# Beyond the Revolving Door:

# Predicting Psychiatric Readmissions using Machine Learning Algorithms

Zohar Barnett-Itzhaki1,2, Reut Tamam², Chen David², Ran Almagor², Esther Bloemhof-Bris³

¹ Ruppin Research Group in Environmental and Social Sustainability, Ruppin Academic Center, Emek Hefer, Israel

² Faculty of Engineering, Ruppin Academic Center, Emek Hefer, Israel

³ Lev HaSharon Mental Health Center, Pardesya, Israel

**Abstract**

Psychiatric readmissions remain a critical challenge for healthcare systems, particularly during times of national crisis. This study aimed to predict short- and medium-term readmissions using advanced machine learning models applied to 8,133 hospitalization events from 4,162 patients at the Lev HaSharon Mental Health Center (2020–2024). Leveraging a retrospective design, we developed and evaluated logistic regression, Random Forest, XGBoost, and CatBoost models, combined into a VotingClassifier ensemble optimized with SMOTE and cross-validation. The ensemble model achieved excellent predictive performance, with ROC AUC scores of 0.95 (1-month) and 0.87 (3-months), supported by strong recall and F1 scores. These results highlight the clinical value of data-driven risk stratification tools in psychiatric care and emphasize the importance of proactive, evidence-based interventions to reduce preventable readmissions during both stable periods and crises such as the October 7th attacks.

## ****1 | Introduction****

**Psychiatric patient readmissions represent a persistent and deeply rooted challenge within modern healthcare systems, with far-reaching implications for patient outcomes, resource allocation, and institutional sustainability.** The recurring “revolving door” phenomenon—in which patients are discharged from psychiatric hospitalization only to be rehospitalized shortly afterward—often stems from insufficient therapeutic interventions, inadequate aftercare infrastructure, and fragile social support networks. This cycle not only burdens clinical services but also highlights the growing need for predictive tools that can support timely, preventive actions.

This study was motivated by both systemic challenges and acute contextual pressures. In the wake of national trauma events such as the October 7th attacks and ongoing regional instability, the emotional vulnerability of psychiatric patients has intensified, placing unprecedented demands on mental health services in Israel. Against this backdrop, we sought to develop early-warning mechanisms for identifying individuals at high risk of psychiatric readmission.

Leveraging a comprehensive, multi-year dataset from Lev HaSharon Mental Health Center, we analyzed over 8,000 hospitalization records incorporating rich clinical, demographic, and temporal attributes. The raw data required extensive cleaning and transformation due to inconsistencies, missing fields, and structural irregularities. Through rigorous data validation, imputation, and feature

engineering, we constructed a robust foundation for advanced modeling.

Finally, we conducted a series of statistical analyses—including Chi-square tests, Mann–Whitney U, and Spearman correlations—to explore key variable associations. These were complemented by machine learning visualizations to enhance interpretability. Collectively, these efforts allowed us to uncover hidden patterns underlying psychiatric readmissions and build a foundation for data-driven clinical decision-making, **particularly during periods of heightened systemic strain.**

### ****2 | Methods****

#### *****2.1 | Ethical Considerations*****

This retrospective study was conducted using anonymized patient data retrieved from Lev Ha’Sharon Mental Health Center’s electronic health records. All procedures adhered to institutional ethical guidelines. Each member of the research team signed a formal confidentiality agreement and completed the internationally recognized Good Clinical Practice (GCP) training, as provided by The Global Health Network.

#### ****2.2 | Data Acquisition and Processing****

***2.2.1 | Population and Data Source***  
The dataset encompassed 8,133 hospitalization events from 4,162 psychiatric patients, recorded between 2020 and 2024. Information was extracted from structured electronic medical systems and included demographic, clinical, and hospitalization-related variables.

