applied sciences

MDPI

Article

Predicting ICU Readmission Afte Intracerebral Hemorrhage: A Deep Learning Framework Using MIMIC Time-Series Data

Sergio Celada-Bernal *, Alejandro Piñán-Roescher, Ruyman Hernández-López and Carlos M. Travieso-González *

Signals and Communications Department (DSC), Institute for Technological Development and Innovation in Communications (IDeTIC), University of Las Palmas de Gran Canaria (ULPGC), 35017 Las Palmas de Gran Canaria, Spain; alejandro.pinan@ulpgc.es (A.P.-R.); ruyman.hernandez@ulpgc.es (R.H.-L.)

* Correspondence: sergio.celada@ulpgc.es (S.C.-B.); carlos.travieso@ulpgc.es (C.M.T.-G.)

Abstract

Intensive Care Unit (ICU) readmissions following Intracerebral Hemorrhage (ICH) are associated with increased mortality and resource burden. Current prediction models predominantly rely on static admission features, failing to capture the temporal evolution of physiological instability. This study proposes a novel deep learning framework to predict ICU readmission by leveraging high-resolution time-series data from the MIMIC-III and MIMIC-IV databases. We developed a Stacked Gated Recurrent Unit (GRU) Architecture Ensemble, integrated with Time-series Generative Adversarial Networks (TimeGAN) to address the inherent class imbalance of readmission events. Our model achieved a state-of-the-art Area Under the Receiver Operating Characteristic Curve (AUC) of 0.912, significantly outperforming traditional machine learning baselines and static feature models. The sensitivity of 88.1% highlights the model's efficacy in minimizing unsafe premature discharges. Furthermore, interpretability analysis using SHAP values identified Length of Stay, MELD Score, and Monocytes as critical predictors, revealing that readmission risk is driven by a complex interplay between systemic organ dysfunction and inflammatory response. These findings demonstrate that incorporating temporal dynamics and generative data augmentation significantly enhances risk stratification, offering a robust clinical decision support tool to optimize discharge timing in neurocritical care.

Keywords: intracerebral hemorrhage; ICU readmission; deep learning; time series analysis

Academic Editors: Maria Samarakou, Eleni Tsalera and Keun Ho Ryu

Received: 30 January 2026
Revised: 15 February 2026
Accepted: 22 February 2026
Published: 26 February 2026

1. Introduction

Intracerebral hemorrhage (ICH) constitutes a major neurological emergency, accounting for approximately 10–15% of all strokes worldwide yet contributing disproportionately to the global burden of mortality and severe disability [1,2]. In contrast to ischemic stroke, ICH is characterized by early clinical instability due to hematoma expansion and perihematomal edema, necessitating intensive monitoring in the Intensive Care Unit (ICU) [3]. Although acute-phase survival has modestly improved over the past decade through optimization of blood pressure control and anticoagulation reversal [4], the transition from the ICU to the general ward remains a critical point of vulnerability and inefficiency within the continuum of care [5].

Premature ICU discharge is a major determinant of serious adverse events. Patients readmitted to the ICU face significantly worse prognoses, with higher in-hospital mortality

rates and a drastic increase in healthcare costs [6,7]. Currently, the lack of standardized guidelines for discharging patients with ICH renders this decision a subjective clinical challenge, in which balancing the avoidance of premature transfer with the optimization of limited resources is critical. Therefore, it is imperative to develop objective predictive tools capable of accurately identifying the risk of readmission.

The primary objective of this study is to develop, validate, and compare a robust predictive framework for ICU readmission in patients with ICH. Unlike previous studies that rely predominantly on static admission features, this work introduces a dynamic deep learning framework specifically designed to capture the temporal evolution of physiological instability. The novelty of our approach lies in the integration of sequential modeling architectures within an ensemble strategy, combined with generative data augmentation to address class imbalance in a clinically coherent manner.

