Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf

AI-based personalized real-time risk prediction for behavioral management in psychiatric wards using multimodal data

Ri-Ra Kang [a], Yong-gyom Kim [a], Minseok Hong [b], Yong Min Ahn [b], KangYoon Lee [a,*]

[a] Department of Computer Engineering, Gachon University, Seoul 13120. Republic of Korea

[b] Department of Neuropsychiatry, Seoul National University Hospital, 101 Daehak-ro, Jongno-Gu, Seoul 03080, Republic of Korea

ARTICLE INFO

Keywords:
Suicide prevention
Behavioral risk prediction psychiatric wards
Temporal Fusion Transformer (TFT)
Time-series analysis
Machine learning
Wearable sensors

ABSTRACT

Background: Suicide is a major global health issue, with approximately 700,000 deaths annually (WHO). In psychiatric wards, managing harmful behaviors such as suicide, self-harm, and aggression is essential to ensure patient and staff safety. However, psychiatric wards in South Korea face challenges due to high patient-to-psychiatrist ratios and heavy workloads. Current models relying on demographic data struggle to provide real-time predictions. This study introduces the Temporal Fusion Transformer (TFT) model to address these limitations by integrating sensor, location, and clinical data for predicting harmful behaviors. The TFT model's advanced features, such as Variable Selection Networks and temporal attention mechanisms, make it particularly suitable for capturing complex time-series patterns and providing interpretable results in psychiatric settings.

Methods: Data from 145 patients across three hospitals were collected using wearable devices that tracked heart rate, movement, and location. The data were aggregated hourly, preprocessed to handle missing values, and standardized. A binary classification model using TFT was developed and evaluated with accuracy, recall, F1 score, and AUC. Bayesian optimization was employed for hyperparameter tuning, and 5-fold cross-validation was performed to ensure generalizability.

Results: The TFT model outperformed Multi-LSTM and Multi-GRU models, achieving 95.1% accuracy, 74.9% recall, an F1 score of 78.1, and an AUC of 0.863. The Variable Selection Network effectively identified key predictive factors, such as daily entropy and heart rate variability, improving interpretability. Incorporating location and biometric data enhanced prediction accuracy and enabled real-time risk assessments.

Conclusion: This study is the first to use the TFT model for predicting behavioral risks in psychiatric wards. The model's ability to integrate diverse data sources, prioritize cirtical variables, and capture temporal dependencies make it highly suitable for psychiatric environments. While the TFT model performed well, challenges remain with recall due to the limited dataset. Future research will focus on expanding datasets, optimizing variable selection, and standardizing data through a multimodal Common Data Model (CDM) to further improve performance and clinical utility.

1. Introduction

Suicide remains a critical global public health concern, with approximately 700,000 deaths recorded annually (WHO) [1]. In psychiatric wards, it is essential to prevent and manage harmful behaviors such as suicide, self-harm, and aggression to ensure the safety of both patients and healthcare professionals [2]. However, psychiatric wards in South Korea face significant challenges due to high patient-to-psychiatrist ratios and heavy workloads, impeding effective management of these issues [3,4]. International studies indicate that roughly 40 % of inpatient suicides occur within three days of admission, identifying previous suicide attempts, persistent affective symptoms, and a failure to notify staff as key predictive factors for suicide risk [5,6]. Moreover, a Norwegian study reported that approximately 51.9 % of acute psychiatric inpatients experience suicidal impulses [7]. These findings highlight the urgent need for robust risk assessment strategies that can predict and intervene in harmful behaviors at an early stage. In this context, leveraging artificial intelligence (AI) for predictive modeling and real-time data integration shows considerable potential to enhance suicide prevention strategies and patient safety in psychiatric settings

* Corresponding author.

E-mail addresses: rirakang@gachon.ac.kr (R.-R. Kang), gyom1204@gachon.ac.kr (Y.-g. Kim), garam21th@gmail.com (M. Hong), aym@snu.ac.kr (Y. Min Ahn), keylee@gachon.ac.kr (K. Lee).

https://doi.org/10.1016/j.ijmedinf.2025.105870
Received 31 October 2024; Received in revised form 31 January 2025; Accepted 5 March 2025
Available online 12 March 2025

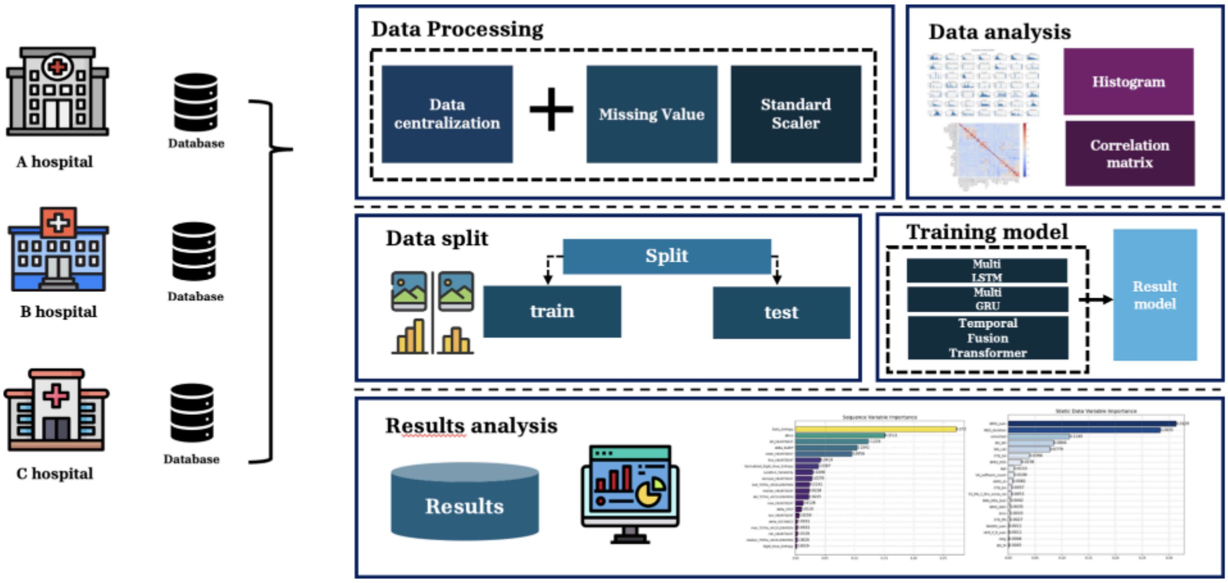

Fig. 1. Psychiatric Prediction flow.

[8–10]. Despite this promise, existing predictive models often rely heavily on demographic and basic clinical data, limiting real-time intervention capabilities within the ward environment [11–15]. To address these limitations, a comprehensive approach that integrates multiple data sources—such as wearable sensor data, location information, and demographic data—is critical [16–19]. In this study, we seek to overcome the shortcomings of the previously utilized Multi-LSTM model by developing a binary classification model based on the latest Temporal Fusion Transformer (TFT). While the Multi-LSTM model faced significant challenges in extracting variable importance after predictions [20], the TFT model capitalizes on advanced features—such as Variable Selection Networks, sequence-to-sequence layers, and temporal self-attention—to capture complex time-series patterns and effectively pinpoint key predictive factors [21,22]. By focusing on the most relevant variables and minimizing noise, the TFT model facilitates more accurate and efficient predictions, particularly in psychiatric wards where extensive biosignal data are collected. Although this research addresses systemic issues prevalent in Korean psychiatric wards—such as resource constraints and heavy workloads—its contributions have broader implications. The methodology developed here can be adapted to various cultural and clinical contexts with differing healthcare infrastructures and resource availability. Furthermore, insights derived from integrating biosignals, location data, and demographic information lay a foundational framework for implementing personalized risk prediction systems worldwide. In summary, this study's primary objective is to determine whether the TFT model can offer superior predictive accuracy and interpretability compared to the existing Multi-LSTM model for identifying harmful behaviors in psychiatric ward settings.