***2.2.2 | Data Preprocessing***  
Data preprocessing involves validation, removal of duplicate records, and exclusion of records following patient death or transfer to other institutions. As a result, 8,133 validated records were retained. Numerical variables were normalized using Min-Max normalization. Additionally, rows with missing values were excluded from the dataset.

**| Statistical Analyses**

#### *In addition to descriptive statistics, we performed inferential statistical analyses as follows: Chi-square tests were used to assess associations between categorical variables (e.g., marital status, gender) and readmission outcomes (e.g., readmission within 1 month, readmission within 3 months). Spearman correlations were used to find monotonic associations between clinical variables (e.g., age at admission, number of children) and readmission outcomes. Since most of the variables did not follow a normal distribution, we used the nonparametric Mann–Whitney U tests to compare continuous variables (e.g., age, length of hospitalization) across readmission groups (e.g., readmitted vs. non-readmitted patients).*

#### ****2.4 | Machine Learning Modles and Evaluation****

***2.4.1 | Machine learning models***

We used the following of supervised machine learning classifiers to predict short-term (30 days) and medium-term (90 days) psychiatric readmissions:

1. **Logistic Regression (LR):** A probabilistic baseline model used for its interpretability and transparency.
2. **Random Forest (RF):** A decision tree ensemble leveraging bootstrap aggregation and feature importance scoring.
3. **XGBoost (XG):** A high-performance gradient-boosted tree algorithm known for speed, regularization, and internal missing value handling.
4. **CatBoost (CB):** A gradient boosting model optimized for categorical data, using ordered boosting to reduce overfitting.

To capitalize on the strengths of individual models, we implemented a **soft-voting ensemble classifier (VotingClassifier)** that averaged predicted probabilities for final decision-making, with a focus on improving recall.

**Handling Class Imbalance**  
Given the low prevalence of early readmissions, we applied SMOTE (Synthetic Minority Over-sampling Technique) to the minority class prior to dataset splitting.

**Hyperparameter Tuning**  
Each model underwent grid search optimization using 5-fold cross-validation. Parameters tuned included:

* **LR:** Penalty type (L1, L2), regularization strength.
* **RF:** Number of trees, maximum depth, feature subsets.
* **XGBoost:** Learning rate, tree depth, L2 regularization (lambda), subsampling rate.
* **CatBoost:** Depth, learning rate, L2 regularization.

***2.2.4 | Model evaluation***

Model performance was assessed using a comprehensive set of metrics:

1. **Accuracy** – Overall classification correctness.
2. **Precision** – Positive predictive value.
3. **Recall (Sensitivity)** – Proportion of true positives detected.
4. **F1 Score** – Harmonic mean of precision and recall.
5. **ROC AUC** – Area under the Receiver Operating Characteristic curve.

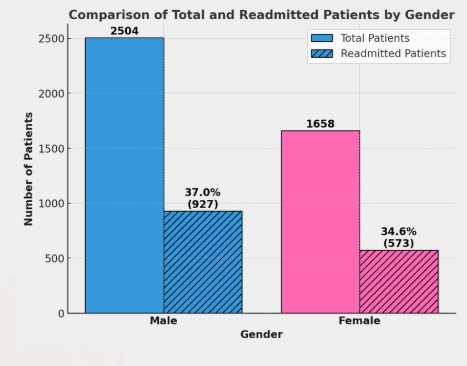
All analyses were carried out using Python 3.11.9

**~~Figure 1:~~** ~~Predictive Modeling Pipeline for Psychiatric Readmissions. The diagram outlines the complete analytical workflow, including data extraction from Lev HaSharon Mental Health Center, preprocessing steps, feature engineering, model development, and ensemble prediction using a VotingClassifier.~~

