




Leveraging deep neural network and language models for predicting long-term hospitalization risk in schizophrenia

Yihang Bao^{1,4} , Wanying Wang^{1,4}, Zhe Liu¹, Weidi Wang^{1,2}, Xue Zhao¹, Shunying Yu^{1,2}  and Guan Ning Lin^{1,2,3} 

Early warning of long-term hospitalization in schizophrenia (SCZ) patients at the time of admission is crucial for effective resource allocation and individual treatment planning. In this study, we developed a deep learning model that integrates demographic, behavioral, and blood test data from admission to forecast extended hospital stays using a retrospective cohort. By utilizing language models, our developed algorithm efficiently extracts 95% of the unstructured electronic health records data needed for this work, while ensuring data privacy and low error rate. This paradigm has also been demonstrated to have significant advantages in reducing potential discrimination and erroneous dependencies. By utilizing multimodal features, our deep learning model achieved a classification accuracy of 0.81 and an AUC of 0.9. Key risk factors identified included advanced age, longer disease duration, and blood markers such as elevated neutrophil-to-lymphocyte ratio, lower lymphocyte percentage, and reduced albumin levels, validated through comprehensive interpretability analyses and ablation studies. The inclusion of multimodal data significantly improved prediction performance, with demographic variables alone achieving an accuracy of 0.73, which increased to 0.81 with the addition of behavioral and blood test data. Our approach outperformed traditional machine learning methods, which were less effective in predicting long-term stays. This study demonstrates the potential of integrating diverse data types for enhanced predictive accuracy in mental health care, providing a robust framework for early intervention and personalized treatment in SCZ management.

Schizophrenia (2025)11:35; <https://doi.org/10.1038/s41537-025-00585-2>


MAIN

Schizophrenia (SCZ) is a chronic and severe mental disorder that profoundly affects an individual's thoughts, feelings, and behaviors. Patients with SCZ often experience symptoms such as delusions, hallucinations, disorganized thinking, and impaired functioning, leading to significant distress and disability^{1,2}. These symptoms pose substantial risks not only to the patients but also to society at large. Due to the increased risk of suicide and other health problems, individuals with SCZ have a higher mortality rate, resulting in a reduced lifespan of ~15–20 years³. Given these risks, inpatient treatment is crucial for effectively managing and treating individuals with SCZ⁴. However, the rising prevalence of SCZ⁵, coupled with a large patient population, exacerbates the challenges related to the allocation of healthcare resources⁶. Healthcare institutions must allocate resources based on the individual conditions of patients upon admission to enhance treatment efficiency and ensure early intervention for those at risk of treatment resistance. A critical aspect of this process is the reliable early identification of patients at risk of prolonged hospitalization and the understanding of the factors that significantly impact the length of stay (LOS)⁷.

Long-stay patients, defined as those requiring hospitalization for an extended period due to the severity and complexity of their condition, present a significant challenge for healthcare systems. Identifying these patients at the time of admission is crucial. Studies have shown that early identification of long-stay patients can enhance treatment efficiency and reduce the financial burden of patients^{8–11}. For instance, a study on post-admission medical interventions for COVID-19 patients demonstrated that targeted interventions based on specific patient risk factors could

significantly shorten hospital stays¹². Similarly, a retrospective study in the Philippines indicated that inappropriate allocation of medical resources at the time of admission led to prolonged hospitalizations and increased financial burden for some patients⁸. Several studies have also indicated that early prediction of LOS in the field of psychiatry is beneficial for resource optimization. For instance, La et al. developed a Cox proportional hazards model based on data from three state psychiatric hospitals in North Carolina to estimate patients' lengths of stay, incorporating these estimates into a discrete event simulation model¹³. This model can simulate patient flow within the hospital, providing decision support for hospital management and facilitating more effective adjustments in resource allocation and bed management. Thus, early identification of long-stay patients can facilitate better allocation of resources, improve efficiency, and reduce costs from the perspective of healthcare institutions¹⁴.

Hospitalization durations for SCZ are notably long. For example, in South Korea, the average hospitalization duration is 78 ± 76 days¹⁵, in China, it is 73.3 ± 42.2 days¹⁶, and in Canada it is 96.6 days¹⁷. These surveys are all based on acute care institutions. This phenomenon highlights the importance of early identification of long-stay patients. To address this need, studies have explored the differences between long-stay and short-stay SCZ patients at the time of admission. Demographic variables, such as age, gender, marital status, employment, and race have been found to influence the length of hospital stay^{14,18–20}. Additionally, behavioral variables, such as the presence of aggressive behavior or suicidal ideation, have been identified as factors affecting hospitalization duration^{21–23}. Both types of variables can be obtained at the time of patient admission.

¹Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China. ²Shanghai Key Laboratory of Psychotic Disorders, Shanghai, China. ³Engineering Research Center of Digital Medicine of the Ministry of Education, Shanghai, China. ⁴These authors contributed equally: Yihang Bao, Wanying Wang. email: yushunying@smhc.org.cn; nickgnlin@sjtu.edu.cn

Research has also explored the association between clinical variables, such as discharge medication dosage, and the length of hospital stay^{19,24}. However, these clinical variables are not available at the time of admission. Building on these insights, researchers have developed computational models to predict long-stay patients using multi-perspective admission information. For instance, Kirchebner et al., combined patient characteristics, criminal history, and medical history to construct a machine learning-based forensic psychiatry hospitalization duration prediction model²⁵. Similarly, Cheng et al. developed a regression model for predicting hospitalization duration using demographic and biochemical variables, finding an association between marital status, thyroid-stimulating hormone levels, and prolonged hospital stays²⁶. While these studies have made valuable contributions, they also faced several challenges. The study by Kirchebner et al. employed manual methods to extract unstructured data, which are not applicable to large-scale datasets. The work by Cheng et al. was limited by the number of features and algorithm efficiency, resulting in a model with an AUC of only 0.577, thereby reducing its applicability. In light of this, there is a need for an advanced algorithmic framework that combines automatic unstructured data extraction to achieve more accurate long-term hospitalization risk prediction for SCZ patients.

The rapid development of deep neural networks and language models (LMs) has enabled more efficient and feasible feature extraction and learning. In this study, we leveraged these advanced techniques to predict long-stay risks in SCZ patients based on various types of admission variables. Specifically, we extracted unstructured behavioral features from a large set of retrospective electronic health records (EHRs) of SCZ patients using a semi-automated process based on LMs. This extraction process can be quickly completed in a few-shot setting, outperforming methods based on regular expressions or BERT^{27,28}, an advanced LMs that uses bidirectional pre-training to thoroughly understand the context of words in sentences. Our model predicts the risk of prolonged hospitalization for SCZ by simultaneously integrating behavioral features, demographic characteristics, and admission blood test indicators for the first time. Building on this, we offer a new perspective on the risk factors associated with prolonged hospitalization in schizophrenia.

This study aims to investigate whether the latest deep learning paradigms and algorithms can provide more accurate predictions of the risk of prolonged hospitalization for schizophrenia. We hypothesize that the integration of behavioral features extracted by LMs, along with demographic characteristics and blood test indicators, will more comprehensively and accurately capture the risk patterns of patients with prolonged hospitalization for schizophrenia. Furthermore, we explore which features are significant risk factors for extended hospitalization, providing guidance for precision medicine strategies.

RESULTS

Patient characteristics and basic analysis

Our retrospective dataset, sourced from the Shanghai Mental Health Center (SMHC), comprises EHRs of SCZ inpatients discharged between 2018 and 2022. Details on the data processing and screening methods are shown in the Methods section and Extended Data Fig. 1. Following the commonly used methods from previous work²⁹, we utilized the mean plus one standard deviation as the threshold for prolonged hospitalization and the mean minus one standard deviation as the threshold for short hospitalization. Specifically, we defined hospitalization durations exceeding 120 days as long-stay and durations shorter than 30 days as short-stay. Patients with a hospital stay longer than one year were excluded, as their LOS is typically influenced by non-medical factors. This yielded a dataset of 1546 short-stay

Table 1. Descriptive statistics on patients.

