. The exclusion criteria were as follows: death or transfer during the previous hospitalization, severe amounts of missing data, and planned readmission for various reasons. The study ultimately included a cohort of 1,123 patients of whom we had relatively complete demographic and clinical information during the specified period.

Outcomes

The main outcome was the occurrence of UPR within 30 days of discharge. UPR referred to non-predicted readmission due to the same disease or factors associated with discharge disease, as defined by the National Health Insurance Administration.[9] For patients with repeat admissions, the first consideration was whether there were two admission records of less than 30 days apart. The attending physician excluded various planned readmissions, as well as patients with multiple admissions records but more than 30 days apart, to obtain case information for outcome events. Patients who experience multiple 30-day UPRs, regardless of the stage of their illness, are categorized as readmission group. The primary goal of this study is to identify key predictors of 30-day UPR. Given that the demographic and clinical information of these patients remains relatively stable across different readmission events, we uniformly retained and analyzed their EMRs from their first 30-day UPR.

Predictor Variables

This study conducted a preliminary selection of the risk factors for 30-day UPR based on prior research and clinical experience. It classified candidate predictors into demographic and clinical variables. For this study, data from the first admission or first UPR were selected for analysis for each predictor variable. First, this study cleaned and sorted the data on patient attributes. It excluded the variables that could not be analyzed, such as names, hospitalization, and discharge dates. Finally, 16 predictors were selected as independent variables to further explore their relationship with 30-day UPR in subsequent analyses.

ML Algorithms

This study selected five ML techniques as the prediction model-based algorithms. LR, a generalized linear regression analysis model suitable for classification problems, estimates the probability of event occurrence based on a given dataset of independent variables.[30] The decision tree (DT) classification process can be optimized by recursively selecting features and dividing the training data into data subsets based on these features.[31] Random forest (RF) is a supervised ML algorithm based on a DT, reducing the one-sidedness and judgment inaccuracy caused by a single DT.[32] It can handle both classification regression and dimension reduction problems. The extreme gradient boosting (XGB) algorithm has strong scaling, and it is convenient as well, since it can be constructed in parallel.[33] XGB can also balance prediction accuracy and model complexity. It is used for classification, regression, and ranking problems. Support vector machine (SVM) is a supervised learning algorithm used for classification and regression analysis.[34] It implements a classification or regression task by constructing an optimal hyperplane and then uses support vectors to determine its location.

Model Training and Evaluation

This study utilized R 4.3.2 for data analysis, development, and validation. During the data preprocessing phase, this study coded the non-numeric data to ensure compatibility with the algorithm models. In order to deal with the problem of missing data, relevant data in EMRs are used to fill in the missing values. Specifically, for numerical variables, the median is used to fill in the missing values; for categorical variables, the mode is used to fill in the missing values. It processed these data via random downsampling to obtain category-balanced datasets for modeling. It then used the tableone package in R to analyze the demographic information.

The initial feature screening was based on the entire dataset. To identify independent factors capable of predicting 30-day UPR, variables that showed significant differences in univariate logistic regression analysis were included in the multivariate logistic regression analysis. Subsequently, the importance of features was analyzed using SHAP values, and

the results were compared with those from the multivariate analysis. The final results were utilized for model construction, with all subsequent modeling steps strictly segregated between the training and test sets.

This study utilized stratified random sampling to divide the dataset into training (70%) and testing (30%) sets, ensuring that the ratio of patients was consistent in both subsets. We constructed ML models by employing the five distinct algorithms: LR, DT, RF, SVM, and XGB. The training set was utilized for both model training and hyperparameter tuning, with the accuracy from 10-fold cross-validation serving as the key metric for identifying the optimal parameters for each model. To implement this, the caret package in R was used to evenly distribute the training set data into 10 groups. Each iteration involved using one group as the validation data and the remaining nine groups for training to assess the prediction stability of the model.

In the model evaluation stage, the mlr3 and tidymodel libraries in R were used to remove the result labels from the test set data and input them into the trained ML model after parameter tuning. The model made judgments and obtained results, and the predicted results were compared with the real result labels using a confusion matrix. The predictive and evaluative performance of the model was evaluated based on the accuracy, precision, recall, F1 score, and area under the receiver operating characteristic curve (AUC). The SHAP tool was used to conduct post hoc interpretation analysis on the most accurate model, determining the importance of the model features in the decision-making process. Finally, the ggplot2 package in R was used to visualize the model data.

Model Interpretation

The universal SHAP method is used to interpret the output of any ML model.[28] This innovative approach links game theory to local interpretation and provides the only consistent, locally accurate method for calculating additive features. Its introduction is an important milestone in practical clinical decision support. It enables healthcare professionals to more effectively understand and interpret complex predictions made by ML models and recognize how specific factors drive individual patient predictions, thereby improving the accuracy and reliability of clinical decision support.

Results

Participant Characteristics

This study included 1,123 Chinese adult patients with schizophrenia. Among them, 343 patients experienced unplanned readmissions, while 780 did not (30-day unplanned readmission rate: 30.54%). The 30-day unplanned readmission group displayed sex differences ($P<0.001$) (70.0% males, 30.0% females). The most represented age group was 45–59 years (54.8%), followed by 18–44 years (22.7%) and ≥60 years (22.4%). Table 1 presents the descriptive statistics.