2. Methods

This study developed a risk prediction model for harmful behaviors among psychiatric inpatients by integrating data from four wards across three hospitals, as illustrated in Fig. 1, which shows the flow of creating the prediction model. Participants with missing data exceeding 50 % of

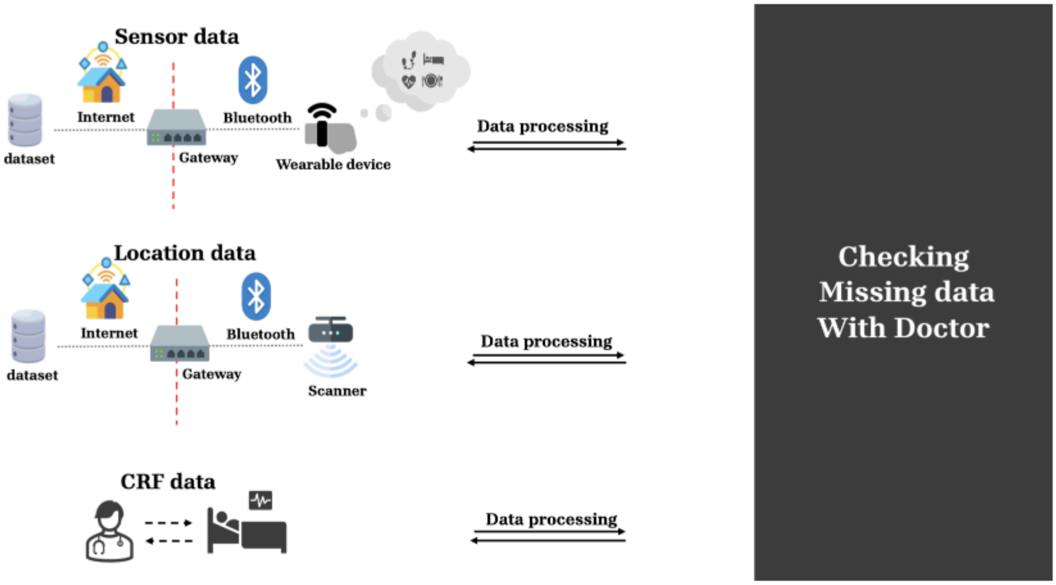

Fig. 2. Data preparation and analysis procedure.

the total time range were excluded, and the remaining missing values were imputed using the softImpute method. softImpute is an iterative matrix completion technique that approximates missing values by fitting a low-rank matrix and applying a nuclear norm penalty to reduce overfitting, effectively preserving the underlying data structure and making it suitable for large-scale datasets [23]. After standardizing the data, which transforms features to have a mean of zero and a standard deviation of one to remove scale differences and improve model convergence and comparability, histograms and correlation matrices were employed to explore key features. The dataset comprised 145 patients, who were randomly sampled into training and testing sets using an 8:2 ratio. This dataset was then used to evaluate the performance of Multi-LSTM, Multi-GRU, and Temporal Fusion Transformer models. Finally, the results were analyzed to assess model performance, highlighting the most significant predictive features and ensuring interpretability.

2.1. *Data collection*

This study collected data from four psychiatric wards across three hospitals: Seoul National University Hospital, Yongin Mental Hospital, and Dongguk University Hospital. All participants provided informed consent prior to the study and were compensated 100,000 KRW (approximately USD 70) per week during the study period. They were instructed to continuously wear the device except when showering or leaving the ward, but were free to remove it at any time if they wished. A total of 145 patients from three hospitals provided consent to participate in the study were instructed to continuously wear a wearable device (URBAN HR, Partron Co., Ltd.), except during showers, tests, or outings. The device monitored heart rate, three-axis acceleration, and location data to calculate metrics such as calories burned, sleep index, steps, and distance moved. Data were transmitted in real-time to hospital servers via BLE gateways. Health Connect Co., Ltd. developed the system for data collection and precise movement analysis, integrating biometric and location data to enhance patient management. Weekly clinical assessments and additional evaluations after interventions were conducted, and we adjusted for missing and outlier data to ensure accuracy with researchers. Adjustments for outliers are described in detail in the Appendix.

2.2. *Data Definition*

Sensor data, location data, and patient information data are collectively defined as multimodal data, which were utilized to develop the model. (See Fig. 2)

2.2.1. *Location data*

In this study, various methods were employed to process location data efficiently for model application. Instead of using simple latitude and longitude values, location data were transformed into more meaningful features for model input [24].

1. Location Entropy: This metric measures the diversity of a patient's visited locations. Higher entropy suggests a wider range of places visited. It is calculated daily from 24-hour location data and also at 8-hour intervals (midnight, 8 AM, 4 PM) for parallel model input.
2. Normalized Location Entropy: Represents location entropy normalized to a specific range. Calculated daily and at 8-hour intervals (midnight, 8 AM, 4 PM) from 24-hour data and used for model input.
3. Longitude and Latitude Variability: Measures changes in patient locations using a sliding window approach, indicating the variability in longitude and latitude over time.
4. Semantic Location: Identifies the most frequent location within a sliding window, highlighting where the patient spends most of their time, crucial for understanding movement patterns and input into the model.

Table 1
Summary of wearable sensor data variables.

Data Type	Variable name	Description
Acceleration Data	First_Total_Acceleration	First recorded total acceleration value
	Last_Total_Acceleration	Last recorded total acceleration value
	Mean_Total_Acceleration	Mean total acceleration value over the entire period
	Max_Total_Acceleration	Maximum total acceleration value over the entire period
	Min_Total_Acceleration	Minimum total acceleration value over the entire period
	Std_Total_Acceleration	Standard deviation of total acceleration
	Nunique_Total_Acceleration	Number of unique total acceleration values
	Median_Total_Acceleration	Median total acceleration value over the entire period
Heartbeat Data	First_Heartbeat	First recorded heartbeat value
	Last_Heartbeat	Last recorded heartbeat value
	Mean_Heartbeat	Mean heartbeat value over the entire period
	Median_Heartbeat	Median heartbeat value over the entire period
	Max_Heartbeat	Maximum heartbeat value over the entire period
	Min_Heartbeat	Minimum heartbeat value over the entire period
	Std_Heatbeat	Standard deviation of heartbeat
	Nunique_Heartbeat	Number of unique heartbeat values
Other Sensor Data	Delta_Distance	Change in distance over time intervals
	Delta_Sleep	Change in sleep patterns over time intervals
	Delta_Step	Change in step count over time intervals
	Delta_calories	Change in calorie consumption over time intervals

2.2.2. *Sensor data*

In this study, wearable sensor data were processed in various ways and applied to a risk predictions for harmful behaviors model. Through this processing, it was possible to analyze which variables played the most crucial role in the predictions [25]. The primary sensor data used and their processing methods are shown in Table 1. The angular velocity data stored in the wearable devices were excluded due to excessive missing values. Additionally, the battery data from the wearable devices were deleted as they were not related to the patients' biometric data.

2.2.3. *CRF data*

In this study, several data were used to develop a risk predictions for harmful behaviors model. These data reflect the patients' demographic information and psychological state, and include the variables shown in Table 2.

2.3. *Data preprocessing*

The data was collected in seconds, while the occurrences of self-harm and aggression were measured on a daily basis. Due to the differing and irregular timestamps of the sensor and location data, we aggregated the data into hourly intervals. To address this irregularity, each second-level data point was integrated based on the mode of the hourly data [26,27]. This allowed us to analyze the changes in each patient's vital sign data and location data on an hourly basis throughout the day. This procedure is illustrated in Fig. 3. Additionally, to maintain data consistency required for model training, all static variables were converted to numeric data, and missing values were filled with 0 to establish a pattern, considering the possibility of data not being collected due to the

Table 2
Summary of patient data variables.