LOS	Short	Long	P value
N	1546	812	
Age of admission (years)	35.9 ± 18.0	48.7 ± 17.3	<1e ⁻⁶
Disease duration (days)	3802 ± 5016	7932.1 ± 5651.6	<1e ⁻⁶
Sex			2.6e ⁻²
Female	866(56.0%)	415(51.1%)	
Male	680(44.0%)	397(48.9%)	
Marital status			<1e ⁻⁶
Married	468(30.3%)	179(22.0%)	
Divorced/Widowed	116(7.5%)	117(14.4%)	
Unmarried	962(62.2%)	516(63.5%)	
Hemoglobin	135.3 ± 17.6	132.3 ± 17.3	4.6e ⁻⁵
Neutrophil	4.5 ± 1.9	4.4 ± 1.8	9.9e ⁻¹
Neutrophil%	61 ± 11.5	63.1 ± 10.4	1.2e ⁻⁵
Eosinophil	0.125 ± 0.137	0.12 ± 0.14	7.1e ⁻²
Eosinophil%	1.87 ± 2.06	1.8 ± 2.1	4.5e ⁻¹
Basophil	0.026 ± 0.017	0.024 ± 0.018	2.2e ⁻²
Basophil%	0.38 ± 0.25	0.37 ± 0.25	2.5e ⁻¹
Lymphocyte	2.0 ± 0.7	1.8 ± 0.7	<1e ⁻⁶
Lymphocyte%	29.4 ± 10.4	27.4 ± 9.3	1.8e ⁻⁵
Monocyte	0.5 ± 0.2	0.5 ± 0.2	6.1e ⁻⁴
Monocyte%	7.4 ± 1.9	7. ± 2.1	3.4e ⁻¹
Platelet	265.7 ± 73.6	253.6 ± 74.5	1.7e ⁻⁴
C-reactive protein	4.8 ± 12.1	5.3 ± 11.3	3.1e ⁻⁵
Albumin	43.0 ± 4.2	41.2 ± 4.0	<1e ⁻⁶
Globulin	28.5 ± 4.6	28.5 ± 4.7	8.8e ⁻¹
Total bilirubin	14.9 ± 7.7	13.4 ± 6.6	3.1e ⁻⁶
Alanine aminotransferase	24.4 ± 22.5	24.6 ± 21.2	4.8e ⁻²
Alkaline phosphatase	84.8 ± 53.3	77.4 ± 26.6	6.7e ⁻¹
γ-glutamyl transferase	24.2 ± 28.3	26.3 ± 32.7	1.7e ⁻¹
Sodium	140.8 ± 2.8	141.2 ± 3.0	5.4e ⁻⁴
Potassium	3.9 ± 0.6	3.8 ± 0.4	3.5e ⁻³
Chlorine	104.9 ± 3.1	105.5 ± 3.2	1.1e ⁻⁶
NLR	2.7 ± 2.0	2.9 ± 2.2	1.3e ⁻⁵
MLR	0.3 ± 0.2	0.3 ± 0.2	1.9e ⁻²
PLR	148.4 ± 66.3	160.1 ± 98.0	4.9e ⁻⁴
SII	697.3 ± 526.0	737.1 ± 623.9	4.1e ⁻²

^aAll tests conducted in this study are two-sided tests.

records and 812 long-stay records, along with their corresponding 30-dimensional admission features (see Table 1). These features were categorized into behavioral variables, demographic variables, and admission blood test indicators. Additionally, based on prior research linking inflammatory factors with SCZ^{30,31}, we included calculated inflammatory markers in the blood test indicators. These markers include the neutrophil-to-lymphocyte ratio (NLR), monocyte-to-lymphocyte ratio (MLR), platelet-to-lymphocyte ratio (PLR), and Systemic immune-inflammation index (SII).

For these features, we performed statistical analysis, significance testing, and correlation analysis, as detailed in Table 1 and Extended Data Figs. 2–4. Continuous variables were analyzed using the Mann-Whitney U test, while categorical variables were assessed using the chi-square test for homogeneity. The results revealed significant univariate associations between admission age (p value < 1e⁻⁶), disease duration (p value < 1e⁻⁶), and

long-stay hospitalization in the demographic data, corroborating previous research findings³². Other significant variables, such as total bilirubin (p value = $3e^{-6}$) and lymphocytes (p value < $1e^{-6}$), also have research support for their associations with SCZ^{33,34}. Through correlation analysis, we observed that apart from variables with direct computational relationships, the remaining variables exhibited low correlations with each other (average Absolute Pearson Correlation Coefficient < 0.16). This indicates the diversity of perspectives offered by these features.

Optimizing LMs for confidential unstructured data extraction

Recent advancements in LMs such as ChatGPT³⁵ have demonstrated performance on par with state-of-the-art supervised learning methods across various natural language processing tasks, such as question answering³⁶, entity recognition³⁷, and relation extraction³⁸. These LMs offered significant advantages over traditional supervised methods, including the ability to perform few-shot learning or even zero-shot learning capabilities and enhanced interpretability³⁹. In our study, we aimed to identify behavioral variables, including aggression, self-destruction, and elopement, from unstructured text. The detailed definitions of the behavioral variables can be found in Extended Data Note. The selection of these three metrics was based not only on their association with the disease but also on the fact that, despite being unstructured, they had a low missing rate in our records and the information was generally quite clear. While other features, such as educational level, have also been shown to have a significant association with prolonged hospitalization in SCZ, these features were either unavailable in our EHRs or had a high missing rate. Given the unstructured format nature of EHRs data, employing LM-based methods presents a viable solution for addressing the unstructured text classification challenges.

However, due to stringent confidentiality policies, we could not directly apply state-of-the-art models like ChatGPT. Conversely, locally deployed models often suffered from performance constraints. To circumvent this, we proposed a prompt optimization method using LMs for unstructured information extraction, ensuring data remained fully localized. Detailed implementation steps are outlined in the Method section. In brief, we assessed the performance of local LMs on a small validation set and iteratively optimized the prompts using advanced LMs or manual adjustments based on evaluation results. The workflow and examples of optimized prompts are illustrated in Fig. 1a, b. evaluation results during the optimization process, depicted in Fig. 1c–e and Supplementary Table 1, showed that our iterative strategy enables the prompts to achieve within approximately five rounds on a validation set of 100 samples. Notably, Fig. 1d exhibited a performance fluctuation, attributed to GPT-4 occasionally following incorrect optimization paths due to the lack of specific data knowledge. Once this decline was identified, the optimization direction was promptly corrected.

To validate the superiority of our unstructured feature extraction method over supervised learning approaches, we manually labeled 40 records as a test set. The comparative results, shown in Fig. 1f–h and Supplementary Table 2, highlighted our method's performance against Linear Regression and fine-tuned BERT²⁷ on tasks including aggression recognition, self-destruction recognition, and elopement recognition. Specifically, both supervised learning models were pre-trained on the validation set before being tested on the labeled test set. Our method not only achieved the highest accuracy (average 94.2%) but also identified records where the LM could not provide a judgment. These few uncertain records can be manually annotated, ensuring high accuracy while significantly reducing the manual workload.

In summary, our information extraction method based on LMs utilizes iteratively optimized prompts to extract data while

maintaining data confidentiality. Our approach, validated through multiple perspectives, significantly reduced manual workload and achieved near-perfect accuracy in extracting behavioral variables, demonstrating the efficacy of LMs in unstructured information extraction.

Predicting long-term hospitalization risk in SCZ patients using integrated demographic, behavioral, and blood test indicators

We hypothesized that integrating demographic variables, behavioral variables, and blood test indicators would enhance the prediction of long-term hospitalization risk in SCZ patients. The reasons for selecting these as input features can be found in the Supplementary Note. The workflow of our model is depicted in Fig. 2a. Structured demographic variables and blood test indicators underwent preprocessing operations, including normalization and variable transformation. For unstructured data, we utilized an LM-based information extraction strategy. These features were consolidated into a 33-dimensional vector and fed into an MLP network architecture with attention mechanisms for feature extraction and binary classification of long and short hospitalizations. Comprehensive details on data preprocessing and the model process are available in the methods section.

During training, we applied leave-one-out cross-validation⁴⁰ to optimize our hyperparameters. Post-optimization, we evaluated our model's performance on an independent test set and benchmarked it against other machine learning methods. The independent test set, partitioned prior to training, comprised the most recent 50 long-term and short-term hospitalization patients from the dataset. This temporal segmentation strategy, grounded in the latest data, has been validated as both effective and reasonable⁴¹. Figure 2b illustrates the ROC curves of our model versus other machine learning methods on the independent test set, while Fig. 2c–g display the corresponding confusion matrices. Precision, recall, and other metrics derived from the confusion matrices are detailed in Supplementary Table 8. Our model achieved an AUC of 0.9 and an accuracy of 0.81, demonstrating its efficacy in predicting long-term hospitalization risk for SCZ patients based on multi-faceted admission variables. By analyzing the confusion matrices of each model, we observed that conventional machine learning models well in predicting accuracy for short-stay patients, their predictions for long-stay patients were close to random guessing. Conversely, our model effectively identified the characteristics of long-stay patients, demonstrating a marked improvement over traditional methods.

Comprehensive interpretability analysis revealed associations between admission variables and prolonged hospitalization for SCZ

We hypothesized that a comprehensive interpretability analysis of our black-box model would not only validate its effectiveness but also elucidate the admission risk factors associated with prolonged hospitalization for SCZ. Firstly, we conducted representation visualization using Principal Component Analysis to reduce the dimension of both input vectors and final hidden vectors to two dimensions, as shown in Fig. 3a, b. The projections of the original input vectors in the two-dimensional space did not form discernible clusters. In contrast, the final hidden vectors computed by the model exhibited distinct classification boundaries, indicating that our model effectively identified a hyperplane that separated the dataset.