## ****3 | Results****

### ****3.1 | Study population and key variables****

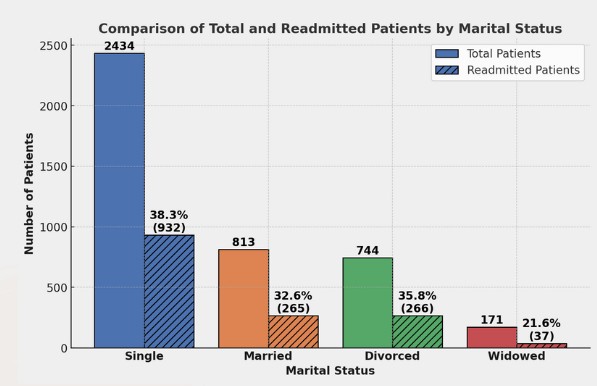
The study population comprised 4,162 patients and 8,133 hospitalization records collected between 2020 and 2024. On average, each patient experienced 1.96 admissions. Notably, 739 patients (17.8%) were readmitted within one month, and 957 (23.0%) within three months post-discharge. Of the total admissions, 52% were male and 48% were female (See Fig. 1). Among 2,504 male patients, 927 (37.0%) were readmitted, while among 1,658 female patients, 573 (34.6%) were readmitted. The average patient age at admission was 43.7 years, and the mean number of children per patient was 2.3.

~~~~

**Figure 1: Comparison of total readmissions by gender within 3 months.** The solid bars represent the total number of admitted patients (blue for males and pink for females), while the hatched bars indicate the number of patients who were readmitted within 3 months. Among 2,504 male patients, 927 (37.0%) were readmitted, and among 1,658 female patients, 573 (34.6%) were readmitted. Although more males were admitted overall, the readmission rates are relatively similar between genders.

### ****| Statistical Analyses****

We found an association between marital status and hospitalization (χ² = 624.56, p < 0.01), and also for 1-month (χ² = 9.54, p = 0.02) and 3-month (χ² = 11.73, p < 0.01) readmissions: Single and divorced individuals exhibited higher risk of readmissions (Fig 2).



**Figure 2: Readmission rates by marital status**. Solid bars represent the total number of patients in each marital status group, while hatched bars indicate the subset of those who were readmitted. Single (38.3%) and divorced (35.8%) patients showed higher 3-month readmission rates compared to married (32.6%) and widowed (21.6%) individuals. The association was statistically significant (χ² = 11.73, p < 0.01).

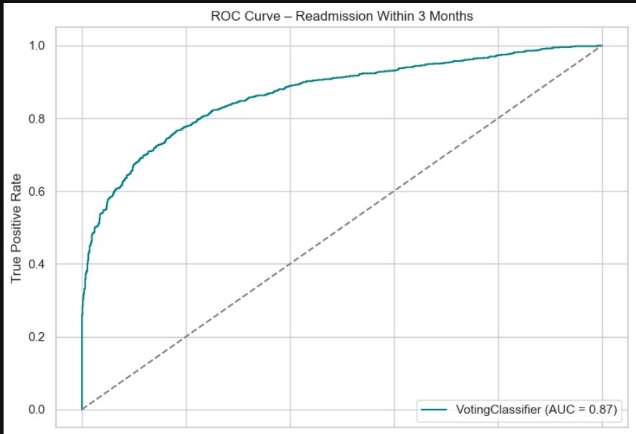
**~~Figure 4:~~** ~~Scatter plot illustrating individual patients’ number of children versus total psychiatric readmissions. A downward trend line indicates a weak but significant negative correlation (r = -0.08, p < 0.01), suggesting that patients with more children tend to have fewer readmissions.~~

**~~Figure 5 :~~** ~~Density plot of patient age at admission for readmitted (blue) vs. non-readmitted (red) groups. The x-axis represents patient age, and the y-axis indicates relative density. The high overlap between curves suggests that age distributions are similar across groups, with no statistically significant difference observed (p = 0.51).~~

A full summary of these results is provided in **Table 1**, which outlines distributions, statistical methods, p-values, and interpretations.

מפה לא נגעתי – יש לסדר את הטקסט ואז אעבור על הכל