Analysis Of the Influencing Factors of UPR

In the univariate logistic regression analysis, we selected statistically significant variables ($P<0.05$) to be included in the multivariate logistic regression model for in-depth analysis. Only the variables showing an extremely high significance level were retained ($P<0.001$) (Table 2). The SHAP method was used for the feature importance analysis (Figure 1). Five common features of the two methods were significant, including in the final risk prediction model: number of somatic comorbid diseases, duration of the disease course, length of the latest hospital stay, drug withdrawal history, and sex.

Model Evaluation And Validation

The LR, RF, XGB, SVM, and DT methods were combined with the five statistically significant variables to establish the ML models. The AUC, accuracy, precision, recall, and F1 values were calculated for the five models. The results showed that in the testing set, the AUC values were 0.794, 0.717, 0.823, 0.830, and 0.810 for the LR, DT, RF, XGB, and SVM models, respectively (Table 3 and Figure 2). The XGB model performed the best, with the highest AUC (0.830), accuracy (0.7768), recall (0.7255), and F1 (0.6637) values. Overall, the XGB model outperformed the other models in identifying the risk factors for the 30-day UPR of patients with schizophrenia.

Table 1 Baseline Characteristics of 30-Day Readmission Vs Non-Readmission Patients (n=1123)

Variable	Non-Readmission (n=780)	Readmission (n=343)	*P*-Value
Sex			<0.001
Male	429 (55.0%)	240 (70.0%)	
Female	351 (45.0%)	103 (30.0%)	
Age (years)			0.014
18–44	176 (22.6%)	78 (22.7%)	
45–59	366 (46.9%)	188 (54.8%)	
≥60	238 (30.5%)	77 (22.4%)	
Educational status			0.071
Junior high school or below	149 (19.1%)	47 (13.7%)	
High school and technical secondary school	371 (47.6%)	167 (48.7%)	
College degree or above	260 (33.3%)	129 (37.6%)	
Marital status			0.134
Unmarried / divorced / widowed	651 (83.5%)	299 (87.2%)	
Married	129 (16.5%)	44 (12.8%)	
Employment status			0.209
Unemployed / retired	571 (73.2%)	264 (77.0%)	
Employed	209 (26.8%)	79 (23.0%)	
Medical insurance			0.049
At one's own expense	91 (11.7%)	23 (6.71%)	
Employee medical insurance	398 (51.0%)	181 (52.8%)	
Residents' medical insurance	198 (25.4%)	87 (25.4%)	
Commercial insurance	93 (11.9%)	52 (15.2%)	
Drug withdrawal history			<0.001
No	268 (34.4%)	61 (17.8%)	
Yes	512 (65.6%)	282 (82.2%)	
History of smoking			0.067
No	458 (58.7%)	222 (64.7%)	
Yes	322 (41.3%)	121 (35.3%)	
History of drinking			0.226
No	287 (36.8%)	140 (40.8%)	
Yes	493 (63.2%)	203 (59.2%)	
Drug history			0.426
No	722 (92.6%)	312 (91.0%)	
Yes	58 (7.44%)	31 (9.04%)	
Main diagnosis			0.009
F20.000	247 (31.7%)	110 (32.1%)	
F20.100	41 (5.26%)	27 (7.87%)	
F20.300	95 (12.2%)	26 (7.58%)	
F20.500	152 (19.5%)	67 (19.5%)	
F20.600	67 (8.59%)	16 (4.66%)	
F20.900	178 (22.8%)	97 (28.3%)	
Number of somatic comorbid diseases			<0.001
≤1	337 (43.2%)	49 (14.3%)	
2–4	319 (40.9%)	104 (30.3%)	
≥5	124 (15.9%)	190 (55.4%)	
Age at first disease onset (years)			0.034
<18	73 (9.36%)	41 (12.0%)	
18–44	619 (79.4%)	279 (81.3%)	
45–59	88 (11.3%)	23 (6.71%)	

(*Continued*)

Table 1 (Continued).

Variable	Non-Readmission (n=780)	Readmission (n=343)	*P*-Value
Duration of the disease course (years)			<0.001
<10	82 (10.5%)	39 (11.4%)	
10–19	206 (26.4%)	153 (44.6%)	
20–29	240 (30.8%)	93 (27.1%)	
≥30	252 (32.3%)	58 (16.9%)	
Number of previous hospitalizations			<0.001
≤1	83 (10.6%)	18 (5.25%)	
2–4	143 (18.3%)	43 (12.5%)	
≥5	554 (71.0%)	282 (82.2%)	
Length of the latest hospital stay (days)			<0.001
≤7	53 (6.79%)	3 (0.87%)	
8–29	396 (50.8%)	268 (78.1%)	
≥30	331 (42.4%)	72 (21.0%)	

Notes: Continuous variables are described using means and standard deviations, while categorical variables are represented using frequencies and percentages.
Abbreviations: Non-readmission, Non-30-day unplanned readmission; Readmission, 30-day unplanned readmission.