To further dissect the significant features in our classification process, we employed SHapley Additive exPlanations (SHAP) for model interpretability analysis⁴². SHAP provides a unified approach to interpreting machine learning models by calculating the average marginal contribution of each feature value over all possible coalitions. Figure 3c and Extended Data Fig. 5–6 illustrate the SHAP analysis results on our test set. The colors in the figures

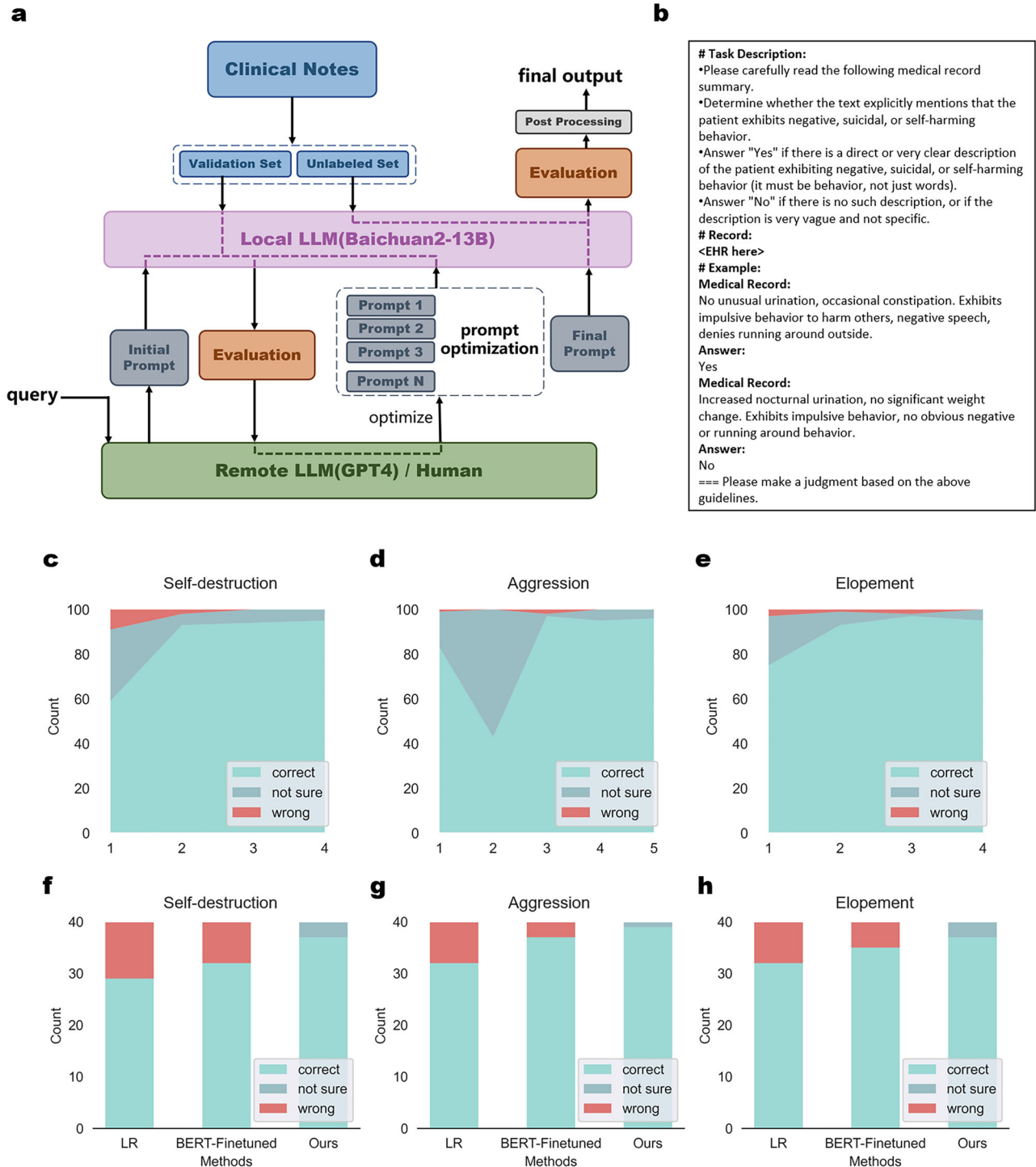


Fig. 1 Methodology and comprehensive evaluation of our local LM-based few-shot extraction framework. a Workflow of our few-shot unstructured information extraction method based on local LM. **b** Typical structure of the final optimized prompt in this study. **c–e** Accuracy on the validation set during prompt optimization. **f–h** The comparison of our methods with linear regression and BERT on an independent test set. Results show that our methods can achieve the lowest error rate.

represent the values of the corresponding features, and the x-axis denotes the SHAP values of the features. Key observations included: 1) Demographic Variables: Age of Admission (positively correlated), Disease Duration (positively correlated), and Marital Status (negatively correlated) emerged as three critical risk factors for prolonged hospitalization, aligning with previous studies^{20,26,43}; 2) Blood Test Indicators: Blood test indicators such as

inflammatory markers, Lymphocyte%, and Albumin were significant. Previous studies had linked these indicators with the severity of SCZ symptoms⁴⁴ and the effectiveness of medication⁴⁵, and this work is the first to associate them with prolonged hospitalization; 3) Behavioral Variables: Behavioral variables like elopement and self-destructive behaviors also correlated with extended hospital stays.

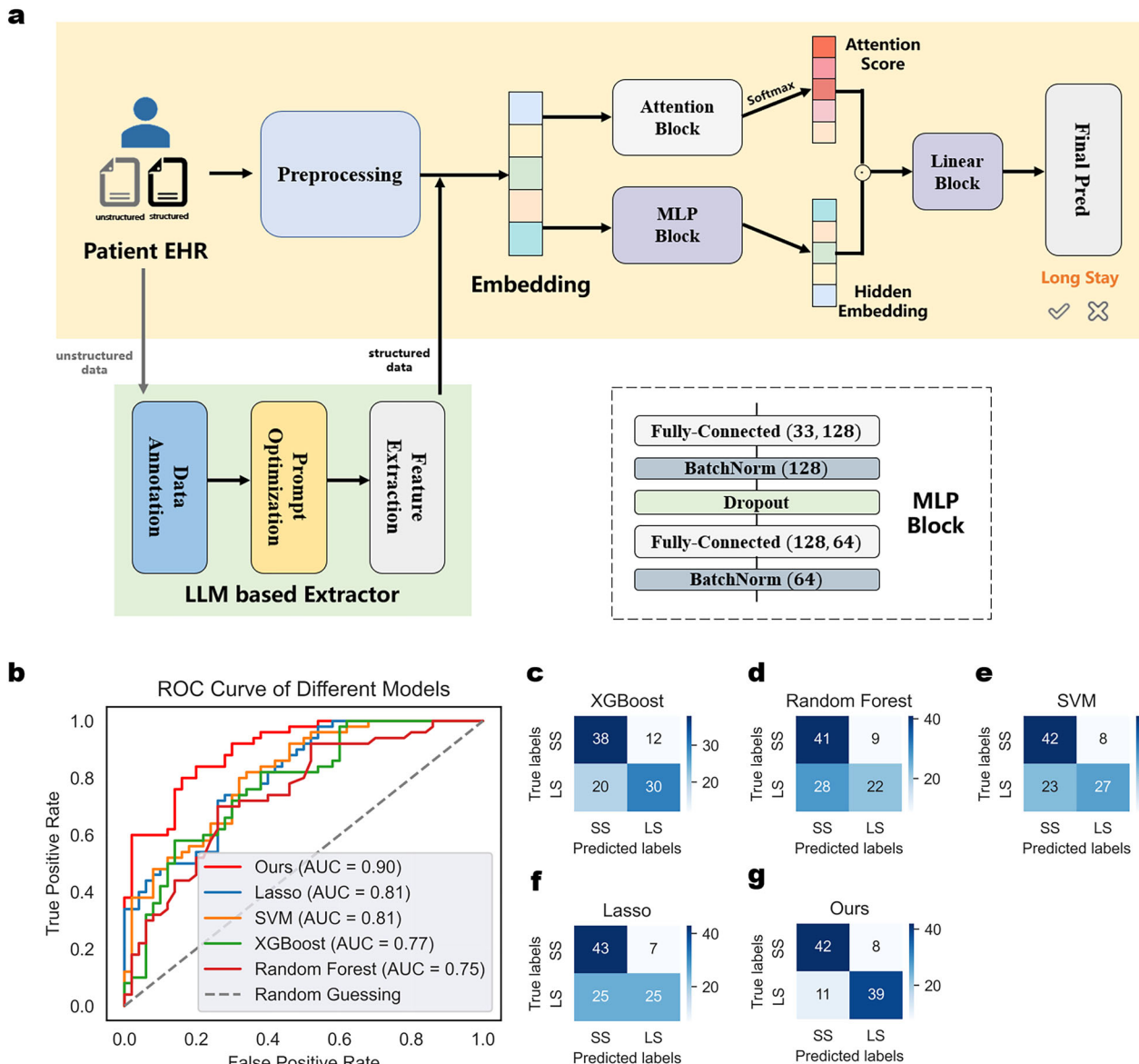


Fig. 2 Architectural design and evaluation of our long-term hospitalization risk prediction system. a Architecture of our long-term hospitalization risk prediction model. **b** ROC curve of our model and machine learning methods on the test set. **c–g** Confusion matrix of our model and machine learning methods on the test set.