The main contributions of this study are threefold:

(i) The implementation and systematic comparison of advanced time-series architectures (GRU, Stacked GRU, TCN, and Transformer) for ICH-specific ICU readmission prediction.
(ii) The incorporation of a deep ensemble strategy to enhance predictive stability and generalization.
(iii) The integration of generative augmentation to improve minority-class representation while preserving temporal physiological structure.

We hypothesize that modeling the full temporal dynamics of vital signs and laboratory trajectories enables the identification of latent risk patterns, thereby outperforming conventional static models and providing a more reliable clinical decision-support tool.

The remainder of this manuscript is organized as follows. Section 2 presents a structured review of related work. Section 3 describes the materials and methodological framework. Section 4 reports the experimental results. Section 5 discusses the clinical implications and limitations. Finally, Section 6 concludes the study and outlines future research directions.

2. Related Work

Recent advances in Artificial Intelligence have significantly improved outcome prediction in critical care. Integrated optimization and machine learning approaches have been developed that outperform traditional clinical scoring systems [8,9]. Within this context, it is important to highlight the architectural diversity in time-series modeling. To capture complex temporal dependencies and dynamic changes in patient physiology, it is necessary to transition from static models toward advanced deep learning architectures capable of processing the evolving sequence of data [10].

Human physiology is intrinsically dynamic; therefore, applying Deep Learning to time-series data offers a superior theoretical advantage. Recurrent Neural Network (RNN) architectures such as the Gated Recurrent Unit (GRU) enable the capture of sequential dependencies to predict readmissions and critical events with greater accuracy than static methods [11,12]. In parallel, Temporal Convolutional Networks (TCN) have gained traction as an efficient alternative, as they excel in processing long-range patterns more rapidly than RNNs, facilitating their implementation in continuous monitoring systems [13,14]. Finally, the recent incorporation of Transformers has revolutionized clinical data analysis. Attention mechanisms have been shown to dynamically weight distant events in a patient's history, achieving unprecedented accuracy in predicting adverse outcomes [15–17].

3. Materials and Methods

3.1. Data Sources

This retrospective study was conducted using data extracted from two widely recognized open-access critical care databases: MIMIC-III (version 1.4) [18] and MIMIC-IV (version 3.1) [19]. Both repositories, developed by the MIT Laboratory for Computational Physiology, contain detailed clinical information collected at the Beth Israel Deaconess Medical Center (Boston, MA, USA).

These databases provide a comprehensive compendium of anonymized clinical information, including demographic data, vital signs (monitored on an hourly basis), laboratory results, medication histories, imaging studies, clinical notes, and survival records. The integration of both datasets enables leveraging their scale and diversity, supporting a robust analysis of ICU readmission risk in a heterogeneous cohort of patients with ICH, thereby improving the generalization and reliability of the findings. These databases represent a cornerstone in critical care informatics and have been extensively validated in the recent literature for developing predictive models of ICU readmission and mortality in patients with Intracerebral Hemorrhage, as demonstrated by the works [8,9]. Their widespread adoption in this specific domain confirms their suitability and robustness for training deep learning algorithms on high-resolution physiological time-series.

In accordance with privacy regulations, all data were pre-deidentified to ensure strict compliance with patient protection standards (HIPAA), and therefore informed consent requirements were waived. Access to the databases was granted after completing the required ethics training for research involving human subjects (CITI Program) and obtaining the corresponding certification (Record ID: 67993327).

3.2. Study Population

The study population consisted of patients diagnosed with Intracerebral Hemorrhage (ICH). Identification was performed through Structured Query Language (SQL) queries on the Google BigQuery platform, utilizing the International Classification of Diseases, Ninth Revision (ICD-9) code 431, and Tenth Revision (ICD-10) codes I610–I616 and I618–I619. Patients were included if ICH was recorded as either their primary or secondary diagnosis.

To ensure a well-defined and clinically relevant cohort, specific exclusion criteria were applied. We applied specific exclusion criteria to ensure cohort homogeneity and data quality:

- Patients younger than 18 years were excluded due to physiological differences in pediatric populations.
- Patients with an initial ICU length of stay shorter than 24 h were removed to ensure sufficient time-series data density for temporal modeling.
- Patients who died during or immediately following their first ICU discharge were excluded to account for competing risks.