Table 2 Multivariate Logistic Regression Analysis Results

Variable	β	SE	Waldχ^2	*P*	OR	95% CI
Sex						
Male	0.782	0.175	20.094	<0.001	2.186	1.558~ 3.090
Drug withdrawal history						
Yes	1.221	0.216	31.895	<0.001	3.391	2.237~ 5.228
Number of somatic comorbid diseases						
2–4	0.920	0.218	17.828	<0.001	2.510	1.645~3.871
≥5	2.844	0.232	150.255	<0.001	17.184	11.019~27.389
Age at first disease onset (years)						
45–59	−1.802	0.517	12.155	<0.001	0.165	0.060~ 0.460
Duration of the disease course (years)						
20–29	−1.076	0.310	12.006	<0.001	0.341	0.185~0.626
≥30	−1.658	0.332	24.957	<0.001	0.190	0.099~0.364
Length of the latest hospital stay (days)						
8–29	2.861	0.700	16.694	<0.001	17.485	5.238~87.246

Model Interpretation

The SHAP is an additive explanatory model inspired by Shapley values. This study performed an overall visualization of the features using the shap.summary_plot method (Figure 3). Among the top five contributing model features, each row represented a feature with an abscissa SHAP value. The dots represented the samples. More yellow indicates larger features, and more purple indicates smaller features. The influencing factors of the XGB model based on SHAP showed that the number of somatic comorbidities was the most important feature affecting 30-day UPR of patients with schizophrenia; the risk of UPR increased with an increase in the number of comorbidities. Long disease duration is a protective factor against 30-day UPR in patients with schizophrenia. Male patients with a history of discontinuation had an increased risk of 30-day UPR. In addition, SHAP can not only analyze the influencing factors of the prediction model at the overall level, but also analyze the influencing factors for individuals.

As shown in Figure 4A, the SHAP value for Patient A was 0.00163, which was less than the model-predicted base value of 0.305; hence, the model predicted that the patient would not have a 30-day UPR. Patient A was female, with

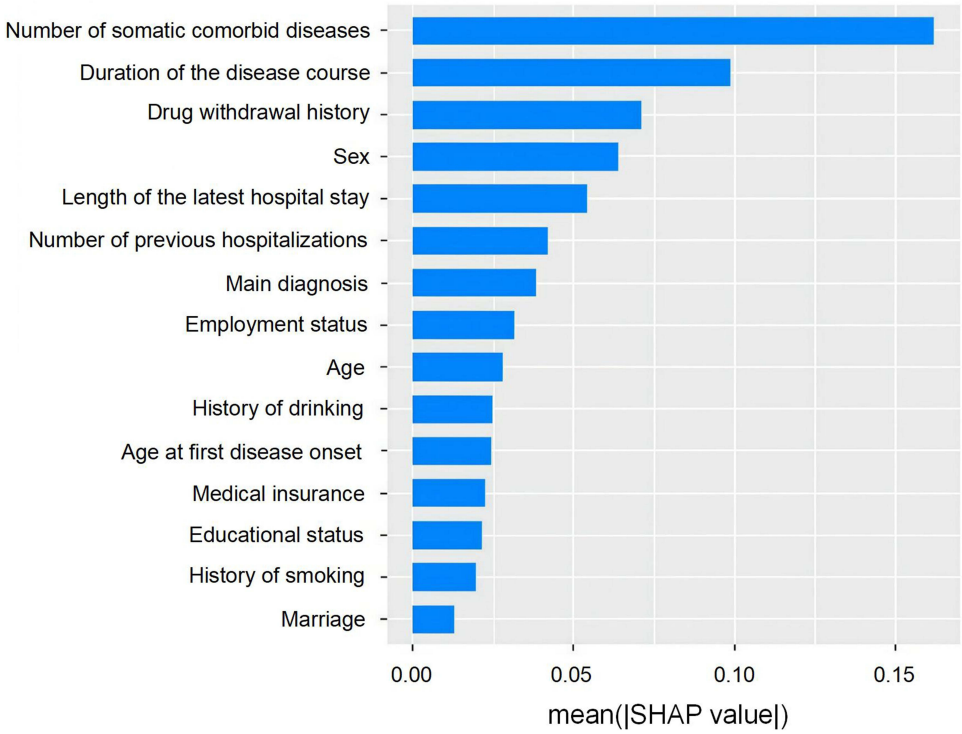

Figure 1 Feature importance analysis.
Notes: The higher the features rank, the more significant its impact on the outcome.

a length of latest hospital stay ≤7 days, duration of disease course ≥30 years, and with a "without drug withdrawal history" value that was blue, indicating that these features reduce the patient's SHAP value and thus reduce the risk of 30-day UPR. Conversely, patient A with five or more somatic comorbid diseases had the opposite effect on the prediction results compared to other characteristics, indicating that they were at a higher risk of 30-day UPR in the future.

As shown in Figure 4B, the SHAP value of Patient B was 0.47, which was greater than the base value predicted by the model; hence, the model predicted that the patient would undergo a 30-day UPR. The number of somatic comorbid diseases of Patient B was blue, which played a positive role in reducing the risk of 30-day UPR. This patient was male, had a drug withdrawal history, the latest hospital stay duration of 8–29 days, and a duration of disease course of 10–19 years. These factors placed the patient in a higher category of risk for 30-day UPR. In addition, of all the relevant variables, the duration of the disease course had the widest range and the greatest impact, which should be paid attention to.

Discussion

This study is the first to construct a 30-day UPR model of patients with schizophrenia using multiple predictors extracted from EMRs data via ML, with SHAP values used to interpret the findings. While prior research has used ML to predict

Table 3 Performance Metrics of the Machine Learning Models Based on the 10-Fold Cross-Validation Testing Set

Models	Accuracy	Precision	Recall	F1	AUC
LR	0.7262	0.5379	0.6961	0.6068	0.794
DT	0.7530	0.5960	0.5784	0.5871	0.717
RF	0.7738	0.6140	0.6863	0.6481	0.823
XGB	0.7768	0.6116	0.7255	0.6637	0.830
SVM	0.7500	0.5818	0.6275	0.6038	0.810

Abbreviations: LR, logistic regression; DT, decision tree; RF, random forest; XGB, extreme gradient boosting; SVM, support vector machine; AUC, area under the receiver operating characteristic curve.