Ablation study and case analysis revealed the effectiveness of multi-perspective variables in aiding long-term hospitalization risk prediction

To illustrate the effectiveness of using multi-perspective variables, we conducted ablation experiments by incrementally adding blood test indicators and behavioral variables to the demographic variables. The detailed procedure is described in the Methods section. Figure 4b–e and Supplementary Table 9 present the results of these experiments. Using demographic variables alone achieved an AUC close to 0.8, demonstrating their strong predictive foundation. Adding blood test indicators and behavioral variables significantly improved model performance to an AUC of 0.85 and 0.83, with the combined inputs of all three types of variables further enhancing the performance to match the original model with an AUC of 0.9. This indicated that while demographic variables provide a solid predictive base, the inclusion of blood test indicators and behavioral variables adds additional, valuable information, improving the prediction of

long-term hospitalization risk by an AUC of 6.25% and 3.75%. Furthermore, the examination of confusion matrices revealed that introducing blood test indicators significantly reduced the false positive rate by 31% (from 0.32 to 0.22). This is likely due to the direct health status information these indicators provided, which reduced the likelihood of misclassification. Additionally, we explored how our information extraction paradigm contributes to this research by replacing the behavioral variables extracted by our LM with those extracted using BERT. We observed a decline in model performance (from 0.901 to 0.882). The modest decrease may be attributed to the fact that behavioral variables do not dominate the predictive power in this context; however, the advantages of using our extraction paradigm will become more pronounced when applied to other dominant features in information extraction tasks.

To further illustrate the importance of multi-perspective variables in predicting long-term hospitalization from an individual perspective, we conducted a case study analyzing SHAP

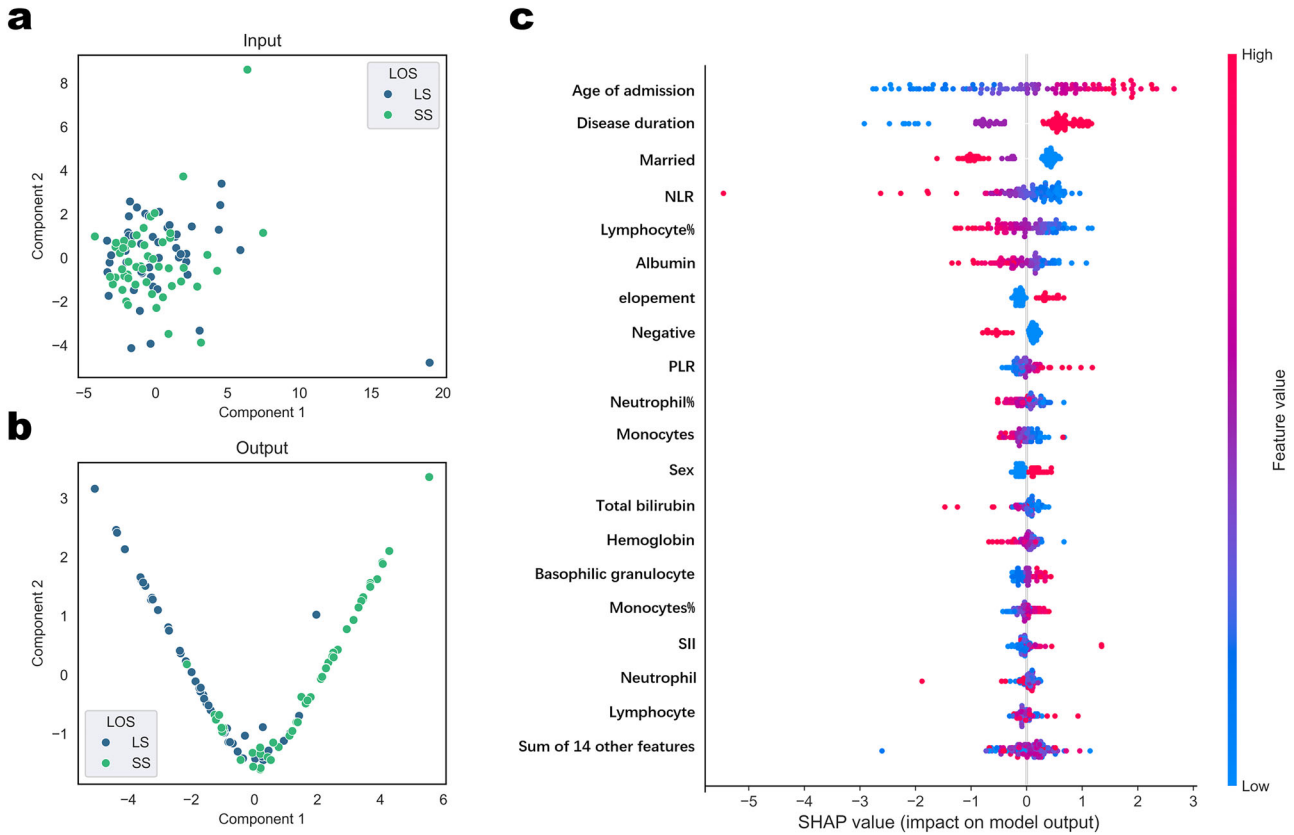


Fig. 3 Interpretability validation of our predictive model through complementary analytical lenses. a–b Interpretability analysis performed on test set using Principal Component Analysis. **c** Model interpretability analysis on test set using SHAP. The color bar corresponds to the raw values of the variables for each instance. In this beeswarm plot, points are distributed horizontally along the x axis according to their SHAP value.

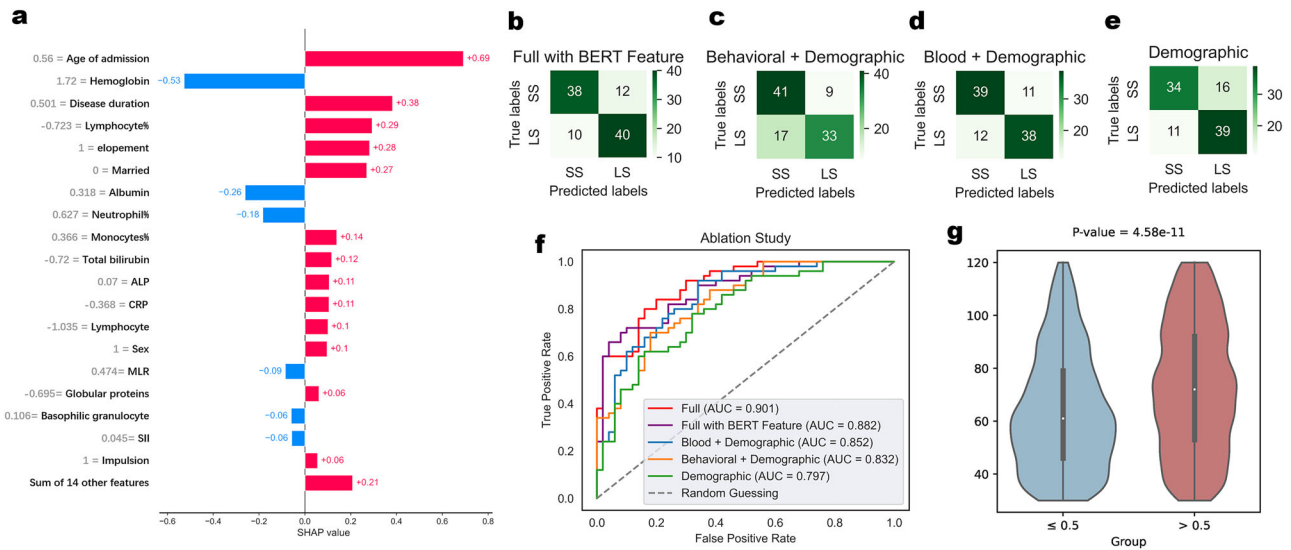


Fig. 4 The demonstration of prediction enhancement through multi-perspective variable integration using ablation analysis and clinical validation. a Interpretability analysis performed on a single patient using SHAP. **b–e** Confusion matrix of the models with BERT extracted behavioral features. **c–e** Confusion matrix of the models with different data variables were ablated. **f** ROC curve of the models with different data variables was ablated. **g** Patients with intermediate length of stay were grouped according to the model prediction. The significant differences shown in the figure illustrate the effectiveness of our model.