Variable name	Description
MED AP	Medication usage (Anxiety medication)
MED AD	Medication usage (Antidepressants)
MED MS	Medication usage (Mood stabilizers)
MED SH	Medication usage (Sleep aids)
BPRS AFF	Brief Psychiatric Rating Scale (Affective symptoms)
BPRS POS	Brief Psychiatric Rating Scale (Positive symptoms)
BPRS_NEG	Brief Psychiatric Rating Scale (Negative symptoms)
BPRS_RES	Brief Psychiatric Rating Scale (Resistance symptoms)
BPRS_sum	Brief Psychiatric Rating Scale (Total score)
YMRS_sum	Young Mania Rating Scale (Total score)
MADRS_sum	Montgomery-˚Asberg Depression Rating Scale (Total score)
HAMA_sum	Hamilton Anxiety Rating Scale (Total score)
CS status 1 score cal	Severity of suicidal ideation
CS status 2 bhv score cal	Severity of suicidal behavior
CS status 2 NSSI	Non-suicidal self-injury score
Age	Patient's_Age
Sex	Patient's Sex
Occu	Patient's Occupational status
Relig	Patient's Religious affiliation
Insurance	Insurance status
Ho_inc_month	Monthly household income
convicted	Legal issues (convicted)
PH_no_physical	Physical health issues
PH tx status	Current treatment status
MED onmed	Currently on medication
MED duration	Duration of medication use
MINI MDx text	MINI psychiatric diagnosis
DIG text	Diagnostic test text
DIG cat	Diagnostic test category
VH selfharm	Visualization of self-harm
VH selfharm count	Number of self-harm visualizations
crime	Criminal record
CS life 1 score cal	Life score (1st assessment)
CS life 2 NSSI	Non-suicidal self-injury score (2nd assessment)
CS life 2 bhv score cal	Behavioral score (2nd assessment)
HCR P H sum	HCR-20 Past (H) section score
HCR P C sum	HCR-20 Clinical (C) section score
HCR P R sum	HCR-20 Risk (R) section score
HCR P sum	HCR-20 Total score
CTQ EA	Childhood Trauma Questionnaire (Emotional abuse)
CTQ PA	Childhood Trauma Questionnaire (Physical abuse)
CTQ SA	Childhood Trauma Questionnaire (Sexual abuse)
CTQ EN	Childhood Trauma Questionnaire (Emotional neglect)
CTQ PN	Childhood Trauma Questionnaire (Physical neglect)
CTQ MD	Childhood Trauma Questionnaire (Neglect score)
BIS M	Behavioral Inhibition System (Motor)
BIS NP	Behavioral Inhibition System (Non-motor)
BIS sum	Behavioral Inhibition System (Total score)
ASRS IA	Adult ADHD Self-Report Scale (Inattention)
ASRS HM	Adult ADHD Self-Report Scale (Hyperactivity-Impulsivity)
ASRS HV	Adult ADHD Self-Report Scale (Hyperactivity-Impulsivity & Antisocial behavior)
ASRS sum	Adult ADHD Self-Report Scale (Total score)
AUDIT sum	Alcohol Use Disorders Identification Test (Total score)

patient's risky behaviors. After the data integration, normalization was performed.

2.4. Data Splitting

In this study, the data were preprocessed and split into training and validation sets to train the risk predictions for harmful behaviors model. The Suicide Dataset class was defined by inheriting from the Dataset class of PyTorch [28]. This class takes data in the form of a Data Frame and converts the static variables, sequence variables, and target variables into tensors that can be input into the model. Employing this defined class, dataset instances were created and then split into training and validation sets. 80 % of the dataset was used for training, and the remaining 20 % was used for validation

2.5. Software

In this study, the PyTorch software framework was utilized to build a risk predictions for harmful behaviors model. PyTorch is a widely used deep learning library that provides a robust and flexible environment for constructing neural networks and handling various data operations [28]. The model was developed using Python version 3.9.17, ensuring compatibility and stability throughout the data processing and model training phases.

2.6. Implementation of the TFT model

In this study, a classification model based on the core concepts of the TFT model proposed by Lim et al. (2019) was implemented. Although the TFT model was not fully reproduced, several key elements were applied and modified [21]. Fig. 4. illustrates the structure of the TFT model.

The Temporal Fusion Transformer (TFT) is highly suitable for processing various types of time-series data such as sensor data, location data, and patient information that we are using. This model leverages the self-attention mechanism to learn both complex patterns and long- and short-term dependencies. It effectively handles multivariate time-series data, allowing it to better predict interactions between different variables. Additionally, the model offers interpretability, enabling us to analyze which variables are most important in making predictions, particularly for critical data like patient information. With its ability to process data in parallel, TFT can efficiently handle large volumes of sensor and location data, while maintaining strong performance even with long sequences of data.

2.6.1. Variable selection Networks (VSNs)

VSNs learn and select the importance of each static and time-series variable, allowing the model to focus on the most relevant features at each time step. The output of the VSN is used as the importance weight for each variable. This adaptive selection process ensures that the model dynamically prioritizes features based on the patient's unique behavioral patterns, enabling tailored predictions and effective interventions. The formula for VSN is as follows, where W_1 and W_2 are learnable weight matrices.

$$(VSN(x) = Softmax(ReLU(W_1 x) W_2)) \tag{1}$$

2.6.2. LSTM

Long Short-Term Memory (LSTM) layers were used to model the sequences of time-series data, helping to capture temporal dependencies in the data. The formula for LSTM is as follows where h_t and c_t represent the hidden state and cell state at the current time step, respectively.

$$(h_t, c_t) = LSTM(x_t, (h_{t-1}, c_{t-1})) \tag{2}$$

2.6.2.1. Multi-Head attention mechanism. The multi-head attention mechanism was introduced to model the temporal dependencies of the time-series data, allowing the model to focus on different parts of the input sequence simultaneously. The formula for the attention mechanism is as follows where Q, K, and V are the query, key, and value matrices, respectively, and dk is the dimension of the key vectors.

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{3}$$

2.6.2.2. Static covariate encoder. A static covariate encoder was used to embed the static features and combine them with the time-series data, ensuring that the static information is effectively reflected in the temporal dynamics. The formula for the static covariate encoder is as follows where W_S and b_s are learnable weights and biases.

$$Static_{Encoder(x)} = RELU(x + b_s) \tag{4}$$

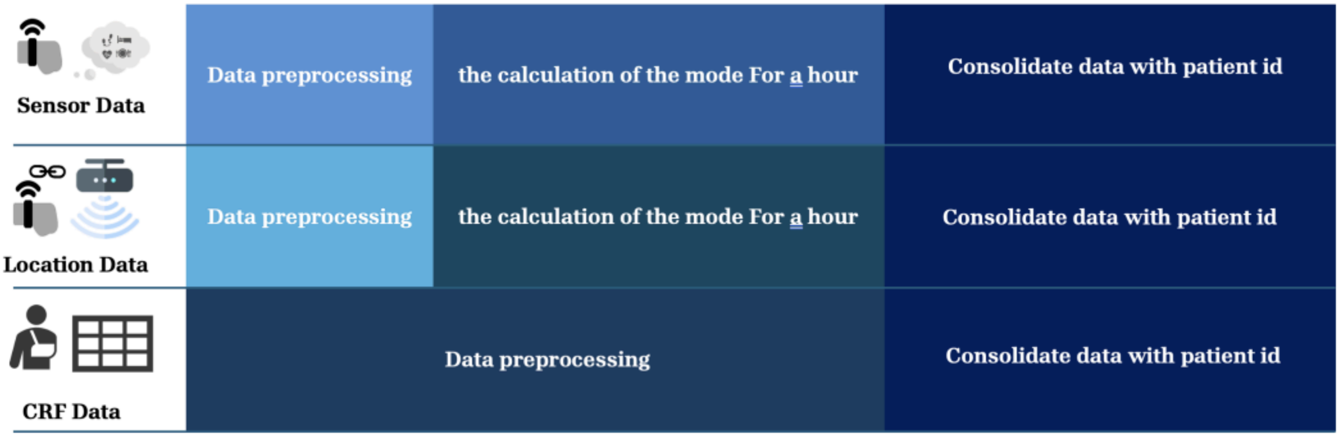
Sensor Data
Data preprocessing
the calculation of the mode For a hour
Consolidate data with patient id
Location Data
Data preprocessing
the calculation of the mode For a hour
Consolidate data with patient id
CRF Data
Data preprocessing
Consolidate data with patient id

Fig. 3. The figure shows preprocessing and consolidation of sensor, location, and CRF data.