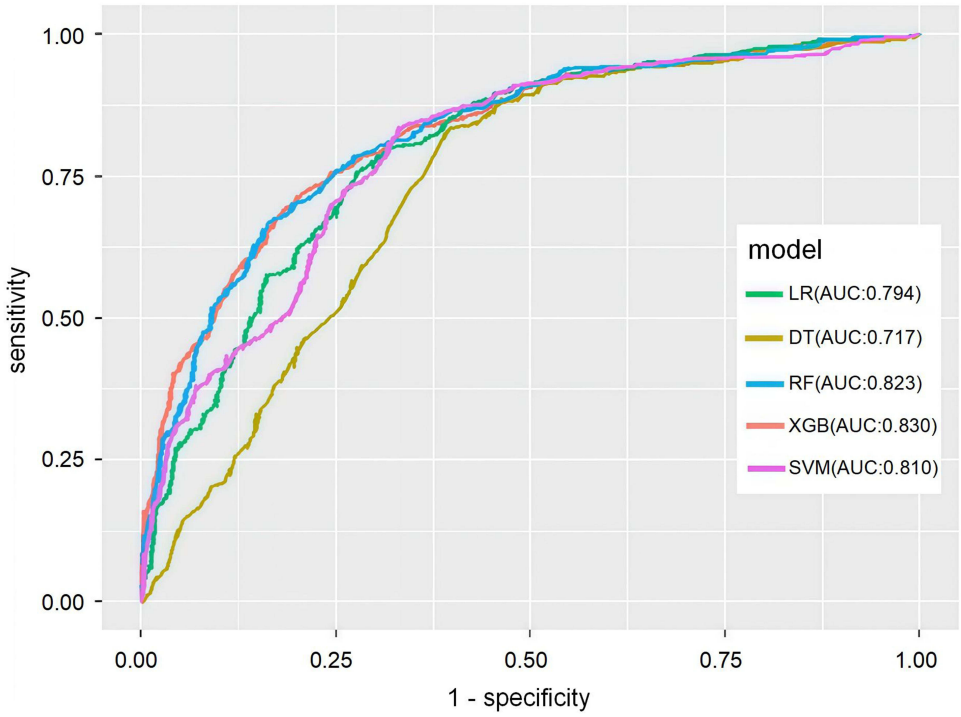

Figure 2 Comparison of the prediction performance of machine learning models based on the 10-fold cross-validation test set.

Figure 3 SHAP summary plot of the 5 risk factors.
Notes: Among the top five contributing model features, each row represented a feature with an abscissa SHAP value. The dots represented the samples. More yellow indicates larger features, and more purple indicates smaller features.

the risk of all-cause readmissions or UPR,[25–27] studies predicting the risk of readmission for patients with schizophrenia remain limited.[10,29] Therefore, further exploration is warranted. The results of this study showed that the 30-day UPR rate of patients with schizophrenia was 30.54% (343/1,123), which aligns with the findings from comparative studies;

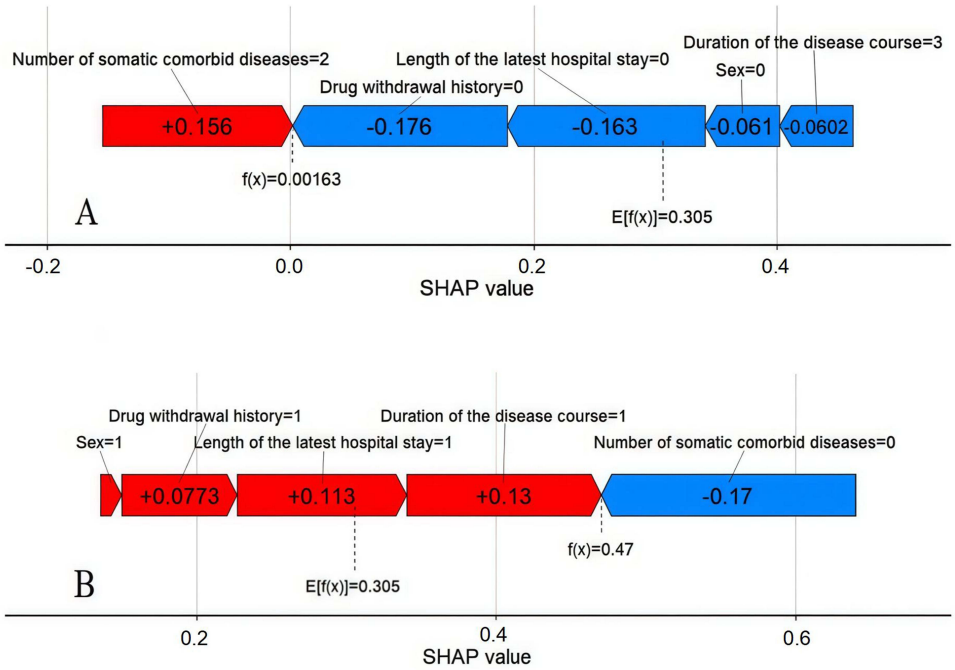

Figure 4 Examples of SHAP interpretation for patients (**A** and **B**).
Notes: The red bars represent positive significance for the outcome index, and blue bars represent negative significance for the outcome index; Number of somatic comorbid diseases=0 is ≤1, Number of somatic comorbid diseases=1 is 2–4, Number of somatic comorbid diseases=2 is ≥5; Duration of the disease course=0 is <10 years, Duration of the disease course=1 is 10–19 years, Duration of the disease course=2 is 20–29 years, Duration of the disease course=3 is ≥30 years; Length of the latest hospital stay=0 is ≤7 days, Length of the latest hospital stay=1 is 8–29 days, Length of the latest hospital stay=2 is ≥30 days; Drug withdrawal history=0 is No, Drug withdrawal history=1 is Yes; Sex=0 is female,Sex=1 is male.

however, this rate was lower than the 52.56% readmission rate in 11 Spanish hospitals (1,611/3,065).[4,29] Variations in rehospitalization rates may be influenced by the specific policies in countries or regions as well as by differences in psychiatric hospital bed turnover and the provision of community care services after discharge.