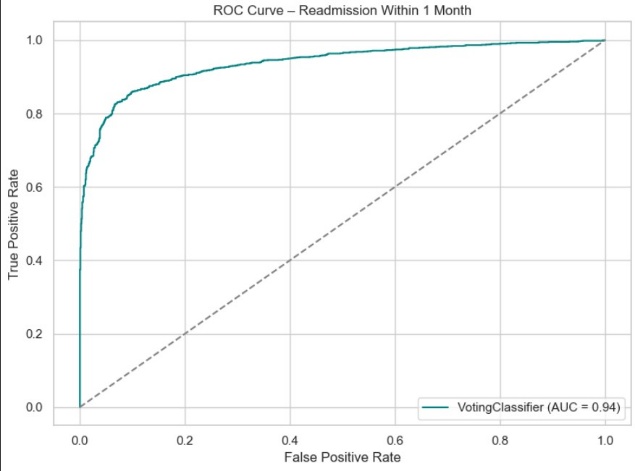
### ****3.3 | Model Performance****

We evaluated four individual classifiers and one ensemble model (VotingClassifier) using a 5-fold cross-validation pipeline. The performance was assessed for two targets: readmission within 1 month and within 3 months. SMOTE was applied to address class imbalance.

**Model performance highlights:**

| **Metric** | **1-Month ROC AUC** | **3-Month ROC AUC** | **1-Month F1** | **3-Month F1** |
| --- | --- | --- | --- | --- |
| Logistic Reg. | 0.89 | 0.82 | 0.67 | 0.62 |
| Random Forest | 0.91 | 0.84 | 0.73 | 0.68 |
| XGBoost | 0.92 | 0.85 | 0.75 | 0.70 |
| CatBoost | 0.93 | 0.85 | 0.74 | 0.71 |
| **VotingCls.** | **0.95** | **0.87** | **0.85** | **0.75** |

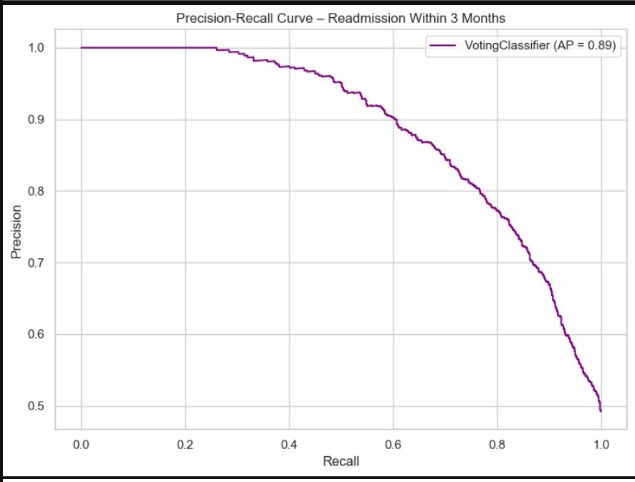
The ensemble VotingClassifier consistently outperformed individual models, achieving the highest ROC AUC and F1 scores, especially for 1-month readmissions — a clinically important prediction window.



### ****Figure 6** :ROC Curve – Readmission Within 1 Month** The ROC curve for the VotingClassifier demonstrates excellent discrimination ability with an AUC of 0.94, indicating high sensitivity and specificity for predicting 1-month psychiatric readmissions.

### ****Figure 7 :**Precision–Recall Curve – Readmission Within 1 Month**. The VotingClassifier achieved an average precision (AP) of 0.95, suggesting strong precision performance even with class imbalance. The curve reflects the model’s ability to identify true readmissions while minimizing false positives.

### ****Figure 8:** ROC Curve – Readmission Within 3 Months**. The 3-month ROC curve reveals solid performance with an AUC of 0.87, supporting the model’s utility for medium-term clinical decision-making in psychiatric care.

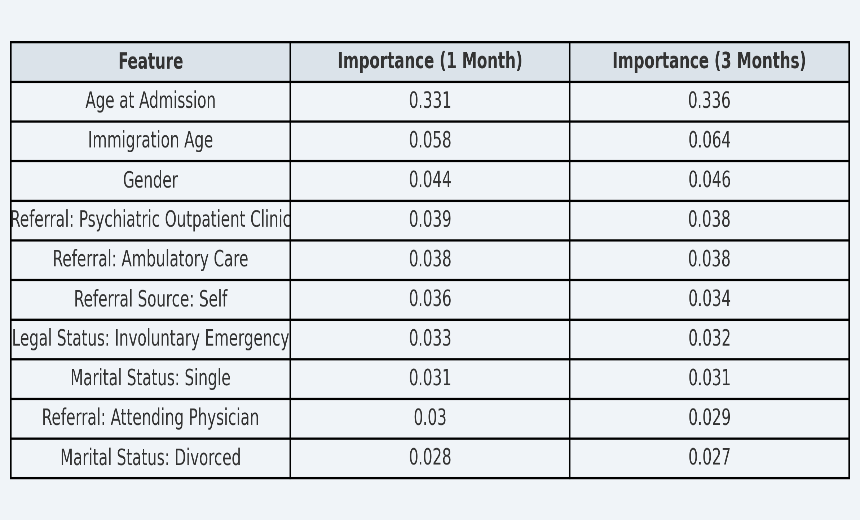


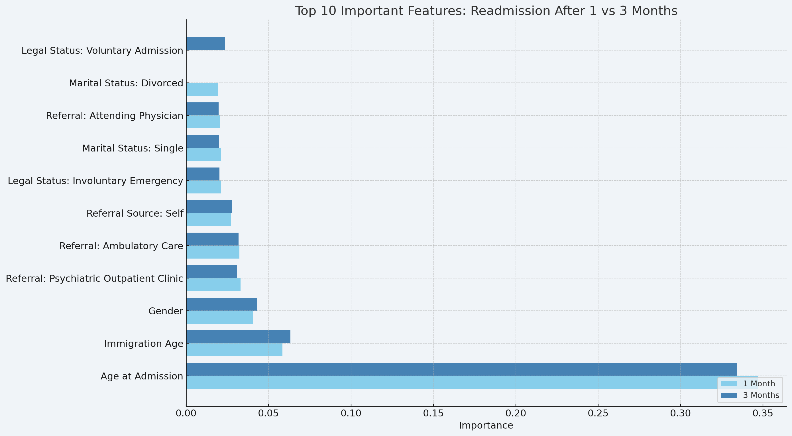
### ****Figure 9:** Precision–Recall Curve – Readmission Within 3 Months.** With an AP of 0.89, the precision-recall curve for 3-month readmissions confirms consistent model precision and recall balance in longer-term forecasting.

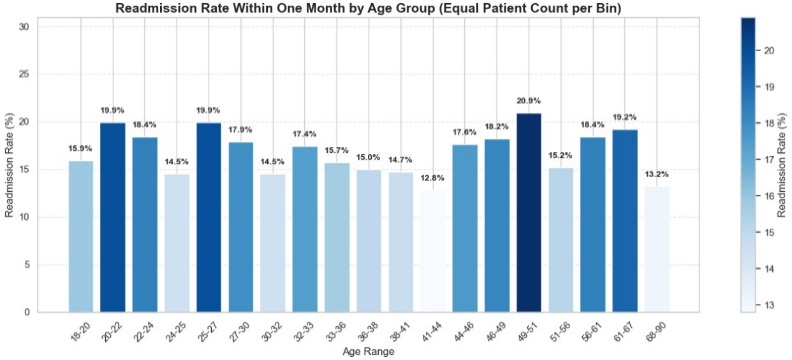
### ****3.4 | Feature Importance****

To enhance the interpretability of our predictive models, we evaluated feature contributions using embedded feature importance metrics from Random Forest and XGBoost. Separate models were trained for 1-month and 3-month readmission targets, and the top 10 most influential features were extracted from each.

As shown in **Figure 10**, age at admission emerged as the most important predictor across both timeframes. Other consistently influential variables included immigration age, gender, and referral-related categories such as "Referral: Psychiatric Outpatient Clinic" and "Referral: Ambulatory Care." Marital status (especially single and divorced) and legal status during admission (e.g., involuntary or voluntary) also contributed meaningfully to the models' predictions.