values for a specific patient—a 52-year-old male with a hospitalization duration of 161 days (Fig. 4a). Age and disease duration were found to have the most significant impact on predicting hospitalization duration, consistent with previous

studies^{46,47} that correlated older age with longer hospital stays. SCZ, as a severe chronic illness, often involves long disease durations, correlating with recovery difficulty and extended hospitalization periods. Reviewing the EHRs revealed poor

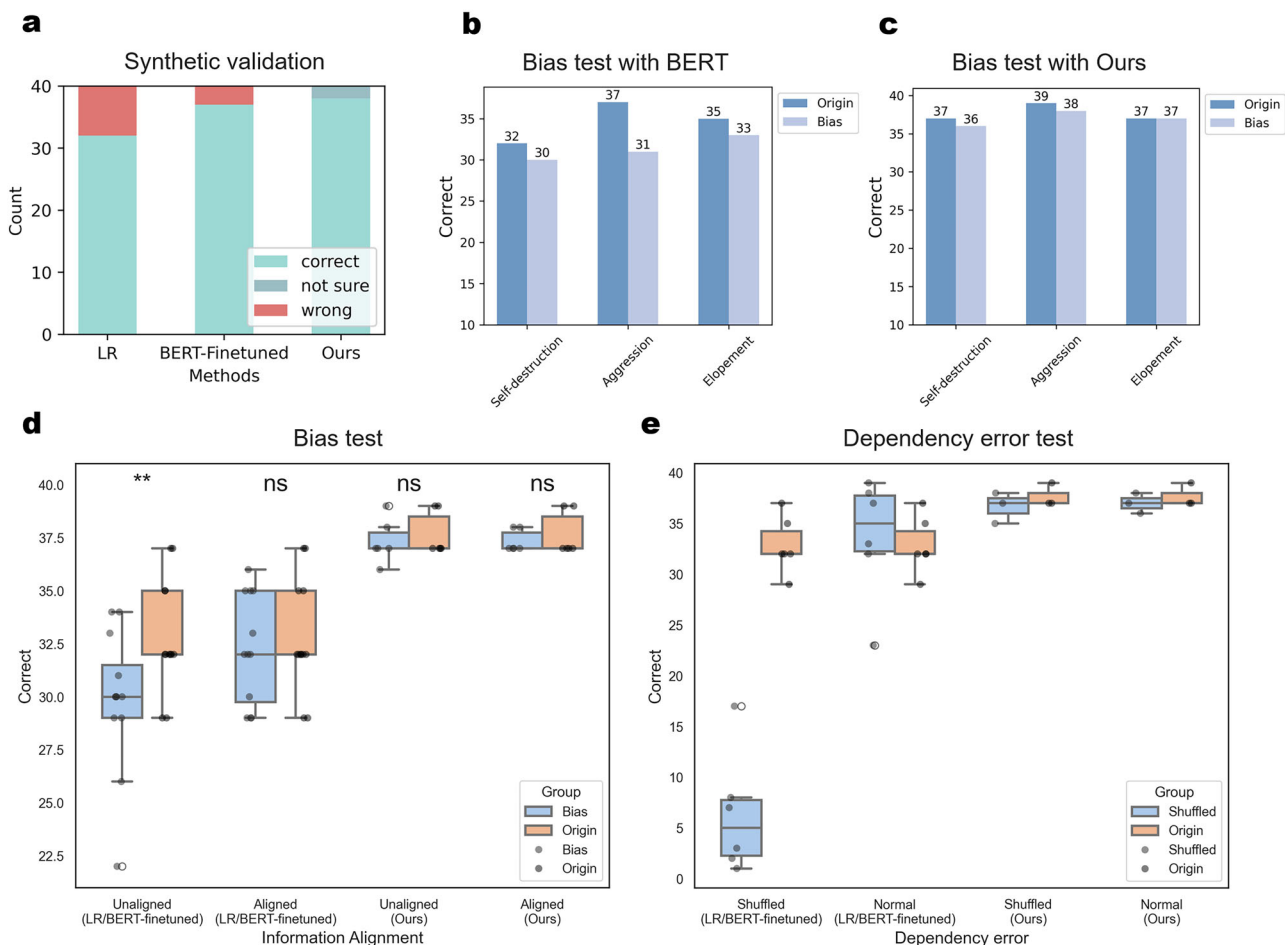


Fig. 5 Our model extraction framework demonstrates advantages in performance and bias mitigation. **a** The comparison of our methods with linear regression and BERT on the synthetic test set. Results show that our methods can achieve the lowest error rate. **b** BERT model performance on original and racial bias datasets. **c** Our model performance on original and racial bias datasets. **d** A study on the performance degradation of the model on racial bias and gender bias datasets shows a significant performance decline of the supervised model, as indicated by the Wilcoxon test ($p = 4.5e^{-3}$). **e** A study on the performance degradation of the model on dependency error datasets shows a significant performance decline of the supervised model.

medication adherence, strongly associated with condition severity and longer hospital stays. In addition, aggressive and elopement behaviors noted in the records indicated poor adherence. Blood test indicators related to blood cells, liver function, and electrolytes supported the prediction of prolonged hospitalization, as detailed in the Extended Data Note. This case study demonstrated that incorporating multi-perspective variables not only enhances the accuracy and comprehensiveness of the prediction but also allows for a multi-faceted analysis of patient conditions upon admission. This approach facilitated more efficient allocation of medical resources and the development of tailored treatment plans.

Our information extraction paradigm exhibited superiority in enhancing accuracy, reducing information bias, and minimizing erroneous information dependency

To demonstrate the superiority of our LM extraction paradigm, we conducted a multi-faceted study. First, to validate the effectiveness of our paradigm on external cohorts, we performed tests on an external synthetic dataset. The results, shown in Fig. 5a, indicate that our method maintains optimal accuracy, achieving a 95% correct rate. From another perspective, we assessed the performance of our approach in the presence of biases. We introduced two bias scenarios: racial bias and gender bias.

Detailed testing methodologies can be found in the Methods section, with results illustrated in Fig. 5b–d, Extended Data Figs. 8 and 9 and Supplementary Table 3–6. Notably, supervised learning methods exhibited a significant performance decline in the presence of bias (Wilcoxon test, $P = 4.5e^{-3}$), whereas our method remained unaffected. This indicates that our approach effectively mitigates potential unfairness related to racial and gender discrimination. Additionally, we considered another scenario involving erroneous information dependencies, where the model might focus on variables unrelated to our actual conclusions. For instance, when extracting whether a patient has a history of self-harm, it is inappropriate to infer this based on age or gender, even if strong correlations exist. We constructed such a scenario by inserting correlated information based on labels to influence the information extraction task. The results, presented in Fig. 5e, Extended Data Fig. 10 and Supplementary Table 7, show that supervised models were adversely affected by the misleading correlated information, incorrectly using age as a characteristic to determine self-harming behavior, while our method maintained performance comparable to the original task.

Correlation analysis in the population with intermediate LOS

In addition to predicting the hospitalization duration for long-stay and short-stay patients, we also aimed to capture the LOS

Table 2. Performance between different models in the population with intermediate length of stay.

Methods	Features	Correlation
-	Age of admission	0.149
-	NLR	0.032
Random forest	All	0.187 ± 0.023
SVM	All	0.195 ± 0.018
XgBoost	All	0.154 ± 0.037
Lasso	All	0.183 ± 0.018
Ours (Zero-shot)	All	0.213

characteristics of intermediate stays patients using correlation analysis. Specifically, a higher correlation between the predicted scores and the actual LOS indicates better predictive capability. As detailed in Table 2 and the Methods section, our model applied in a zero-shot manner⁴⁸ to intermediate-stay patient group showed a significant positive correlation of 0.213 (p value $< 1e^{-6}$) between the predicted probabilities and the actual hospitalization duration. Here the zero-shot manner refers to making predictions for the category of intermediate-stay patients without having previously encountered any data specific to this category. This result was compared with univariate analysis and machine learning-based multivariate analysis. The machine learning methods, trained using five-fold cross-validation on the intermediate-stay patients dataset, exhibited lower correlation, with an average Pearson Correlation Coefficient of 0.179, compared to our model's average Pearson Correlation Coefficient of 0.213, even in a zero-shot scenario. This suggested that our model effectively captured hospitalization duration characteristics for SCZ patients. Results shown in Fig. 4f further supported our finding by categorizing intermediate-stay patients into two groups based on the predicted scores. Patients with predicted scores greater than 0.5 had significantly longer hospitalization durations than those with scores less than 0.5, demonstrating the effectiveness of our model.