This study produced AUC values that ranged from 0.717 to 0.830. A previous study utilized generalized linear models via lasso and elastic-net regularization (AUC=0.697) and XGB (AUC=0.738) models to predict the readmission of patients with mental disorders.[27] Further, a previous study evaluated six ML algorithms using the hospital EMRs of 10,358 patients aged >65 years undergoing surgery with general anesthesia; they found that all of the models have unremarkable predictive abilities for 30-day UPR rates (AUC: 0.6865–0.8654, precision: 0.2683–0.3977, recall: 0.1550–0.3289, and F1: 0.1957–0.4909).[35] A prior study used seven ML models to predict 30-day UPR rates after spinal fusion in 59,145 patients, but only general performance was observed (AUC: 0.63–0.66).[36] Overall, the ML models in this study demonstrated promising predictive performance during internal validation, indicating the feasibility of constructing such predictive models. These findings also provide support for future large-scale external validation studies to explore the potential application value of the models in identifying and preventing adverse outcomes in patients with schizophrenia.

This study found that the overall predictive power of the XGB model exceeded that of the other four models. The predictive power of the included features, and the explanatory features between them, is interesting: few studies have considered easily collected features for modeling the 30-day UPR of patients with schizophrenia. In addition, this study utilized multivariate logistic regression and SHAP values to quantify the importance of model features and ultimately identified the top five key features: number of somatic comorbid diseases, duration of the disease course, length of the latest hospital stay, drug withdrawal history, and sex.

Within the dataset, the number of somatic comorbid diseases showed the most pronounced significance. Extant research has highlighted the close connection between physical health status and mental illness.[16,37] Notably, a cross-sectional study revealed an independent correlation between coexisting chronic somatic disease and higher psychiatric

readmission rates in patients with schizophrenia after excluding interfering factors,[38] reinforcing the argument that somatic comorbid diseases can significantly exacerbate the risk of UPR within 30 days in patients with schizophrenia. Severe somatic comorbidities may complicate treatment and increase the overall vulnerability of the individual.[39] Comorbid multiple somatic disorders may require a combination of multiple drugs, which can result in reduced patient tolerance to psychotropic drugs and an increased risk of a treatment interruption.[40] Therefore, the identification and management of comorbid conditions are crucial to controlling the disease status of schizophrenia patients, preventing readmissions, and improving their quality of life.

Multiple studies have explored the potential association between disease duration and readmission; however, they have not revealed a significant association.[41,42] Conversely, this study found that patients in the intermediate stage of the disease (10–15 years) were at higher risk of 30-day UPR. This finding may relate to attitudes and behavioral changes of the patients at this stage. A meta-analysis noted that the intervention in the early stage of the course of the disease was effective and that the improvement of symptoms after >10 years was relatively limited.[43] A prospective study indicated that the average number of patients hospitalized for psychiatric reasons was 2.3, compared with 1.4 for medical reasons over a study period of 12 to 27 years.[44] In the mid-stage of the disease, some patients may gradually lose confidence in treatment or grow tired of disease management. Therefore, for patients at this stage, hospitals, communities, and families must work together to provide them with comprehensive social support and mental health education to improve their treatment confidence and overall quality of life.

To increase bed turnover and decrease medical costs, medical insurance policies and regulations related to medical institutions tend to encourage minimizing the average hospital stay. However, this study found that patients with schizophrenia with a hospital stay of 8–29 days had an up to a 78.1% higher risk of 30-day UPR. This aligns with a prior study, which found that patients with schizophrenia hospitalized for fewer than 30 days had more frequent readmissions.[45] Furthermore, a 3-year cohort study observed that a shorter hospital stay was significantly associated with an increased 30-day readmission rate.[46] There is still a lack of uniform standards for the optimal length of hospitalization that can reduce medical costs without increasing the risk of readmission. Given the unique nature of schizophrenia, patients require more observation time, treatment plan adjustments, and stabilization before discharge to attain better clinical improvement. Consequently, medical staff should better determine the best time to discharge by carefully assessing the patient's treatment response, disease stability, and post-discharge support measures, thus reducing the risk of UPR and facilitating long-term recovery.

This study revealed that patients with low medication adherence faced a higher risk of 30-day UPR, reinforcing the perspective of previous studies, which not only revealed a strong link between treatment non-adherence and the planned withdrawal of antipsychotics and high readmission rates in schizophrenia,[47,48] but also validated the utility of these predictors in real medical practice. A previous study predicted significant associations between drug non-adherence and treatment outcomes in schizophrenia, based on ML models that account for the complex interactions and mutual influence of multiple factors.[49] This shows that medical staff must regularly monitor patients' conditions, observe the effect of drug treatment, and timely adjust treatment and rehabilitation programs when necessary. In the future, ML could be used to dynamically predict the risk of medication non-compliance in patients with schizophrenia,[50] which would have indispensable value in optimizing management and reducing UPR.