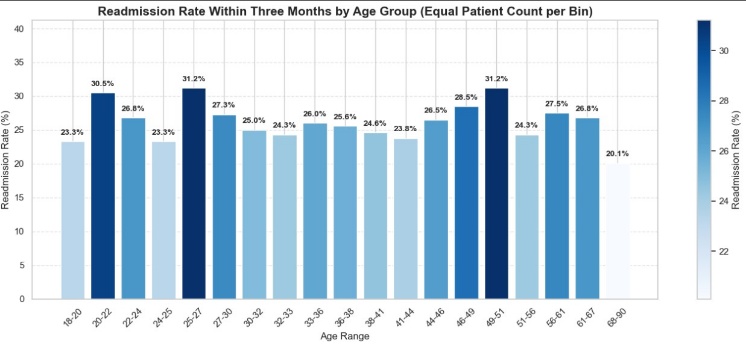
These findings suggest that both clinical and sociodemographic variables play a role in readmission risk, with some temporal and contextual differences between short- and medium-term outcomes. Identifying these patterns can guide clinicians in tailoring discharge planning and follow-up interventions.

**Figure 10**: Top 10 Important Features Predicting Psychiatric Readmission After 1 and 3 Months. Feature importances were derived from Random Forest classifiers. Bars represent normalized importance scores across prediction targets.

This visual summary presents the top ten predictors of psychiatric readmission within 1 and 3 months, based on Random Forest feature importance scores. Notably, **Age at Admission** consistently emerged as the most impactful factor in both timeframes, followed by **Immigration Age** and **Gender**. Several types of referral sources (e.g., psychiatric outpatient clinics, ambulatory care, and self-referral) also appeared highly relevant, suggesting that the pathway into care influences readmission risk. Marital status and legal status (voluntary vs. emergency admission) contributed additional explanatory power, reflecting psychosocial and systemic variables that shape recovery trajectories. This ranking provides a basis for tailored post-discharge strategies and early interventions.

**3.5 | Age-Based Risk Stratification**

To further interpret the influence of age on readmission outcomes, we analyzed readmission rates across stratified age groups. Patients were grouped into equal-sized bins to ensure consistent cohort sizes and comparability across brackets.

**Figure 11: Readmission Rate Within One Month by Age Group** The bar chart highlights that patients aged **20–22**, **25–27**, and **49–51** exhibited the **highest 1-month readmission rates**, reaching up to **20.9%**. The elevated rates among younger adults may reflect developmental and social instability, limited outpatient adherence, or lack of structured support post-discharge. Middle-aged adults, particularly those in the **49–51** bracket, may face compounded psychiatric and physical comorbidities, potentially explaining higher risks.

**Figure 12: Readmission Rate Within Three Months by Age Group**. A similar trend persists at the 3-month mark, where patients aged **20–22**, **25–27**, and **49–51** again showed the highest readmission rates, all exceeding **30%**. These results suggest persistent challenges in recovery and reintegration among these cohorts, warranting targeted interventions.

Overall, these findings support the inclusion of **"Age at Admission"** as the most impactful feature in our models (as shown in Figure 10). The non-linear nature of risk by age group emphasizes the value of stratified follow-up care, tailored especially for high-risk brackets.