DISCUSSION

This study presented a novel and integrative approach for predicting long-term hospitalization risk in SCZ patients by leveraging demographic, behavioral, and blood test indicators using advanced deep learning techniques. Our findings demonstrate the significant potential of using a comprehensive set of admission variables to enhance prediction accuracy and provide deeper insights into the factors influencing prolonged hospitalization in SCZ patients.

One of the key contributions of our study is the application of LMs for extracting unstructured information from EHRs. This approach not only ensures high precision and privacy through local deployment but also significantly reduces the manual workload involved in information extraction. The iterative prompt optimization strategy we employed further enhances the model's accuracy, providing a robust method for extracting relevant behavioral variables with minimal manual intervention. This represents a significant advancement over traditional supervised learning methods. Additionally, we evaluated the effectiveness of our approach against supervised learning methods in two simulated tasks: latent discrimination and erroneous dependency. The results demonstrated that our method significantly reduces the occurrence of both latent discrimination and erroneous dependency issues. This characteristic is particularly crucial in the context of medical-related problems.

Our integrative model achieved a classification accuracy of 0.81 and an AUC of 0.9 on an independent test set, highlighting its efficacy in predicting long-term hospitalization risks. The inclusion

of diverse data types, such as demographic variables (e.g., age at admission, disease duration, marital status), blood test indicators, and behavioral data, proved crucial in enhancing the model's predictive power. In addition, defining the problem as a binary classification issue of short-term versus long-term hospitalization also helps our tool to some extent in achieving cross-institutional application. This is because other healthcare institutions can view the prediction results as indicators of the risk of prolonged hospitalization when applying our prediction model, rather than being constrained by the specific number of days of hospitalization. In all, this comprehensive approach allows for a more nuanced understanding of the various factors contributing to prolonged hospitalization, supporting better resource allocation and personalized treatment planning at the time of admission.

Our interpretability analysis using SHAP values⁴² identified several key risk factors associated with prolonged hospitalization. Age at admission and marital status emerged as significant predictors, consistent with previous findings^{18–20}. Behavioral features like self-destructive behaviors and elopement were also important indicators of patient adherence, corroborating earlier research⁴⁹. Additionally, blood test indicators like neutrophil-to-lymphocyte ratio (NLR), lymphocyte percentage, and albumin levels, which have been previously studied in relation to SCZ severity and treatment efficacy, were found to be significant predictors in our model^{50–52}. This comprehensive analysis not only validates the effectiveness of our model but also provides deeper insights into the factors influencing long-term hospitalization in SCZ patients.

Through our model's interpretability analysis, we identified several risk factors associated with prolonged hospitalization for SCZ patients. Some of these risk factors, such as age at admission and marital status, aligned well with established research findings^{18–20}. However, other risk factors we identified, such as NLR and lymphocyte, though not previously directly linked to the risk of prolonged hospitalization in schizophrenia, have been associated with the disease severity itself in other studies. For example, behavioral features like self-destructive tendencies and elopement may reflect patient adherence, a relationship corroborated by early research⁴⁹. Blood test indicators, such as NLR, lymphocyte, and albumin have also been previously explored in relation to SCZ from perspectives of treatment efficacy and disease severity^{50–52}. Given that LOS is closely tied to the severity of SCZ and the treatment effectiveness, our findings reinforce the relevance of these risk factors. Moreover, our study offers new insights into the relationships between key admission indicators and these critical aspects of SCZ management.

Despite the promising results, our study has a few potential limitations. The LM-based information extraction process struggles with long and complex text entries, potentially affecting the accurate capture of relevant behavioral variables. Besides, our dataset, sourced from a single mental health center, may limit the generalizability of our findings. Future studies should validate our model using more diverse, multi-center datasets to enhance its robustness. In addition, due to the limited personal information, our model could not account for confounding factors such as medication adherence, social support, and post-discharge care, which could significantly impact hospitalization outcomes. Future research should include these factors to provide a more comprehensive understanding of the predictors of long-term hospitalization in SCZ patients. Despite these limitations, our work provides a valuable new paradigm for mental health record processing and personalized medical guidance. By integrating multi-perspective admission variables, we have significantly enhanced the predictive accuracy and interpretability of long-term hospitalization risks in SCZ patients.

In conclusion, our research highlights the effectiveness of using an LM-based paradigm for processing mental health EHRs and incorporating multi-perspective admission variables for long-term hospitalization risk prediction in SCZ patients. This study provides a valuable new paradigm for mental health record processing and personalized medical guidance, contributing to more effective resource allocation and treatment planning at the admission stage.

METHODS

Data extraction and processing

Our retrospective dataset consists of EHRs of SCZ inpatients discharged from the SMHC between 2018 and 2022. SMHC is an acute care hospital, but it also provides long-term care. Patients were diagnosed according to the ICD-10 classification system, with disease codes ranging from F20 to F29. After admission, these patients were all assigned to the same type of ward, which means that ward assignments are not based on the severity of illness or expected LOS. Patients will not be required to be discharged due to the treatment duration being too long. The information dimensions include admission blood test data, demographic information, and current medical history records containing behavioral data. The use of this data was approved by the Ethics Committee of the Shanghai Mental Health Center (Approval No. 2022-79), and strict adherence to data confidentiality regulations was maintained throughout the data processing process.

The data processing and screening process is illustrated in Extended Data Fig. 1. Firstly, we excluded patients with a hospitalization duration of less than three days. This is because we considered that a hospitalization duration of less than three days is not typical for acute phase inpatient treatment of SCZ, as these patients may not have received sufficient treatment due to early discharge. Therefore, such a short hospitalization duration does not accurately reflect the treatment efficacy of the patients. Secondly, we excluded patients with outcomes other than cure or improvement upon discharge. These patients may have not yet met the clinical criteria for cure and require extended hospitalization for further treatment, but have been discharged prematurely for non-medical reasons.

The final step involved grouping the records based on hospitalization duration, and converting them into binary data representing long-stay and short-stay categories⁵³. A study conducted in 2019 indicated that the hospitalization duration for SCZ patients in China is 73 ± 42 days¹⁶. Building upon this, we defined records with a hospitalization duration exceeding 120 days as long-stay, and those with a duration shorter than 30 days as short-stay. This categorization provided the initial data required for our study.

Few-shot unstructured information extraction based on local LM

Figure 1a illustrates the workflow of our method. For the selection of local LM, as our dataset is in Chinese, we used Baichuan2-13B⁵⁴. This model has demonstrated performance that is not only close to GPT-3.5 Turbo on various medical question-answering datasets but also comprehensively outperforms LLaMA2⁵⁵ across these test sets. Detailed test results can be found at <https://github.com/baichuan-inc/Baichuan2>. We downloaded the version released on 2023-12-29 and deployed it on a local server equipped with three Nvidia GeForce 3090 GPUs. For the remote advanced LMs, we used GPT-4³⁵, the most advanced model available at the time of our study. Due to external network access restrictions on our local server, we accessed this model via the web interface (accessed on 2024-04-19). We also developed an API-based automated version for users with network access, providing a more automated prompt optimization experience.

Below, we describe our prompt optimization process. First, we annotated 100 examples for each of the three behavioral variables, ensuring a 1:1 ratio of positive to negative samples, meaning there were 50 entries with the corresponding feature and 50 without it. These datasets served as validation sets for adjusting the prompt, which is a hyperparameter for the local LM. Initially, we manually set an initial prompt and had the local LM generate judgments for each entry five times consecutively. If all five judgments were consistent, this was taken as the final judgment; if they were inconsistent, the final judgment was recorded as “not sure”. This process provided an evaluation of the performance of this prompt on the validation set, resulting in counts of correct, incorrect, and uncertain entries. Extended Data Figure 7 shows that this five-repetition paradigm exhibits significantly higher stability and a lower error rate than the paradigm without repetition. Next, we embedded the current prompt and its corresponding performance into a pre-designed query template and requested the remote LM to optimize the prompt. The optimized prompt then replaced the current prompt. This iterative process continued until no errors were produced on the validation set, and the performance of the current prompt ceased to improve after further optimization. An example of the prompt before and after optimization can be seen in Extended Data Figure 11. Specifically, considering the concentration of information in the text segments we needed to extract, we selected the sentences starting from the two clauses preceding the keyword “impulsion” to the end of the text segment as the model input.

Comparison of our LM-based method with supervised learning-based models

To demonstrate the advantages of our LM-based method over supervised learning-based models under the same conditions, we compared our method with Linear Regression and fine-tuned BERT. Specifically, we annotated 40 entries (20 positive and 20 negative) for each of the three behavioral variables as a test set. The previously annotated validation set of 100 entries was used for training the supervised learning-based models. Linear Regression was implemented using sklearn, and BERT employed the “bert-base-chinese”⁵⁶ tokenization strategy and pre-trained parameters. During training, we used a leave-one-out cross-validation strategy⁴⁰ for hyperparameter tuning.