This study found that male had a higher risk of 30-day UPR, consistent with previous findings,[19] and highlighting sex differences in the management of schizophrenia. The data clearly revealed higher prevalence of schizophrenia in male than in female, at a ratio close to 1.4:1.[51] Male patients with schizophrenia often face more severe social distress, coupled with the fact that their treatment compliance is generally lower than that of female patients, aggravating the difficulty of establishing a treatment relationship and resulting in a high involuntary treatment rate, high hospitalization rate, and high suicide rate.[52] Therefore, medical staff should more finely adjust and optimize the needs of male patients in order to reduce the risk of 30-day UPR and significantly improve both their quality of life and their rehabilitation prospects.

Limitations

This study had several limitations. First, the use of only retrospective EHR data for extracting basic clinical information led to the omission of other potential risk factors that are not documented, thus affecting the

comprehensiveness of the prediction. Second, this study was confined to using data from a single tertiary hospital; future studies should use multi-center and multi-region data to enhance the generality and applicability of the results. Third, this study only underwent internal validation. Accordingly, additional external validation is required to establish the credibility of the results.

Conclusion

This study constructed 30-day UPR risk prediction models of patients with schizophrenia using five ML algorithms. After evaluation and comparison, the XGB model demonstrated significantly superior performance compared to other models, achieving an AUC value of 0.830, an accuracy rate of 0.7768, a precision rate of 0.6116, a recall rate of 0.7255, and an F1 score of 0.6637, indicating high robustness. The SHAP method further explained key risk factors, including number of somatic comorbid diseases, duration of the disease course, length of the latest hospital stay, drug withdrawal history, and sex, to support early identification of high-risk patients after discharge and optimization of transitional care plans. This study provides a scientific basis for leveraging ML model to improve the health outcomes and reduce healthcare costs for patients with schizophrenia, which can be further promoted and applied through multi-center external verification and dynamic risk assessment in the future.

Data Sharing Statement

The datasets used in this study are available from the corresponding author named Zhixia Zhang on reasonable request. Due to the sensitivity of the schizophrenia patient data, which may contain personally identifiable information, we could not provide open access to our data.

Ethics Approval

This study obtained approval from the Ethics Review Committee of Tianyou Hospital Affiliated to Wuhan University of Science and Technology (LL2024032102). The study was conducted in accordance with the Declaration of Helsinki. The need for written informed consent was waived by the Ethics Review Committee of Tianyou Hospital Affiliated to Wuhan University of Science and Technology due to the retrospective design of the study. Furthermore, the data sources were medical records filed by patients after discharge. The data collection did not interfere with patient diagnosis or treatment, and the study data did not contain identifiable personal information.

Acknowledgments

The authors thank the participants for their support to this study.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

There is no funding support available for this study.

Disclosure

The authors declare no competing interests.

References

1. WHO. Schizophrenia. https://www.who.int/news-room/fact-sheets/detail/schizophrenia. Accessed October 20, 2024.
2. Institute of health Metrics and Evaluation (IHME). Global Health Data Exchange(GHDx). Available from: http://ghdx.healthdata.org/gbd-results-tool?params=gbd-api-2019-permalink/27a7644e8ad28e739382d31e77589dd7. Accessed October 25, 2024.