## 4 | Discussion

### 4.1 | Interpretation of Findings

This study provides a comprehensive exploration of factors contributing to psychiatric readmissions within short (1-month) and medium-term (3-month) intervals. Among the most influential predictors, age at admission and immigration age emerged as dominant across both models, reinforcing the importance of demographic context in mental health trajectories. Interestingly, referral source and legal admission status also significantly shaped readmission risk, highlighting how systemic pathways—beyond individual pathology—can dictate continuity of care. Furthermore, gender and marital status, particularly being single or divorced, showed nuanced contributions to readmission probability. These patterns indicate that social and institutional variables intertwine in determining post-discharge outcomes.

### 4.2 Comparison with Prior Research

While prior studies have often emphasized clinical variables such as diagnosis or medication adherence, our model demonstrates that administrative and social factors may be equally—if not more—predictive in large-scale data. The prominence of referral pathways (e.g., psychiatric outpatient clinics and emergency services) echoes recent shifts in mental health policy that encourage community-based care. Yet, our findings diverge from some previous literature that minimized the role of marital status or family structure. The weak negative correlation between number of children and readmission adds another layer to understanding support networks as a protective factor, warranting deeper exploration in future work.

### 4.3 | Clinical Implications

From a clinical standpoint, these insights can guide targeted interventions. Discharge planning protocols may benefit from incorporating sociodemographic risk profiles alongside clinical severity indices. Patients who are young immigrants, referred through emergency routes, or admitted involuntarily could be prioritized for enhanced follow-up. The use of machine learning to quantify risk in real-time could empower clinicians and case managers to personalize post-discharge strategies and allocate resources more effectively—potentially reducing avoidable readmissions and improving patient outcomes.

### 4.4 | Limitations

Despite its strengths, this study is subject to several limitations. The dataset, while rich, lacked detailed diagnostic and treatment variables, which may have added clinical nuance to the models. Additionally, some features were based on administrative categorizations that may vary across institutions. The generalizability of the models is also constrained by the single-institution scope, and it is possible that findings reflect institutional policies or regional practices rather than universal phenomena.