Data preprocessing before deep neural network

Before feeding the data into the deep neural network, we followed conventional data preprocessing and encoding steps. For continuous data, we applied a standardization strategy as shown in Eq. (1). Specifically, for the duration of illness data, we employed a data binning strategy, dividing the duration into categories of less than 30 days, 30–600 days, and over 600 days, encoded as 0, 0.5, and 1, respectively. Other binary features, such as behavioral variables, were encoded as 0 or 1. Three-category features, such as marital status, were encoded as 0, 0.5, and 1. In order to avoid information leakage, we did standardization after data partitioning, and the standardization parameters of the test set followed the training set.

$$X_{norm} = \frac{X - X.mean}{X.std} \quad (1)$$

For data partitioning, we employed a time-sliced data-splitting approach. This method uses the most recent data in time as an independent test set to simulate the future predictive ability of our model and reduce interference from confounding factors in result evaluation. Here, we designated the last 50 long-term hospitalization records and the last 50 short-term hospitalization records, totaling 100 records, as the independent test set, while

the remaining data served as the training set. We did not separately partition a validation set because we utilized leave-one-out cross-validation for hyperparameter selection. This method involves iteratively leaving out one data point as validation data, and by exhaustively enumerating this validation data point, we ultimately obtain the proportion of correct predictions as the evaluation performance on the validation set. The main advantage of this method is its ability to fully utilize the data during training. To expedite computation efficiency, we set the sampling interval of leave-one-out cross-validation to 5, and conducted model training in parallel.

Prediction model implementation and performance evaluation

As showed in Fig. 2a, our model incorporates a self-attention mechanism with the MLP block. The MLP block can be represented as follows:

$$E^l = \sigma(E^{l-1}W^l + b^l) \quad (2)$$

where E^{l-1} represents the input vector, W^l is the trainable weight matrix, b^l is the bias term, and σ denotes the non-linear activation function. The attention block can be represented as follows:

$$Q = E \cdot W_Q, K = E \cdot W_K, V = E \cdot W_V \quad (3)$$

$$\text{Attention Scores} = \frac{Q \cdot K^T}{\sqrt{d_k}} \quad (4)$$

where E represents the input vector, W_Q , W_K , and W_V are trainable weight matrixes. d_k is the dimension of K . Finally, the vectors are weighted summed and fed into a fully connected layer for classification.

We implemented our model using the PyTorch⁵⁷ and executed it on a server equipped with three Nvidia GeForce 3090 GPUs. During model training, we employed techniques such as batch normalization⁵⁸ and dropout⁵⁹. Cross-entropy was utilized as the loss function, while Adam served as the optimizer⁶⁰, with an initial learning rate set to 0.0001 and weight decay to 1e-5. The batch size was set to 32 during training.

We evaluated our model on the independent test set using common machine learning evaluation metrics, as indicated in Supplementary Table 8. To further compare the superiority of our model over machine learning models, we implemented a series of machine learning models using sklearn. To ensure fairness, the training process of all machine learning models followed the same procedure as our model.

Interpretability analysis

We used sklearn to perform Principal Component Analysis dimensionality reduction and visualization analysis on the input vectors and final layer hidden vectors of the test set, using the default parameters. For the SHAP analysis, we used GradientExplainer to calculate the SHAP values. GradientExplainer leverages the gradients of the model output with respect to its input to efficiently approximate SHAP values, particularly suited for deep learning models.

Ablation study

To validate the effectiveness of multi-perspective admission features, we conducted feature ablation experiments. The data preprocessing methods, dataset partitioning methods, core model architecture, cross-validation methods, and model evaluation schemes used in these experiments were identical to those in the main experiment. The only difference was the change in the dimensionality of the input vectors, which correspondingly altered the input dimension of the first layer of the model architecture.

Testing on a synthetic external test set

Based on the work of Guevara et al.⁶¹, our synthetic external test set was generated using GPT-4o and reviewed by professional clinicians. The prompts used for generation can be found in the extended data note. To maintain consistency with previous tests, we generated a total of 140 patient history records, with 70 containing negative self-harm information and 70 without. In terms of dataset partitioning, 100 records were allocated for training and validation, while 40 records were reserved for independent testing. The ratio of positive to negative samples in all datasets was kept at 1:1.

Testing for bias and dependency errors

In our study, we incorporated three types of interference information by generating phrases to be added at the beginning of sentences. All phrases were generated by GPT-4o, with a requirement for diverse categorical expressions. For the dataset addressing racial bias, we created phrases describing Caucasian individuals, such as "A Caucasian patient presents for evaluation," and phrases for African American individuals, such as "The patient is an African American female." We conducted two types of tests: the first involved differing phrase types between the training and validation sets, where Caucasian descriptive phrases were included during training and African American phrases during testing. The second test maintained the same phrase types across both sets. This comparison allowed us to analyze the impact of racial bias on the performance of different models. A similar experimental approach was employed for gender bias. For the third type of dependency error testing, we added descriptions related to age, such as "The patient is an elderly female" and "A young patient presents for evaluation," based on the self-harm negative labels. In this case, our strategy for adding phrases at the beginning of sentences was reversed between the training and testing sets: during training, we added "elderly" phrases to sentences with self-harm indicators and "young" phrases to those without, while the testing set followed the opposite pattern. Clearly, an effective information extraction algorithm should remain unaffected by these irrelevant short phrases.

Intermediate-stay patients analysis

Intermediate-stay patients are those with hospitalization durations ranging from 30 to 120 days. After applying the same screening criteria as in the main experiment, a total of 3916 records were included. We chose to use the Spearman correlation coefficient to measure performance because it focuses more on the direct monotonic relationship between variables. For the machine learning models, we used their regression algorithm versions and trained them on the intermediate-stay patients' dataset using five-fold cross-validation for the regression task. All feature extraction and data preprocessing methods were consistent with those used in the main experiment.

DATA AVAILABILITY

Due to ethical restrictions, the data used in this study cannot be publicly shared. For further inquiries about the data, please contact the corresponding author. Data access can be granted under the following conditions: (1) all co-authors agree to share the data; (2) there is a data use agreement and ethical approval in place, following the data application requirements of the SMHC. Once shared, the data can only be used for the specified purpose.

CODE AVAILABILITY

Source code can be found at <https://github.com/RoarBoil/SchizoLOSPrediction>. The software versions can be found in the README of this GitHub repository.

Received: 27 November 2024; Accepted: 15 February 2025;
Published online: 05 March 2025