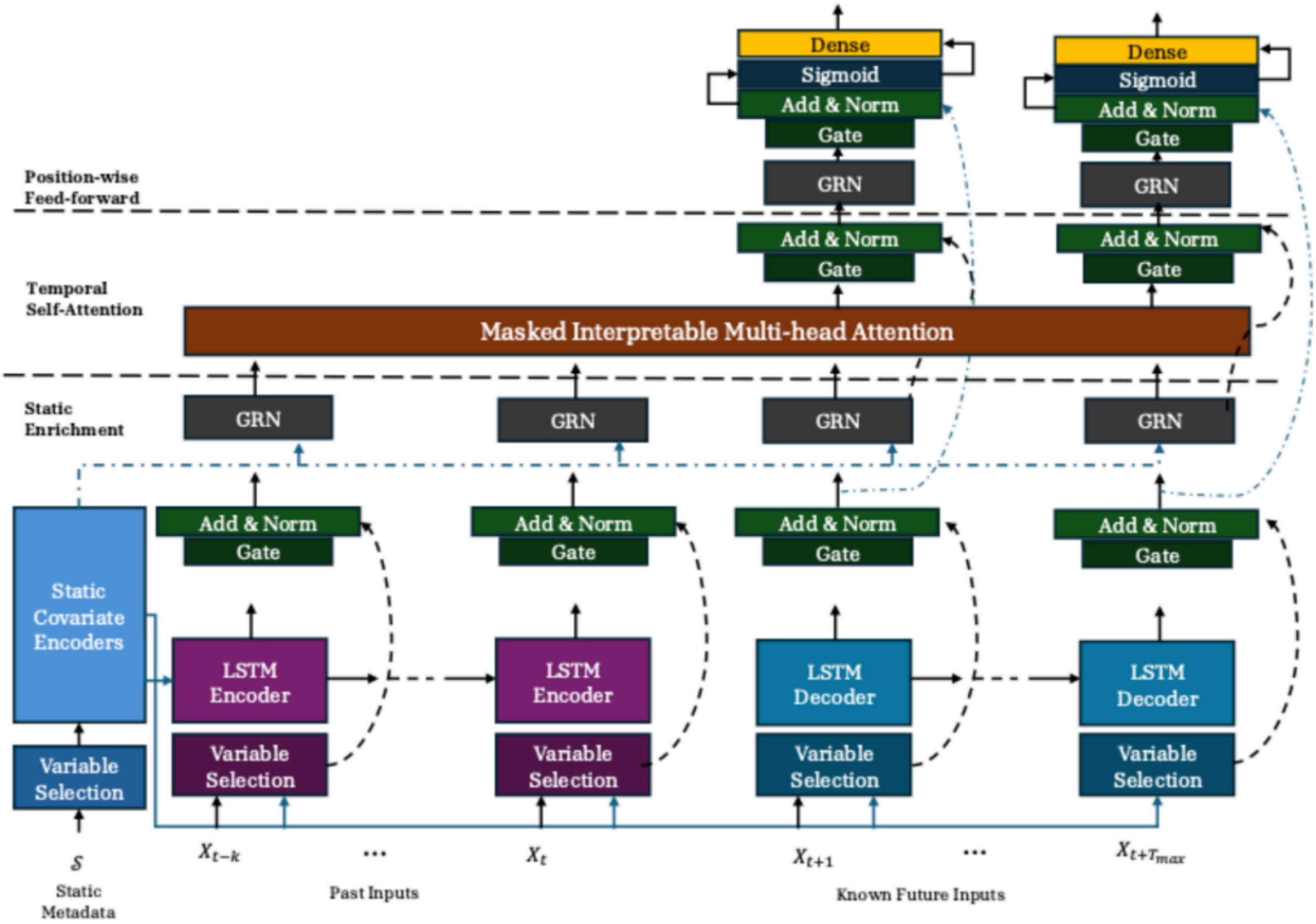

Fig. 4. Temporal Fusion Transformers for Interpretable Multi-horizon Time-Series Forecasting.

2.6.2.3. Sigmoid layer for classification model. A sigmoid activation function was used in the final output layer to compute the predicted probabilities for the binary classification of harmful behaviors risk. The formula for the sigmoid function is as follows.

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{5}$$

2.7. Hyperparameter optimization with Bayesian optimization

Bayesian optimization was used to optimize the performance of the Temporal Fusion Transformer (TFT) model and the hyperparameters [29]. Bayesian Optimization is a powerful method that efficiently explores the hyperparameter space to find the optimal values more quickly [30]. This method converges to the optimal hyperparameters faster than grid search or random search. In this study, hyperparameters such as hidden size, dropout rate, minibatch size, learning rate, maximum gradient norm, and num heads were tuned [31].

Table 3
Hyperparameter range.

Hyperparameter	range
hidden size	(10, 320)
dropout	(0.1, 0.9)
minibatch size	(10, 256)
learning rate	(1e-4, 1e-2)
max gradient norm	(0.01, 200)
num heads	(1, 8)

Table 4
Optional Hyperparmeter.

Hyperparameter	value
hidden size	67
dropout	0.537
minibatch size	241
learning rate	0.0096
max gradient norm	155.02
num heads	7.26

- Hyperparameters and Ranges: The following table shows the range of values considered for each hyperparameter during the optimization process. (See Table 3.).

The optimal combination of hyperparameters found through Bayesian Optimization is shown in Table 4..

3. Results

3.1. Evaluation metrics

The performance of the model was evaluated via the following metrics [32–34]. We evaluated the model's performance using several key metrics, including accuracy, F1 score, Area Under the Curve (AUC), sensitivity, specificity, precision, and recall. These metrics provide a comprehensive assessment of the model's ability to accurately predict risk of harmful behaviors, measure its precision and recall in identifying true positives, and determine its overall robustness and reliability in clinical settings.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{6}$$

$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{7}$$

$$\text{AUC} = \int_0^1 \text{TPR}\, d(\text{FPR}) \tag{8}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{9}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TP} + \text{FN}} \tag{10}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{11}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{12}$$

3.2. Model performance

Using the optimized hyperparameters, the model was tested and a very high performance was recorded. The evaluation metrics are as shown in table Table 5. As seen in the table above, the performance of models using different algorithms demonstrates excellence in various evaluation metrics. Notably, the TFT model recorded the highest accuracy and F1 score overall, along with an outstanding AUC score. This indicates that the TFT model outperforms other models in terms of overall performance. The superior performance of the TFT model is especially highlighted by its higher scores in accuracy, recall, F1 score, and AUC compared to other models.