3. Hjorthøj C, Stürup AE, McGrath JJ, Nordentoft M. Years of potential life lost and life expectancy in schizophrenia: a systematic review and meta-analysis. *Lancet Psychiatry*. 2017;4(4):295–301. doi:10.1016/S2215-0366(17)30078-0
4. Sato A, Moriyama T, Watanabe N, Maruo K, Furukawa TA. Development and validation of a prediction model for rehospitalization among people with schizophrenia discharged from acute inpatient care. *Front Psychiatry*. 2023;14:1242918. doi:10.3389/fpsyt.2023.1242918
5. Schennach R, Riedel M, Obermeier M, et al. What happens with schizophrenia patients after their discharge from hospital? Results on outcome and treatment from a "real-world" 2-year follow-up trial. *Eur Arch Psychiatry Clin Neurosci*. 2020;270(6):661–671. doi:10.1007/s00406-019-01055-4
6. Neto C, Senra F, Leite J, et al. Different Scenarios for the Prediction of Hospital Readmission of Diabetic Patients. *J Med Syst*. 2021;45(1):11. doi:10.1007/s10916-020-01686-4
7. Innes H, Lewsey J, Smith DJ. Predictors of admission and readmission to hospital for major depression: a community cohort study of 52,990 individuals. *J Affect Disord*. 2015;183:10–14. doi:10.1016/j.jad.2015.04.019
8. Doerfler LA, Moran PW, Hannigan KE. Situations associated with admission to an acute care inpatient psychiatric unit. *Psychol Serv*. 2010;7(4):254–265. doi:10.1037/a0020642
9. Zhou H, Della PR, Roberts P, Goh L, Dhaliwal SS. Utility of models to predict 28-day or 30-day unplanned hospital readmissions: an updated systematic review. *BMJ Open*. 2016;6(6):e011060. doi:10.1136/bmjopen-2016-011060
10. Thongkam J, Sukmak V. Enhancing decision tree with adaboost for predicting schizophrenia readmission. *Adv Mat Res*. 2014;931:1467–1471. doi:10.4028/www.scientific.net/AMR.931-932.1467
11. Heo J, Yoon NH, Shin S, Yu SY, Lee M. Effects of the Mental Health and Welfare Law revision on schizophrenia patients in Korea: an interrupted time series analysis. *Int J Ment Health Syst*. 2021;15(1):76. doi:10.1186/s13033-021-00499-3
12. Benjenk I, Shields M, Chen J. Measures of Care Coordination at Inpatient Psychiatric Facilities and the Medicare 30-Day All-Cause Readmission Rate. *Psychiatr Serv*. 2020;71(10):1031–1038. doi:10.1176/appi.ps.201900360
13. Zhu T, Jiang J, Hu Y, Zhang W. Individualized prediction of psychiatric readmissions for patients with major depressive disorder: a 10-year retrospective cohort study. *Transl Psychiatry*. 2022;12(1):170. doi:10.1038/s41398-022-01937-7
14. Patel N, Singh S, Desai R, et al. Thirty-day unplanned readmission in hospitalised asthma patients in the USA. *Postgrad Med J*. 2022;98(1165):830–836. doi:10.1136/postgradmedj-2021-140735
15. Zhong Q, Tan Y, Chen W, et al. Disease burden of schizophrenia patients visiting a Chinese regional mental health centre. *J Comp Eff Res*. 2020;9(7):469–481. doi:10.2217/cer-2019-0129
16. Zhou H, Ngune I, Albrecht MA, Della PR. Risk factors associated with 30-day unplanned hospital readmission for patients with mental illness. *Int J Ment Health Nurs*. 2023;32(1):30–53. doi:10.1111/inm.13042
17. Chen S, Collins A, Anderson K, Kidd S, Kidd S. Patient Characteristics, Length of Stay, and Functional Improvement for Schizophrenia Spectrum Disorders: a Population Study of Inpatient Care in Ontario 2005 to 2015. *Can J Psychiatry*. 2017;62(12):854–863. doi:10.1177/0706743716680167
18. Mortensen PB, Eaton WW. Predictors for readmission risk in schizophrenia. *Psychol Med*. 1994;24(1):223–232. doi:10.1017/s0033291700026982
19. Cook JA, Burke-Miller JK, Jonikas JA, Aranda F, Santos A. Factors associated with 30-day readmissions following medical hospitalizations among Medicaid beneficiaries with schizophrenia, bipolar disorder, and major depressive disorder. *Psychiatry Res*. 2020;291:113168. doi:10.1016/j.psychres
20. Sugisawa S, Kurihara T, Nakano Y, et al. Risk factors for readmission in schizophrenia treated with combined psychoeducation and standard therapy. *Neuropsychopharmacol Rep*. 2022;42(1):77–83. doi:10.1002/npr2.12229
21. Lysaker PH. Schizophrenia, the self and psychotherapy: is there really anything new under the sun? *Clin Psychol*. 2021;77(8):1865–1870. doi:10.1002/jclp.23234
22. Sutter T, Roth JA, Chin-Cheong K, Hug BL, Vogt JE. A comparison of general and disease-specific machine learning models for the prediction of unplanned hospital readmissions. *J Am Med Inform Assoc*. 2021;28(4):868–873. doi:10.1093/jamia/ocaa299
23. Artetxe A, Beristain A, Graña M. Predictive models for hospital readmission risk: a systematic review of methods. *Comput Methods Programs Biomed*. 2018;164:49–64. doi:10.1016/j.cmpb.2018.06.006
24. Ji X, Zhao J, Fan L, et al. Highlighting psychological pain avoidance and decision-making bias as key predictors of suicide attempt in major depressive disorder-A novel investigative approach using machine learning. *Clin Psychol*. 2022;78(4):671–691. doi:10.1002/jclp.23246
25. Howard EP, Morris JN, Schachter E, Schwarzkopf R, Shepard N, Buchanan ER. Machine-Learning Modeling to Predict Hospital Readmission Following Discharge to Post-Acute Care. *J Am Med Dir Assoc*. 2021;22(5):1067–1072.e29. doi:10.1016/j.jamda.2020.12.017
26. Bolourani S, Tayebi MA, Diao L, et al. Using machine learning to predict early readmission following esophagectomy. *Thorac Cardiovasc Surg*. 2021;161(6):1926–1939.e8. doi:10.1016/j.jtcvs.2020.04.172
27. Morel D, Yu KC, Liu-Ferrara A, Caceres-Suriel AJ, Kurtz SG, Tabak YP. Predicting hospital readmission in patients with mental or substance use disorders: a machine learning approach. *Int J Med Inform*. 2020;139:104136. doi:10.1016/j.ijmedinf.2020.104136
28. Lundberg S, Lee SI. A unified approach to interpreting model predictions. In: 31st Conference on neural information processing systems (NIPS). Long Beach, California, USA;2017.
29. Góngora AS, Herrera MI, Ayala JLM, Rodrigues JJPC, Franco-Martín M, de la Torre Díez I. Machine Learning Models to Predict Readmission Risk of Patients with Schizophrenia in a Spanish Region. *Int J Ment Health Ad*. 2023;22(4):2508–2527. doi:10.1007/s11469-022-01001-x
30. Sperandei S. Understanding logistic regression analysis. *Biochem Med*. 2014;24(1):12–18. doi:10.11613/bm.2014.003
31. Osong B, Bermejo I, Lee KC, Lee SH, Dekker A, van Soest J. Prediction of Radiotherapy Compliance in Elderly Cancer Patients Using an Internally Validated Decision Tree. *Cancers*. 2022;14(24):6116. doi:10.3390/cancers14246116
32. Wang KZ, Bani-Fatemi A, Adanty C, et al. Prediction of physical violence in schizophrenia with machine learning algorithms. *Psychiatry Res*. 2020;289:112960. doi:10.1016/j.psychres.2020
33. Dalal S, Onyema EM, Malik A. Hybrid XGBoost model with hyperparameter tuning for prediction of liver disease with better accuracy. *World J Gastroenterol*. 2022;28(46):6551–6563. doi:10.3748/wjg.v28.i46.6551
34. Utkin V. An imprecise extension of SVM-based machine learning models. *Neurocomputing*. 2019;331:18–32. doi:10.1016/j.neucom.2018.11.053
35. Li L, Wang L, Lu L, Zhu T. Machine learning prediction of postoperative unplanned 30-day hospital readmission in older adult. *Front Mol Biosci*. 2022;9:910688. doi:10.3389/fmolb.2022.910688