REFERENCES

- McCutcheon, R. A., Marques, T. R. & Howes, O. D. Schizophrenia—an overview. *JAMA Psychiatry* **77**, 201–210 (2020).
- McCutcheon, R. A., Keefe, R. S. & McGuire, P. K. Cognitive impairment in schizophrenia: aetiology, pathophysiology, and treatment. *Mol. Psychiatry* **28**, 1902–1918 (2023).
- Correll, C. U. et al. Mortality in people with schizophrenia: a systematic review and meta-analysis of relative risk and aggravating or attenuating factors. *World Psychiatry* **21**, 248–271 (2022).
- Teale, A.-L. et al. Delayed discharge in inpatient psychiatric care: a systematic review. *Int. J. Ment. Health Syst.* **18**, 14 (2024).
- Solmi, M. et al. Incidence, prevalence, and global burden of schizophrenia-data, with critical appraisal, from the Global Burden of Disease (GBD) 2019. *Mol. Psychiatry* **28**, 5319–5327 (2023).
- He, H. et al. Trends in the incidence and DALYs of schizophrenia at the global, regional and national levels: results from the Global Burden of Disease Study 2017. *Epidemiol. Psychiatr. Sci.* **29**, e91 (2020).
- Carranza Navarro, F. et al. Predictors of the length of stay of psychiatric inpatients: protocol for a systematic review and meta-analysis. *Syst. Rev.* **10**, 1–6 (2021).
- Besa, J. J. V. et al. Prevalence and predictors of prolonged length of stay among patients admitted under general internal medicine in a tertiary government hospital in Manila, Philippines: a retrospective cross-sectional study. *BMC Health Serv. Res.* **23**, 50 (2023).
- Ismail, H. Duration of hospital admission in severe mental illness: is longer better? *BJPsych Adv.* **27**, 213–218 (2021).
- Jones, S. et al. Association between delays to patient admission from the emergency department and all-cause 30-day mortality. *Emerg. Med. J.* **39**, 168–173 (2022).
- Stone, K. et al. A systematic review of the prediction of hospital length of stay: towards a unified framework. *PLOS Digit. Health* **1**, e0000017 (2022).
- Kilaru, A. S. et al. An accelerated hospital observation pathway to reduce length of stay for patients with COVID-19. *Am. J. Manag. Care* **28**, 262–268 (2022).
- La, E. M. et al. Increasing access to state psychiatric hospital beds: exploring supply-side solutions. *Psychiatr. Serv.* **67**, appips201400570 (2015).
- Aghajani, S. & Kargari, M. Determining factors influencing length of stay and predicting length of stay using data mining in the general surgery department. *Hospital Pract. Res.* **1**, 53–58 (2016).
- Hwang, T.-Y. et al. Analysis of involuntary admissions in Korea through the admission management information system. *Int. J. Law Psychiatry* **68**, 101542 (2020).
- Bian, Y. et al. The optimal length of hospitalization for functional recovery of schizophrenia patients, a real-world study in Chinese people. *Psychiatr. Q.* **90**, 661–670 (2019).
- Chen, S. et al. 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* **62**, 854–863 (2017).
- Newman, L. et al. Factors associated with length of stay in psychiatric inpatient services in London, UK. *Psychiatr. Q.* **89**, 33–43 (2018).
- Masters, G. A. et al. Factors associated with length of psychiatric hospitalization. *Compr. psychiatry* **55**, 681–687 (2014).
- Bruce, M. & Smith, J. Length of stay among multi-ethnic psychiatric inpatients in the United Kingdom. *Compr. Psychiatry* **102**, 152201 (2020).
- DiNapoli, E. A. et al. Predictors in geriatric psychiatry hospital length of stay. *Psychiatr. Q.* **86**, 243–251 (2015).
- Di Lorenzo, R. et al. Demographic and clinical characteristics of patients involuntarily hospitalized in an Italian psychiatric ward: a 1-year retrospective analysis. *Acta Biomed.* **89**, 17 (2018).
- Aydın, M. et al. Suicide attempts and related factors in schizophrenia patients. *Saudi Med. J.* **40**, 475 (2019).
- Goga, L. Y. & Marais, B. S. Schizophrenia and schizoaffective disorder: length of stay and associated factors. *South Afr. J. Psychiatry* **30**, 8 (2024).
- Kirchbner, J. et al. Factors and predictors of length of stay in offenders diagnosed with schizophrenia—a machine-learning-based approach. *BMC Psychiatry* **20**, 1–12 (2020).
- Cheng, P. et al. Factors related to the length of stay for patients with Schizophrenia: a retrospective study. *Front. Psychiatry* **12**, 818254 (2022).
- Devlin, J. et al. Bert: pre-training of deep bidirectional transformers for language understanding. *arXiv* <https://arxiv.org/abs/1810.04805> (2018).
- Bui, D. D. A. & Zeng-Treitler, Q. Learning regular expressions for clinical text classification. *J. Am. Med. Inform. Assoc.* **21**, 850–857 (2014).
- Han, T. S. et al. Evaluation of the association of length of stay in hospital and outcomes. *Int. J. Qual. Health Care* **34**, mzab160 (2021).
- Müller, N. Inflammation in schizophrenia: pathogenetic aspects and therapeutic considerations. *Schizophr. Bull.* **44**, 973–982 (2018).
- Kirkpatrick, B. & Miller, B. J. Inflammation and schizophrenia. *Schizophr. Bull.* **39**, 1174–1179 (2013).
- Gopalakrishna, G., Ithman, M. & Malwitz, K. Predictors of length of stay in a psychiatric hospital. *Int. J. Psychiatry Clin. Pract.* **19**, 238–244 (2015).
- Dornelles, E. P., Marques, J. G. & Ouakinin, S. Unconjugated bilirubin and schizophrenia: a systematic review. *CNS Spectr.* **24**, 577–588 (2019).
- Penedo, M. et al. The role of dopamine receptors in lymphocytes and their changes in schizophrenia. *Brain Behav. Immun. Health* **12**, 100199 (2021).
- Achiam, J. et al. Gpt-4 technical report. *arXiv* <https://arxiv.org/abs/2303.08774> (2023).
- Trivedi, H. et al. Interleaving retrieval with chain-of-thought reasoning for knowledge-intensive multi-step questions. *arXiv* <https://arxiv.org/abs/2212.10509> (2022).
- Wang, S. et al. Gpt-ner: named entity recognition via large language models. *arXiv* <https://arxiv.org/abs/2304.10428> (2023).
- Wan, Z. et al. Gpt-re: in-context learning for relation extraction using large language models. *arXiv* <https://arxiv.org/abs/2305.02105> (2023).
- Sun, X. et al. Text classification via large language models. *arXiv* <https://arxiv.org/abs/2305.08377> (2023).
- Vehtari, A., Gelman, A. & Gabry, J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* **27**, 1413–1432 (2017).
- Maher, M. C. et al. Predicting the mutational drivers of future SARS-CoV-2 variants of concern. *Sci. Transl. Med.* **14**, eabk3445 (2022).
- Lundberg, S. M. & Lee S.-I. A unified approach to interpreting model predictions. In: *Adv. Neural Inf. Process. Syst.* **30** (2017).
- Noohi, S. et al. Determinants of length of stay in a psychiatric ward: a retrospective chart review. *Psychiatr. Q.* **91**, 273–287 (2020).
- Yang, C. et al. Hematological and inflammatory markers in Han Chinese patients with drug-free schizophrenia: relationship with symptom severity. *Front. Immunol.* **15**, 1337103 (2024).
- Osimo, E. F. et al. Predicting treatment resistance from first-episode psychosis using routinely collected clinical information. *Nat. Ment. Health* **1**, 25–35 (2023).
- Broderick, C., Azizian, A. & Warburton, K. Length of stay for inpatient incompetent to stand trial patients: importance of clinical and demographic variables. *CNS Spectr.* **25**, 734–742 (2020).
- Xiao, Y. et al. Hospital spending and length of hospital stay for mental disorders in Hunan, China. *Heliyon* **9**, e14968 (2023).
- Wang, W. et al. A survey of zero-shot learning: settings, methods, and applications. *ACM Trans. Intell. Syst. Technol.* **10**, 1–37 (2019).
- Nyanyiwa, S., Peters, K. & Murphy, G. A scoping review: treatment attitudes and adherence for adults with schizophrenia. *J. Clin. Nurs.* **31**, 3060–3075 (2022).
- Labonté, C. et al. Complete blood count inflammatory markers in treatment-resistant schizophrenia: evidence of association between treatment responsiveness and levels of inflammation. *Psychiatry Res.* **308**, 114382 (2022).
- Özdin, S. & Böke, Ö. Neutrophil/lymphocyte, platelet/lymphocyte and monocyte/lymphocyte ratios in different stages of schizophrenia. *Psychiatry Res.* **271**, 131–135 (2019).
- Yin, X. Y. et al. Associations of decreased serum total protein, albumin, and globulin with depressive severity of schizophrenia. *Front. Psychiatry* **13**, 957671 (2022).
- Chrusciel, J. et al. The prediction of hospital length of stay using unstructured data. *BMC Med. Inform. Decis. Mak.* **21**, 351 (2021).
- Yang, A. et al. Baichuan 2: open large-scale language models. *arXiv* <https://arxiv.org/abs/2309.10305> (2023).
- Touvron, H. et al. Llama 2: open foundation and fine-tuned chat models. *arXiv* <https://arxiv.org/abs/2307.09288> (2023).
- Cui, Y. et al. Pre-training with whole word masking for Chinese bert. *IEEE/ACM Trans. Audio Speech Lang. Process.* **29**, 3504–3514 (2021).
- Paszke, A. et al. Pytorch: an imperative style, high-performance deep learning library. In: *Advances Neural information processing systems*. 32 (2019).
- Ioffe, S. and S. Szegedy. Batch normalization: accelerating deep network training by reducing internal covariate shift. In: *International conference on machine learning*. PMLR (2015).
- Srivastava, N. et al. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**, 1929–1958 (2014).
- Zhang, Z. Improved adam optimizer for deep neural networks. In: *2018 IEEE/ACM 26th international symposium on quality of service (IWQoS)*. IEEE (2018).
- Guevara, M. et al. Large language models to identify social determinants of health in electronic health records. *NPJ Digit. Med.* **7**, 6 (2024).

ACKNOWLEDGEMENTS

This work was supported by grants from STI 2030—Major Projects (no. 2022ZD0209100) and the Medical-Engineering Cross Foundation of Shanghai Jiao Tong University (nos. YG2022ZD026 and 23X010302269).

AUTHOR CONTRIBUTIONS

Y.B. and Wan.W. conceived the idea of this research and wrote the manuscript. Y.B. and Z.L. processed the data, implemented the predictor, and tuned the model. Wan.W., Wei.W., X.Z. conducted the mechanism analysis. S.Y. and G.N.L. supervised the research and reviewed the manuscript. All authors contributed to the article and approved the submitted version.

COMPETING INTERESTS

The authors declare no competing interests.

ETHICS

This retrospective study was approved by the Ethics Committee of Shanghai Mental Health Center (approval number 2022-79).

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41537-025-00585-2>.

Correspondence and requests for materials should be addressed to Shunying Yu or Guan Ning Lin.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025