3.3. K-Fold cross validation

K-fold cross-validation assesses a model's generalization performance by dividing data into K folds, where each fold serves as a test set while the remaining folds are used for training. This process, repeated K times, provides an average performance measure and helps prevent overfitting [35,36]. In this study, 5-fold cross-validation was employed, measuring accuracy, precision, recall, F1 score, AUC, sensitivity, and specificity for each fold. The model achieved an average accuracy of 0.9322 (±0.0086), precision of 0.7841 (±0.0821), recall of 0.5852 (±0.0679), F1 score of 0.6623 (±0.0192), AUC of 0.7807 (±0.0267), sensitivity of 0.5852 (±0.0679), and specificity of 0.9762 (±0.0154). These results demonstrate the model's strong performance in predicting risk of harmful behaviors, particularly its high specificity in correctly identifying patients without risk, alongside adequate sensitivity for detecting patients at risk, as summarized in Table 6. However, the results also reveal a trade-off between precision and recall, where higher precision is often accompanied by lower recall. This can be attributed to how strictly or leniently the model predicts the positive class. For

Table 5
Model Performance Comparison.

Model	Accuracy	Precision	Recall	F1 Score	AUC	Sensitivity	Specificity
TFT	0.9513	0.8155	0.7486	0.7806	0.8632	0.7486	0.9778
Multi-LSTM	0.9317	0.9663	0.4503	0.6143	0.7241	0.4503	0.9978
Multi-GRU	0.9311	0.8254	0.5445	0.6562	0.7643	0.5445	0.9841

Table 6
5-Fold cross-validation.

Fold	Accuracy	Precision	Recall	F1 Score	AUC	Sensitivity	Specificity
Fold 1	0.9292	0.7152	0.6278	0.6686	0.7978	0.6278	0.9679
Fold 2	0.9165	0.6618	0.6818	0.6716	0.8160	0.6818	0.9501
Fold 3	0.9380	0.8134	0.5989	0.6899	0.7905	0.5989	0.9821
Fold 4	0.9380	0.8600	0.5059	0.6370	0.7480	0.5059	0.9901
Fold 5	0.9392	0.8700	0.5118	0.6444	0.7513	0.5118	0.9908
Average	0.9322	0.7841	0.5852	0.6623	0.7807	0.5852	0.9762
Standard Deviation	±0.0086	±0.0821	±0.0679	±0.0192	±0.0267	±0.0679	±0.0154

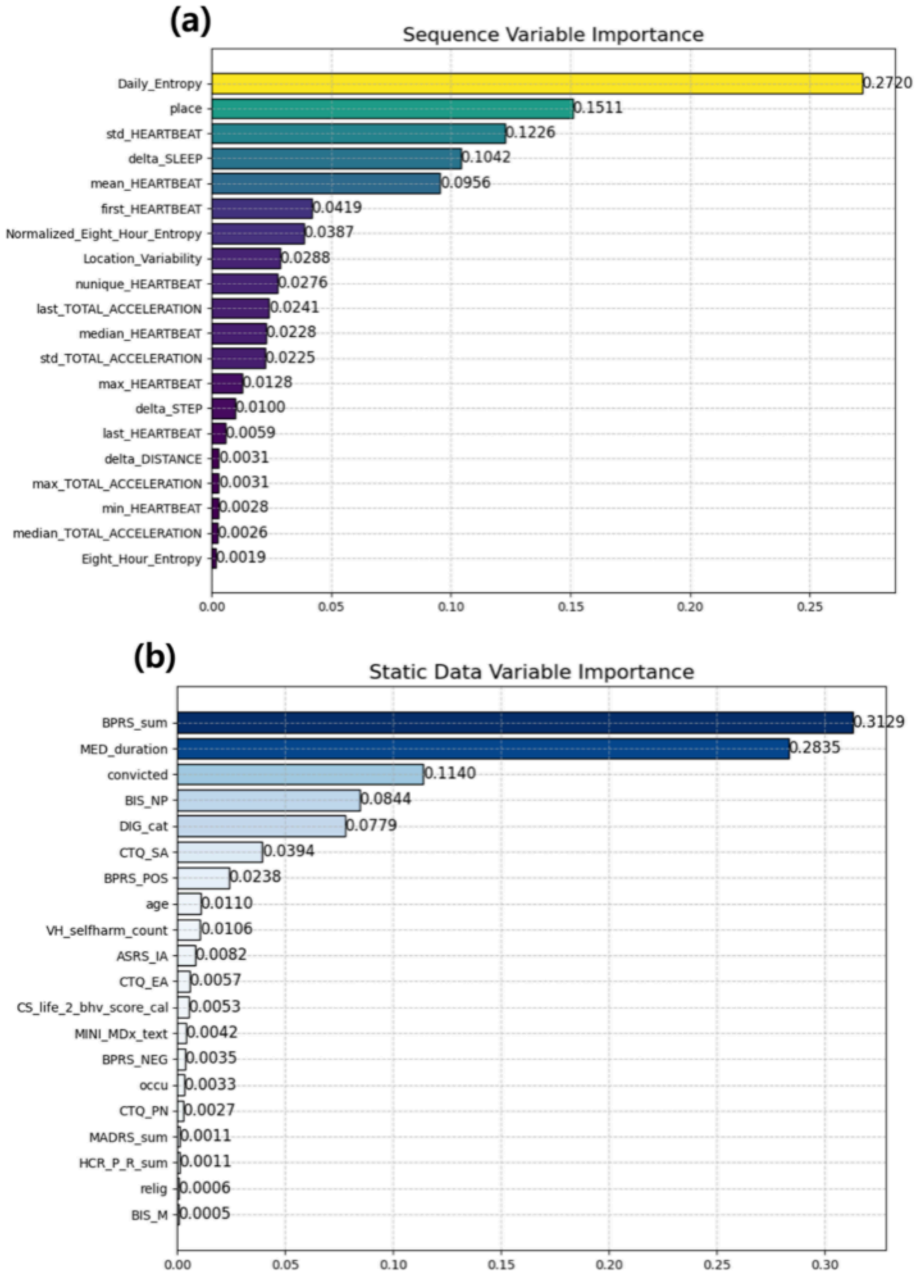

Fig. 5. Variable Importance for (a) sequence variable importance and (b) static data variable importance.

instance, the Temporal Fusion Transformer (TFT) may prioritize certain time points or feature combinations based on temporal information, leading to conservative predictions that lower recall but improve precision, or lenient predictions that increase recall but reduce precision. Furthermore, class imbalance within the dataset exacerbates this trade-off, as folds with fewer positive samples encourage the model to make more conservative predictions, further lowering recall. To mitigate this issue, future work will explore data augmentation techniques to balance the dataset and improve the model's ability to detect positive cases without compromising precision.

3.4. *Import variables*

As shown in Fig. 5, the importance of both time-series and static variables was assessed to evaluate their impact on the model's predictive performance. Among the time-series variables, Daily_Entropy (0.271965) was the most critical factor, indicating that diverse activity patterns play a key role in detecting changes in a patient's condition. Other important variables included Place (0.151081), Std_HEARTBEAT (0.122617), and Delta_SLEEP (0.104202), highlighting the significance of lifestyle, heart rate variability, and sleep duration in mental health. For static variables, BPRS_sum (0.312928) was the most significant, reflecting psychiatric symptom severity as crucial for prediction, while MED_duration (0.283524) and Convicted (0.114017) also contributed significantly.

4. Discussion and Conclusion

This study represents the first attempt to use the Temporal Fusion Transformer (TFT) model to predict harmful behaviors among psychiatric inpatients, thereby demonstrating the model's strengths in clinical settings. The TFT model integrates time-series data with static patient information, learning complex temporal patterns and optimizing predictive performance through automatic feautre selection.