ICU readmission was operationally defined as any unplanned return to the Intensive Care Unit occurring during the same hospitalization after the initial ICU discharge. Only readmissions within the same hospital admission episode were considered. Planned transfers for elective procedures were excluded when identifiable through admission type and procedure records. This definition ensures consistency with prior ICU readmission prediction studies and enhances reproducibility.

Prior to merging MIMIC-III and MIMIC-IV, a formal harmonization process was conducted to ensure dataset comparability. Units of measurement for laboratory and physiological variables were standardized across databases. Timestamp granularity was aligned to a uniform hourly resolution to preserve temporal consistency. Diagnostic codes were harmonized across ICD-9 and ICD-10 definitions to ensure uniform cohort selection.

Following the application of these criteria, a total of 2117 patients were included in the final cohort. This comprised 601 patients from the MIMIC-III database (selected from an initial pool of 1085) and 1516 patients from the MIMIC-IV database (selected from an initial pool of 2424). These datasets were harmonized and merged into a single unified cohort to facilitate robust model training and analysis. The complete workflow of the patient selection and exclusion process is illustrated in Figure 1.

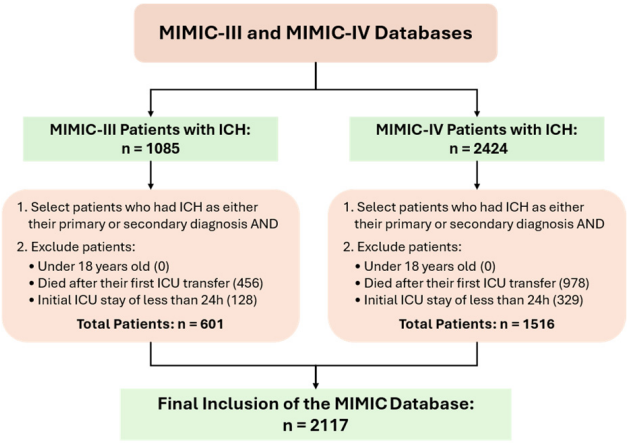

Figure 1. Criteria for study population extraction.

3.3. Feature Extraction

Predictor variable selection was performed using a hybrid approach guided by clinical evidence. Initially, candidate features were identified based on an extensive review of prior literature, prioritizing those that have demonstrated a significant association with intracerebral hemorrhage prognosis and ICU readmission risk in previous studies [8,9,20,21]. Subsequently, this preliminary set was refined and validated through expert clinical consensus to ensure both clinical relevance and availability within the databases.

To address irregular sampling and missing values, a 'Last Observation Carried Forward' (LOCF) imputation strategy was implemented. Specifically, the sequence of the last 24 time steps prior to discharge was extracted for each patient. Missing values within these sequences were imputed using forward filling to preserve physiological continuity, while any initial missing entries were filled with zeros. Subsequently, to ensure equal contribution from features with varying scales, continuous variables were normalized using Z-score standardization (zero mean and unit variance). The scaling parameters were derived exclusively from the training set and applied to the test set to prevent data leakage

All predictor variables were temporally restricted to information available prior to the initial ICU discharge. Mechanical Ventilation status reflected its occurrence during the first ICU admission only. Hydrocephalus was defined based on diagnoses documented before discharge. ICU Length of Stay corresponded exclusively to the duration of the index ICU episode. No post-discharge or post-readmission data were included in feature construction.

As a result of this process, a total of 28 final variables were selected, which are detailed in Table 1. For descriptive analysis of these characteristics, continuous variables were expressed using the median and interquartile range *M(Q1,Q3)* due to the non-normal distribution of physiological data, while categorical variables were reported as absolute and relative frequencies *N()*.

Table 1. Selected clinical features.