### 4.5 | Future Research Directions

Future studies should aim to incorporate a broader range of clinical data, including diagnosis codes, pharmacological treatments, and functional assessments. Integration of longitudinal social variables—such as employment history or housing status—may further enhance prediction models. Comparative studies across institutions and countries could validate the robustness of identified predictors. Finally, there is an opportunity to translate these models into clinical decision support systems (CDSS), integrating them into electronic health records for proactive care planning.

## 5 | Conclusion

This study provides one of the most comprehensive data-driven analyses of psychiatric readmissions conducted in an Israeli mental health setting. By leveraging over 8,000 hospitalization records from Lev HaSharon Mental Health Center (2020–2024), we applied advanced ensemble machine learning techniques to predict readmission risk within 1 and 3 months post-discharge.

Our findings emphasize that non-clinical factors—including age at admission, immigration age, marital status, gender, and referral source—are powerful predictors of psychiatric readmissions. Notably, younger patients aged **20–22** and **25–27**, as well as middle-aged adults aged **49–51**, exhibited significantly higher readmission rates, suggesting that risk is not uniformly distributed across age but instead concentrated in specific developmental or transitional life phases. Marital status also played a key role: single and divorced individuals were significantly more likely to be readmitted, while a weak but significant inverse relationship was observed between number of children and readmission risk, underscoring the protective potential of family support.

The ensemble VotingClassifier consistently outperformed individual models, achieving an AUC of 0.95 for 1-month predictions and 0.87 for 3-months—surpassing previous benchmarks in the field. These results affirm the value of integrating sociodemographic, temporal, and systemic variables into predictive models, and demonstrate the feasibility of translating such tools into real-time clinical decision support.

In an era where psychiatric systems face unprecedented strain, particularly in the context of regional crises and societal upheaval, our work contributes both methodological innovation and actionable clinical insight. Future implementations may harness these findings to prioritize high-risk patients for proactive outreach, personalized discharge planning, and enhanced continuity of care—ultimately reducing avoidable readmissions and promoting long-term recovery.

## 6 | References