By leveraging a Variable Selection Network, the model automatically identifies key elements from both time-series and static variables, enhancing efficiency and interpretability. This approach enables medical staff to pinpoint the variables most relevant to patient risks and utilize thiese insights for developing effective interventions and treatment plans. Moreover, the multi-head attention mechanism captures both short- and long-term dependencies, allowing the model to detect more complex behavioral patterns compared to existing models. The study's findings indicate that the TFT model achieves outstanding performance across various evaluation metrics, particularly accuracy, AUC, and specificity, underscoring its ability to detect risks early while minimizing false alarms. However, the limited dataset size and the small number of harmful behavior cases posed challenges for improving recall (0.7486). To address these constraints, future research will expand the dataset to include more patients and diverse ward environments while further refining the variable selection process to enhance interpretability. Nevertheless, implementing the TFT model in clinical settings faces significant barriers due to the technical complexity of integrating diverse data sources, such as biosignals, loactaion data, and clinical records. To overcome these obstacles, it is essential to employ the Common Data Model (CDM) for standardizing data formats and definitions and to establish a preprocessing pipeline that resolves issues related to time synchronization and missing data. Additionally, a cloud-based infrastructure can offer scalable solutions for real-time processing of large patient datasets, thereby helping to address these challenges. By overcoming these hurdles, the TFT model can strengthen clinical workflows, provide actionable insights, and ultimately contribute to improved patient safety and care in psychiatric wards.

5. Summary table

5.1. What was already known on the topic

- AI models show potential in monitoring psychiatric patients.
- LSTM and GRU models are widely used for time-series forecasting.
- Combining time-series data with patient data enhances risk prediction.

5.2. What this study adds to our knowledge

- The TFT model outperforms previous models in predicting harmful behavior.
- Integrating sensor and patient data improves accuracy and AUC.
- The variable selection feature of TFT enhances prediction performance.
- Bayesian optimization improves model tuning.

Declaration statement

The study protocol was reviewed and approved by the Institutional Review Board of the Seoul National University Hospital (IRB No. 2210–073-1368).

CRediT authorship contribution statement

Ri-Ra Kang: Writing – review & editing, Writing – original draft, Software, Methodology, Conceptualization. **Yong-gyom Kim:** Writing – review & editing, Validation, Methodology. **Minseok Hong:** Writing – review & editing, Validation, Data curation. **Yong Min Ahn:** Writing – review & editing, Supervision, Conceptualization. **KangYoon Lee:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: [KangYoon Lee reports was provided by Gachon University. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper].

Acknowledgements

This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number : HI22C1569)

Appendix A

Outliers were identified through visualization using boxplots, as shown in Fig. 6. However, certain visually striking extreme values in the psychiatric ward data were considered outliers. These values were removed to ensure data quality and accuracy in the analysis.

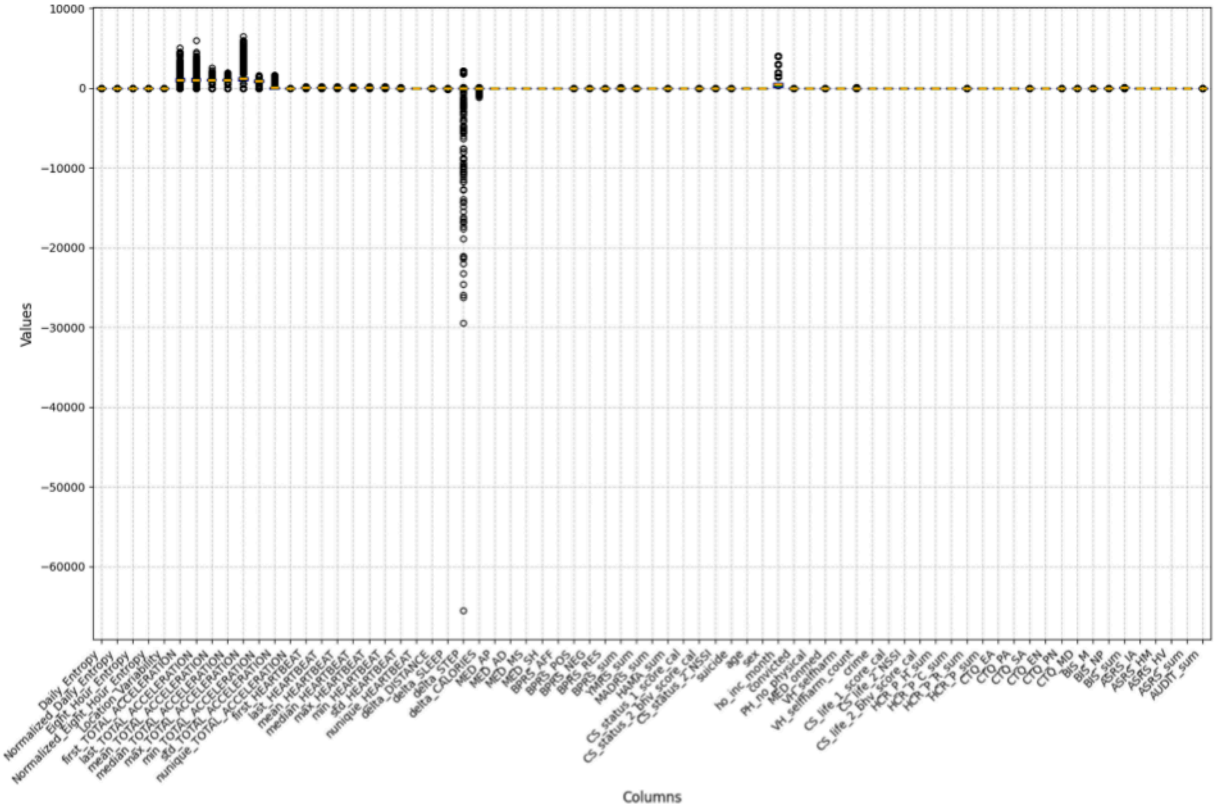

Fig. 6. Boxplot for outlier

Appendix B

Results of Pre-processing operations on the dataset

This heatmap visually represents the correlations between various variables collected from psychiatric wards and provides valuable insights for selecting significant features in a risk predictions for harmful behaviors model. For instance, variables with high correlations, such as 'mean_-HEARTBEAT' and 'max_HEARTBEAT, or 'Daily_Entropy' and 'Eight_Hour_Entropy', offer similar information. Therefore, selecting only one of these variables can reduce the model's complexity. In contrast, variables like 'age' and 'convicted', which show low correlations with other variables, may provide independent predictive power, thus enhancing the model's explanatory capability. This analysis helps minimize unnecessary redundancy and allows for the effective selection of variables that optimize predictive performance. Furthermore, the visual representation of the dataset features is displayed at the top of Fig. 7, while the correlation between the features of the dataset is visualized using a heatmap at the bottom of Fig. 7.