Characteristic	Non-ICU_Readmission (0) (N = 1694)	ICU_Readmission (1) (N = 423)
anchor_age	68.00 (56.00, 79.00)	65.00 (54.00, 75.00)
hospital_stay_days	3.92 (2.07, 8.12)	5.91 (2.76, 11.19)
MELD_score	8.50 (7.50, 10.20)	8.50 (7.50, 11.00)
ALT (IU/L)	22.00 (15.00, 35.78)	23.50 (17.00, 46.23)
Chloride (mEq/L)	103.89 (101.58, 106.33)	103.43 (100.88, 106.00)
Creatinine (mg/dL)	0.81 (0.66, 1.03)	0.80 (0.61, 1.02)
Sodium (mEq/L)	139.84 (137.69, 142.06)	139.72 (137.50, 142.46)
MCHC (%)	32.98 (32.04, 33.80)	32.65 (31.84, 33.52)
Monocytes (%)	6.70 (5.06, 8.84)	7.16 (5.40, 9.30)
Neutrophils (%)	76.50 (69.32, 82.65)	76.70 (70.17, 81.58)
PT (s)	12.50 (11.70, 13.56)	12.84 (11.90, 14.04)
INR	1.14 (1.07, 1.23)	1.17 (1.10, 1.28)
SpO2 (%)	96.60 (95.66, 97.64)	96.86 (95.83, 97.73)
GCS	13.55 (11.08, 14.71)	12.83 (9.96, 14.53)
HR (beats/min)	78.18 (70.65, 86.63)	80.46 (72.20, 87.56)
RR (breaths/min)	18.48 (16.81, 20.55)	18.79 (17.03, 20.92)
Eosinophils (%)	1.00 (0.30, 2.14)	1.20 (0.50, 2.23)
Albumin (g/L)	3.50 (3.10, 3.90)	3.40 (3.00, 3.80)
White Blood Cells (K/uL)	9.68 (7.88, 11.55)	9.99 (8.38, 11.92)
Glucose (mg/dL)	120.48 (107.00, 142.58)	122.43 (110.18, 144.35)
CK (IU/L)	131.00 (67.00, 280.02)	142.25 (69.12, 284.06)
Lymphocytes (%)	13.50 (9.07, 19.10)	12.30 (8.90, 17.24)
Potassium (mEq/L)	3.96 (3.77, 4.18)	3.98 (3.79, 4.22)
Platelets (K/uL)	224.95 (181.39, 275.41)	243.32 (185.10, 305.10)
Bilirubin_total (mg/dL)	0.52 (0.40, 0.80)	0.58 (0.37, 0.83)
Hydrocephalus (Yes)	100 (5.9%)	36 (8.5%)
Hydrocephalus (No)	1594 (94.1%)	387 (91.5%)
Mechanical_ventilation (Yes)	582 (34.4%)	205 (48.5%)
Mechanical_ventilation (No)	1112 (65.6%)	218 (51.5%)
Sex (M)	898 (53.0%)	244 (57.7%)
Sex (F)	796 (47.0%)	179 (42.3%)

3.4. Methodology

3.4.1. Proposed Methodological Framework

The methodological workflow adopted in this study is illustrated in Figure 2. The process begins with the filtered and processed database described in the previous sections and is structured into four sequential phases: data partitioning, balancing through synthetic data generation, modeling via deep architecture ensembles, and comparative evaluation.

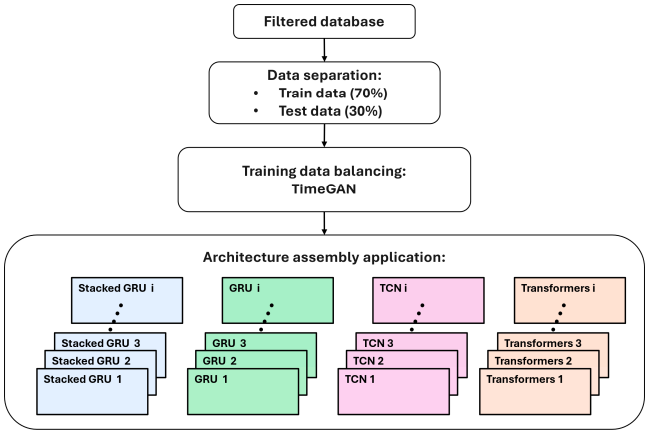

Figure 2. Proposed Methodological Framework.