36. Goyal A, Ngufor C, Kerezoudis P, McCutcheon B, Storlie C, Bydon M. Can machine learning algorithms accurately predict discharge to nonhome facility and early unplanned readmissions following spinal fusion? Analysis of a national surgical registry. *J Neurosurg Spine*. 2019;31(4):568–578. doi:10.3171/2019.3.SPINE181367
37. Ku BS, Lally CA, Compton MT, Druss BG. Neighborhood Predictors of Outpatient Mental Health Visits Among Persons With Comorbid Medical and Serious Mental Illnesses. *Psychiatr Serv*. 2020;71(9):906–912. doi:10.1176/appi.ps.201900363
38. Filipcic I, Simunovic, Filipcic I, et al. Chronic physical illnesses in patients with schizophrenia spectrum disorders are independently associated with higher rates of psychiatric rehospitalization; a cross-sectional study in Croatia. *Eur Psychiatry*. 2017;43:73–80. doi:10.1016/j.eurpsy.2017.02.484
39. Arruda AL, Khandaker GM, Morris AP, Smith GD, Huckins LM, Zeggini E. Genomic insights into the comorbidity between type 2 diabetes and schizophrenia. *Schizophrenia*. 2024;10(1):22. doi:10.1038/s41537-024-00445-5
40. Owen-Smith A, Stewart C, Green C, et al. Adherence to common cardiovascular medications in patients with schizophrenia vs. patients without psychiatric illness. *Gen Hosp Psychiatry*. 2016;38:9–14. doi:10.1016/j.genhosppsych.2015.07.010
41. Fountoulakis KN, Dragioti E, Theofilidis AT, et al. Gender, age at onset, and duration of being ill as predictors for the long-term course and outcome of schizophrenia: an international multicenter study. *CNS Spectr*. 2022;27(6):716–723. doi:10.1017/S1092852921000742
42. Wieselgren IM, Lindstrom LH. A prospective 1-5 year outcome study in first-admitted and readmitted schizophrenic patients; relationship to heredity, premorbid adjustment, duration of disease and education level at index admission and neuroleptic treatment. *Acta Psychiatr Scand*. 1996;93(1):9–19. doi:10.1111/j.1600-0447.1996.tb10613.x
43. de Winter L, Vermeulen JM, Couwenbergh C, et al. Short- and long-term changes in symptom dimensions among patients with schizophrenia spectrum disorders and different durations of illness: a meta-analysis. *J Psychiatr Res*. 2023;164:416–439. doi:10.1016/j.jpsychires.2023.06.031
44. Baltazar L, De Benedictis L, Abdel-Baki A, Lalonde P, Lesage A. Long term course and outcome of first episode schizophrenia: a 27-to-31-year follow-up. *Soc Psychiatry Psychiatr Epidemiol*. 2022;57(7):1319–1328. doi:10.1007/s00127-021-02185-8
45. Barnett BS, Kusunzi V, Magola L, et al. Factors associated with long length of stay in an inpatient psychiatric unit in Lilongwe, Malawi. *Soc Psychiatry Psychiatr Epidemiol*. 2019;54(2):235–242. doi:10.1007/s00127-018-1611-1
46. Lin HC, Tian WH, Chen CS, Liu TC, Tsai SY, Lee HC. The association between readmission rates and length of stay for schizophrenia: a 3-year population-based study. *Schizophr Res*. 2006;83(2–3):211–214. doi:10.1016/j.schres.2006.01.012
47. Morken G, Widen JH, Grawe RW. Non-adherence to antipsychotic medication, relapse and rehospitalisation in recent-onset schizophrenia. *BMC Psychiatry*. 2008;8(1):32. doi:10.1186/1471-244X-8-32
48. Karabulut B, Uslu E. Schizophrenia and medication adherence: associated factors. *Arch Psychiatr Nurs*. 2024;49:47–54. doi:10.1016/j.apnu.2024.01.015
49. Barruel D, Hilbey J, Charlet J, Chaumette B, Krebs MO, Dauriac-Le Masson V. Predicting treatment resistance in schizophrenia patients: machine learning highlights the role of early pathophysiologic features. *Schizophr Res*. 2024;270:1–10. doi:10.1016/j.schres.2024.05.011
50. Koesmahargyo V, Abbas A, Zhang L, et al. Accuracy of machine learning-based prediction of medication adherence in clinical research. *Psychiatry Res*. 2020;294:113558. doi:10.1016/j.psychres.2020.113558
51. Li X, Zhou W, Yi Z. A glimpse of gender differences in schizophrenia. *Gen Psychiatr*. 2022;35(4):e100823. doi:10.1136/gpsych-2022-100823
52. Seeman MV. Sex differences in schizophrenia relevant to clinical care. *Expert Rev Neurother*. 2021;21(4):443–453. doi:10.1080/14737175.2021.1898947