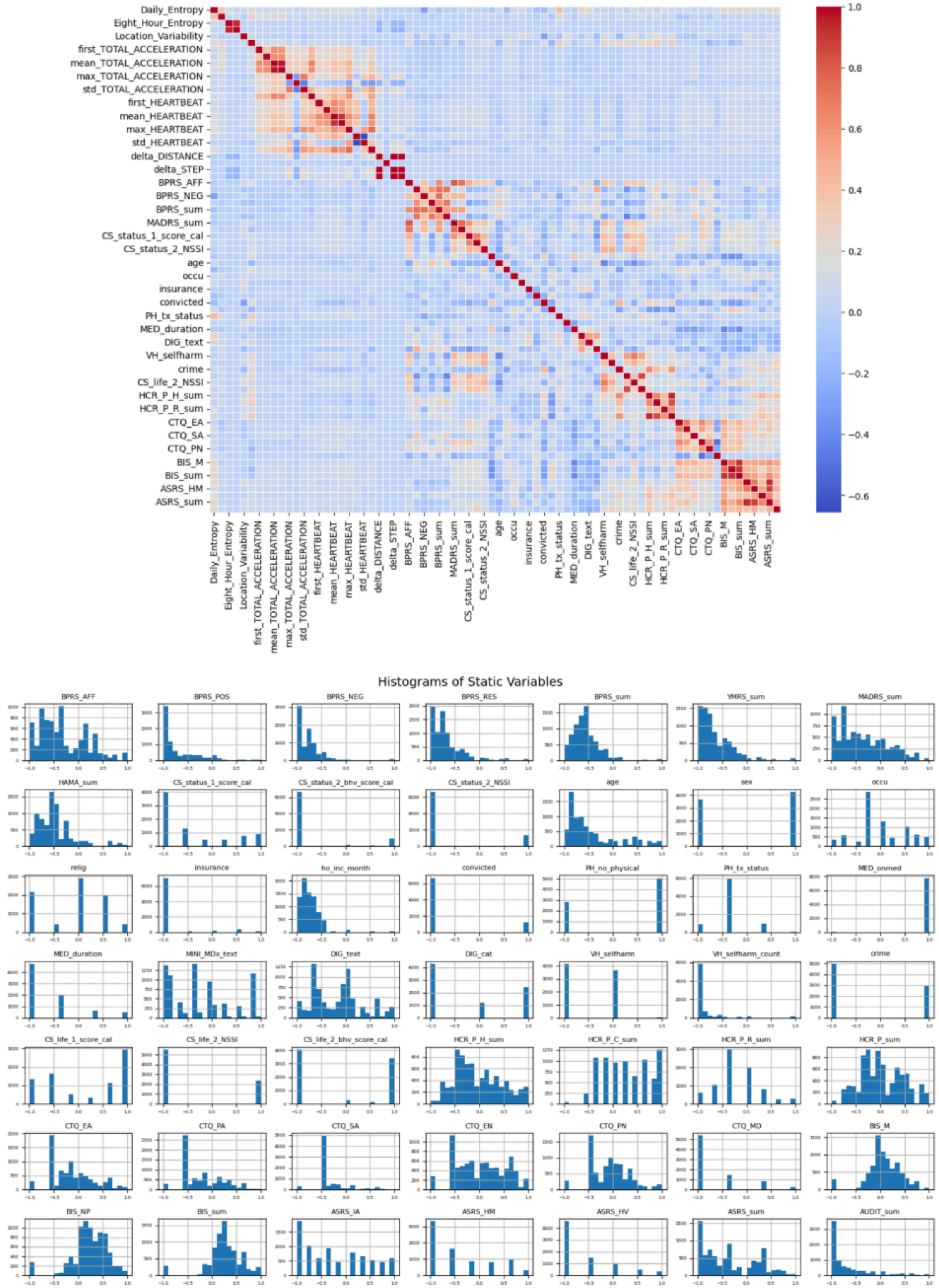

Fig. 7. Data Statistics

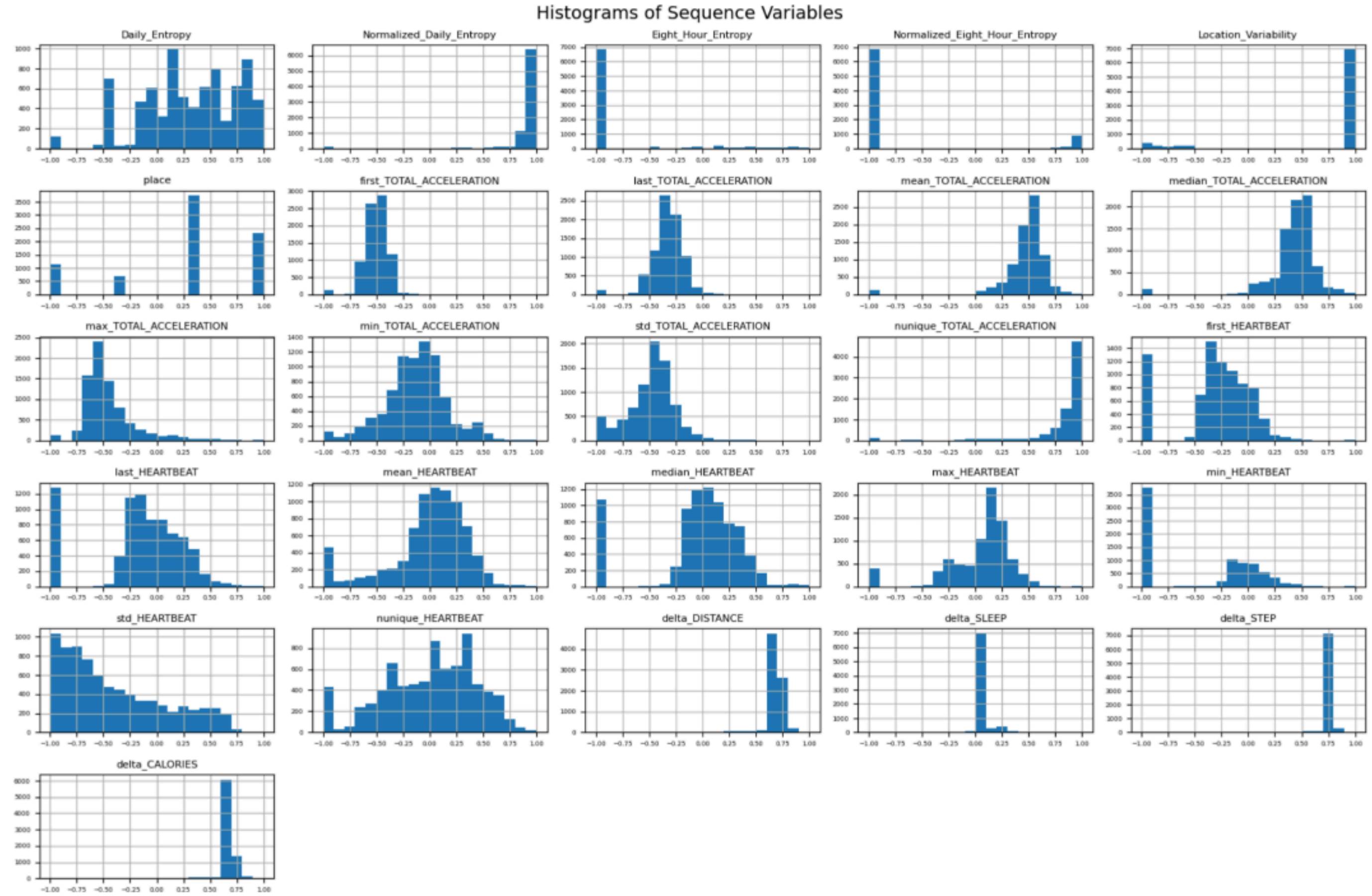

Fig. 7. (*continued*).

References

[1] World Health Organization. One in 100 deaths is by suicide. [Online]. Available: https://www.who.int/news/item/17-06-2021-one-in-100-deaths-is-by-suicide. [Accessed: 20-Aug-2024.

[2] D. Veale, et al., The psychiatric ward environment and nursing observations at night: a qualitative study, J. Psychiatr. Ment. Health Nurs., 27 (4) (2020) 342–351, https://doi.org/10.1111/jpm.12583.

[3] D.-S. Go, et al., A review of the admission system for mental disorders in South Korea, Int. J. Environ. Res. Public Health 17 (24) (2020) 9159, https://doi.org/10.3390/ijerph17249159.

[4] A. Wontorczyk, B. Izydorczyk, M. Makara-Studzińska, Burnout and stress in group of psychiatrists: workload and non-professional-social predictors, Int. J. Occup. Med. Environ. Health 36 (3) (2023) 379, https://doi.org/10.13075/ijomeh.1896.02147.