3.4.2. Data Splitting and Balancing (TimeGAN)

First, the final dataset was randomly divided into two independent subsets: 70% allocated for model training and 30% reserved for testing. Synthetic data generation using TimeGAN was performed exclusively on the training subset and strictly after the dataset was partitioned. The test set was never augmented and contained only real patient data to ensure unbiased performance evaluation. The dataset was partitioned to ensure independent training and testing cohorts. Although MIMIC-III and MIMIC-IV originate from the same institution, their different temporal coverage and database structures allow for cross-version validation within the same healthcare system. It is important to note that this does not constitute true external validation across independent hospitals but rather an assessment of model generalizability across non-overlapping institutional cohorts.

Given the pronounced class imbalance inherent to ICU readmission prediction, we applied a state-of-the-art data augmentation technique: Time-series Generative Adversarial Networks (TimeGAN). TimeGAN has been shown to outperform traditional oversampling methods in biomedical domains by learning the joint temporal distribution of the data and generating synthetic sequences that preserve the integrity of dynamic physiological correlations without introducing artificial noise [22].

The TimeGAN model was configured with the following hyperparameters to stabilize the adversarial training process: a GRU-based generator and discriminator with 3 hidden layers, a hidden dimension of 24 units, and a learning rate of 1×10^{-3}. The model was trained for 10,000 iterations with a batch size of 128. To ensure the reliability of the generated samples, we rigorously monitored the convergence of the reconstruction and adversarial loss functions. The training process was terminated only upon reaching a stable equilibrium between the generator and discriminator, ensuring that the model successfully minimized the distributional discrepancy between real and synthetic data sequences without exhibiting mode collapse.

To assess the fidelity of the generated time series, we conducted a visual dimensionality reduction analysis using t-SNE (t-Distributed Stochastic Neighbor Embedding). As illustrated in Figure 3, the synthetic samples (red) exhibit a high degree of overlap and density consistency with the original clinical trajectories (blue), indicating that the TimeGAN model successfully captured the underlying high-dimensional manifold of the ICU cohort. The lack of isolated clusters for the synthetic data suggests that the generator effectively modeled the global distribution without suffering from mode collapse. Furthermore, the quantitative utility of the data was confirmed through the Predictive Score, yielding a Mean Absolute Error (MAE) of 0.1137. This low error rate demonstrates that the synthetic

sequences preserve the critical temporal dynamics and clinical information necessary for robust predictive modeling.

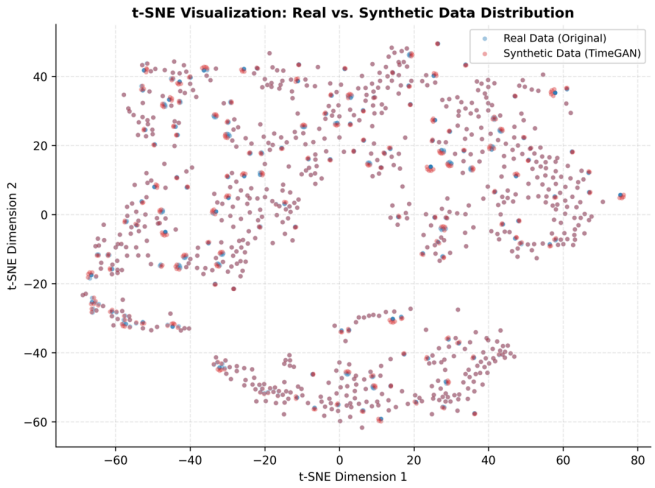

Figure 3. t-SNE visualization.

3.4.3. Deep Learning Architectures and Ensemble Strategy

For risk prediction, we implemented four deep learning architectures specialized in sequential data. To mitigate the variance inherent in stochastic training and enhance generalization, we implemented an ensemble strategy following the methodology described in [23]. The final prediction is computed using a Soft Voting (Averaging) mechanism, where the outputs of all individual networks are averaged to produce a robust consensus score. This approach effectively reduces the prediction error and prevents the model from relying on the idiosyncrasies of a single training run, a technique that has proven effective in handling missing clinical data in intensive care settings. The architectures evaluated under this ensemble scheme were:

- Gated Recurrent Unit (GRU): Captures short-term dependencies in hospital readmission tasks, addressing gradient issues at lower computational cost than LSTMs [24].
- Stacked GRU: A deep configuration stacking multiple hidden layers, which enables the abstraction of complex hierarchical patterns in the evolution of vital signs that a single layer cannot detect [25].
- Temporal Convolutional Network (TCN): Provides parallelization capability and long-range memory through dilated convolutions and has demonstrated superior performance in early detection of adverse events in ICU settings [26].
- Transformer: A model based on self-attention mechanisms. Its inclusion is justified by recent research showing that Transformers capture critical long-range temporal relationships more effectively than any recurrent model [17].
- To process variable-length clinical sequences, a fixed input window of 24 time steps was established. For patients with ICU stays shorter than this window, zero-padding (pre-padding) was applied to the beginning of the sequence to ensure uniform input dimensions for the neural networks. This approach allows the model to handle varying stay durations while maintaining a consistent tensor shape.

The prediction was generated at the time of ICU discharge. All predictor variables were constructed using only information available up to that discharge moment. Time-series inputs were derived exclusively from physiological measurements recorded prior to ICU discharge, and no post-discharge or post-readmission data were incorporated. This design ensures strict temporal separation between predictors and outcome, preventing information leakage.

To guarantee the reproducibility of the study and provide a transparent overview of the methodological design, the specific architectural parameters and training configurations for all implemented models have been consolidated. Table 2 presents a detailed breakdown of the structural specifications (including the number of layers, hidden units, filter sizes, and attention mechanisms) for the GRU, Stacked GRU, TCN, and Transformer architectures. Furthermore, to ensure a fair comparative analysis, the common optimization settings, such as the loss function, optimizer, and batch size applied during the training phase, are also specified.

Table 2. Hyperparameter configuration and detailed model architectures.

Architecture/ Hyperparameter	GRU	Stacked GRU	TCN	Transformers
Input Processing	Masking (value = 0.0)	Masking (value = 0.0)	Masking (value = 0.0)	Layer Normalization
Feature Extraction	2 GRU Layers (64 & 32 units)	3 GRU Layers (64, 32, 16 units) with Batch Normalization	1 TCN Block (64 filters, kernel = 3)	Multi-Head Atten-tion (Heads = 4, Key dim = 32) Residual Connection (Add) Global Average Pooling 1D
Dense Layer	16 units (ReLU)	16 units (ReLU)	16 units (ReLU)	32 units (ReLU)
Output Layer	1 unit (Sigmoid)	1 unit (Sigmoid)	1 unit (Sigmoid)	1 unit (Sigmoid)
Optimizer	Adam (LR = 0.001)	Adam (LR = 0.001)	Adam (LR = 0.001)	Adam (LR = 0.001)
Loss Function	Binary Cross-Entropy	Binary Cross-Entropy	Binary Cross-Entropy	Binary Cross-Entropy
Batch Size	32	32	32	32

3.4.4. Performance Evaluation

Finally, model performance was assessed on the test set (unseen during training) using a comprehensive comparison with standard metrics for imbalanced classification: Area Under the ROC Curve (AUC-ROC), Accuracy, Precision, Recall, Specificity, and the F1-Score.

4. Results

This section presents the experimental findings obtained from the evaluation of the proposed deep learning framework. The results are categorized into four subsections: quantitative model performance, global feature importance analysis, detailed clinical interpretability of non-linear relationships, and a benchmarking comparison against state-of-the-art studies.

4.1. Model Performance Evaluation

The predictive capabilities of the four architectures (GRU, Stacked GRU, TCN, and Transformer) were evaluated on the held-out test set using the ensemble strategy described in the methodology. Table 3 summarizes the performance metrics for each architecture.