[5] I.M. Hunt, et al., Suicide in recently admitted psychiatric in-patients: a case-control study, J. Affect. Disord. 144 (1–2) (2013) 123–128, https://doi.org/10.1016/j.jad.2012.06.019.

[6] E.A. Deisenhammer, et al., Suicide in psychiatric inpatients—a case–control study, Front. Psychiatry 11 (2020) 591460, https://doi.org/10.3389/fpsyt.2020.591460.

[7] J.A. Naifeh, et al., Risk of suicide attempt in reserve versus active component soldiers during deployment to the wars in Iraq and Afghanistan, Suicide Life Threat. Behav. 52 (1) (2022) 24–36, https://doi.org/10.1111/sltb.12770.

[8] K.T. Pham, A. Nabizadeh, S. Selek, Artificial intelligence and chatbots in psychiatry, Psychiatr. Q 93 (1) (2022) 249–253, https://doi.org/10.1007/s11126-022-09973-8.

[9] H. Antonova, et al., Suicide Rate and Factors Analysis: Pre and Post COVID Pandemic Data.Analysis, in: 2022 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), IEEE, 2022, pp. 1–8, https://doi.org/10.1109/IEMTRONICS55184.2022.9795714.

[10] M.M. Tadesse, H. Lin, B. Xu, L. Yang, Detection of suicide ideation in social media forums using deep learning, Algorithms 13 (1) (2019) 7, https://doi.org/10.3390/a13010007.

[11] C. Su, R. Aseltine, R. Doshi, K. Chen, S.C. Rogers, F. Wang, Machine learning for suicide risk prediction in children and adolescents with electronic health records, Transl. Psychiatry 10 (1) (2020) 413, https://doi.org/10.1038/s41398-020-01100-0.

[12] Ben-Ari, Alon, and Kenric Hammond. "Text mining the EMR for modeling and predicting suicidal behavior among US veterans of the 1991 Persian Gulf War." 2015 48th Hawaii international conference on system sciences. IEEE, 2015. https://doi.org/10.1109/HICSS.2015.382.

[13] H. Ehtemam, et al., Role of machine learning algorithms in suicide risk prediction: a systematic review-meta analysis of clinical studies, BMC Med. Inform. Decis. Mak. 24 (1) (2024) 138.

[14] A. Lejeune, et al., Artificial intelligence and suicide prevention: a systematic review, Eur. Psychiatry 65 (1) (2022) e19.

[15] S. Tutun, et al., An AI-based decision support system for predicting mental health disorders, Inf. Syst. Front. 25 (3) (2023) 1261–1276.

[16] S.H. Oh, M. Kang, Y. Lee, Protected health information recognition by fine-tuning a pre-training transformer model, Healthc. Inform. Res. 28 (1) (2022) 16–24, https://doi.org/10.4258/hir.2022.28.1.16.

[17] S. Lee, et al., Current advances in wearable devices and their sensors in patients with depression, Front. Psychiatry 12 (2021) 672347, https://doi.org/10.3389/fpsyt.2021.672347.

[18] M. Elgendi, et al., Mobile and wearable systems for health monitoring, Front. Digit. Health 5 (2023) 1196103, https://doi.org/10.3389/fdgth.2023.1196103.

[19] T. Imbiriba, A. Demirkaya, A. Singh, D. Erdogmus, M.S. Goodwin, Wearable biosensing to predict imminent aggressive behavior in psychiatric inpatient youths with autism, JAMA Netw. Open. 6 (12) (2023) e2348898, https://doi.org/10.1001/jamanetworkopen.2023.48898.

[20] R.R. Kang, Y.G. Kim, M.S. Hong, J.H. Yang, Y.M. An, K.Y. Lee, LSTM-Based multimodal model for self-harm prediction combining mental health case reports and biometric signals, KCTRS 10 (8) (2024) 11–22, https://doi.org/10.47116/apjcri.2024.08.02.

[21] B. Lim, et al., Temporal fusion transformers for interpretable multi-horizon time series forecasting, Int. J. Forecast. 37 (4) (2021) 1748–1764, https://doi.org/10.1016/j.ijforecast.2021.03.012.

[22] R. Phetrittikun, K. Suvirat, T.N. Pattalung, C. Kongkamol, T. Ingviya, S. Chaichulee, Temporal Fusion Transformer for forecasting vital sign trajectories in intensive care patients, IEEE, 2021, pp. 1–5, https://doi.org/10.1109/BMEiCON53485.2021.9745215.

[23] Hong, M.R.-R. Kang, J.H. Yang, S.J. Rhee, H. Lee, Y. Kim, K.Y. Lee, H.G. Kim, Y. S. Lee, T. Youn, S.H. Kim, Y.M. Ahn, Comprehensive symptom prediction in inpatients with acute psychiatric disorders using wearable-based deep learning

models: development and validation study, JMIR 26 (2024) e65994, https://doi.org/10.2196/65994.

[24] B. Balliu, et al., Personalized mood prediction from patterns of behavior collected with smartphones, npj Digit. Med. 7 (1) (2024) 49, https://doi.org/10.1038/s41746-024-01035-6.

[25] M.S. Goodwin, et al., Predicting aggression to others in youth with autism using a wearable biosensor, Autism Res. 12 (8) (2019) 1286–1296, https://doi.org/10.1002/aur.2151.

[26] X. Yan, et al., Missing value imputation based on gaussian mixture model for the internet of things, Math. Probl. Eng. 2015 (2015) 548605, https://doi.org/10.1155/2015/548605.

[27] M. H. Anisi, A. H. Abdullah, and S. A. Razak, "Energy-Efficient Data Collection in Wireless Sensor Networks," vol. 5, no. 2, pp. 61–67, 2015. https://doi.org/10.4236/wsn.2011.310036.

[28] A. Paszke, et al., Pytorch: an imperative style, high-performance deep learning library, Adv. Neural Inf. Process. Syst. 32 (2019), https://doi.org/10.48550/arXiv.1912.01703.

[29] J. Wu, et al., Hyperparameter optimization for machine learning models based on Bayesian optimization, J. Electron. Sci. Technol. 17 (1) (2019) 26–40, https://doi.org/10.11989/JEST.1674-862X.80904120.

[30] A. Rastegarpanah, M.E. Asif, R. Stolkin, Hybrid neural networks for enhanced predictions of remaining useful life in lithium-ion batteries, Batteries 10 (3) (2024) 106, https://doi.org/10.3390/batteries10030106.

[31] A. Stuke, P. Rinke, M. Todorović, Efficient hyperparameter tuning for kernel ridge regression with Bayesian optimization, Mach. Learn.: Sci. Technol. 2 (3) (2021) 035022, https://doi.org/10.1088/2632-2153/abee59.

[32] S.A. Hicks, et al., On evaluation metrics for medical applications of artificial intelligence, Sci. Rep. 12 (1) (2022) 5979, https://doi.org/10.1038/s41598-022-09954-8.

[33] F. Iorfino, et al., Predicting self-harm within six months after initial presentation to youth mental health services: a machine learning study, PLoS One 15 (12) (2020) e0243467, https://doi.org/10.1371/journal.pone.0243467.

[34] A.M. Shah, W. Muhammad, K. Lee, Examining the determinants of patient perception of physician review helpfulness across different disease severities: a machine learning approach, Comput. Intell. Neurosci. 2022 (2022) 8623586, https://doi.org/10.1155/2022/8623586.

[35] A. Chadha, B. Kaushik, A hybrid deep learning model using grid search and cross-validation for effective classification and prediction of suicidal ideation from social network data, New Gener. Comput. 40 (4) (2022) 889–914, https://doi.org/10.1007/s00354-022-00191-1.

[36] D. Anguita, et al., The 'K' in K-fold Cross Validation, ESANN (2012) 441–446.