Table 3. Comparative performance metrics of the deep learning architecture ensembles on the test set.

Architecture	AUC (95% CI)	Accuracy	Precision	Recall	Specificity	F1-Score
Stacked GRU	0.912 (0.888–0.934)	0.845	0.826	0.881	0.807	0.853
GRU	0.899 (0.876–0.922)	0.823	0.819	0.839	0.807	0.829
TCN	0.874 (0.846–0.899)	0.799	0.824	0.770	0.829	0.796
Transformer	0.827 (0.796–0.856)	0.741	0.748	0.743	0.738	0.746

As evidenced in Table 3, the Stacked GRU Architecture Ensemble achieved the highest performance across almost all evaluation metrics. It attained a leading AUC of 0.912,

significantly outperforming the Transformer ensemble and showing a decisive advantage over the standard GRU. Notably, the Stacked GRU demonstrated the highest Recall (0.881). In the context of ICU readmission, high recall is critical as it minimizes false negatives preventing high-risk patients from being incorrectly classified as stable. The TCN ensemble, while achieving the highest Specificity (0.829), exhibited a lower Recall (0.770), suggesting a more conservative behavior in flagging readmission risks.

Figure 4 illustrates the Receiver Operating Characteristic (ROC) curves for the four architecture ensembles. The visual comparison corroborates the tabular results, with the Stacked GRU curve (orange line) consistently dominating the space closer to the top-left corner, indicating superior sensitivity and specificity trade-offs across various decision thresholds.

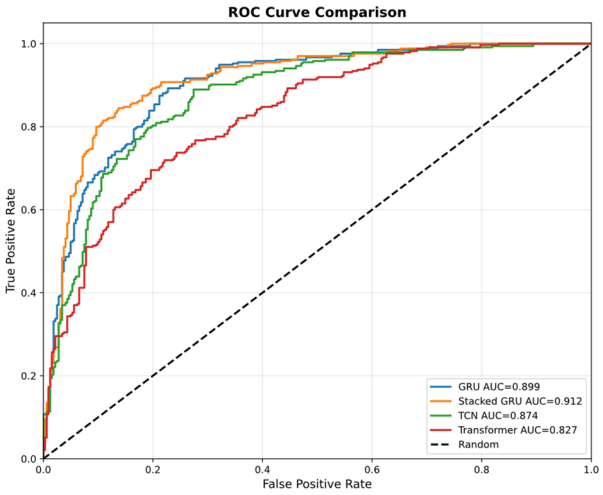

Figure 4. ROC Curve comparison of the four deep learning architecture ensembles.

To provide a more granular view of the classification performance, Figure 5 presents the Confusion Matrix for the optimal framework (Stacked GRU Architecture Ensemble). As shown, the model successfully minimized the rate of False Negatives, which is of paramount importance in clinical settings to ensure patient safety during the discharge transition.

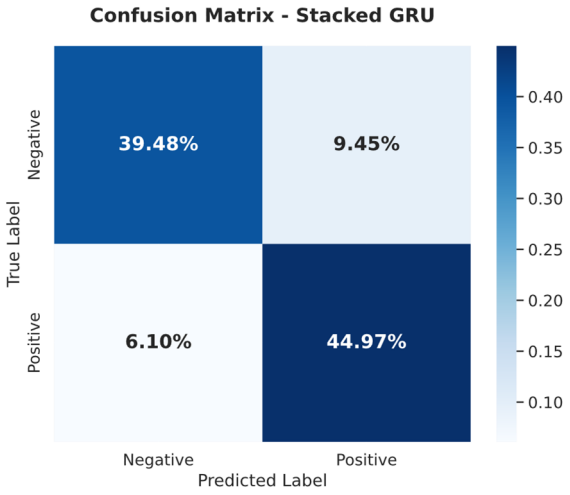

Figure 5. Confusion Matrix for architecture ensemble Stacked GRU.

To further justify the selection of the ensemble architecture over a single-model approach, a stability analysis was performed by comparing the Mean Squared Error (MSE) of standalone networks against the final ensemble (see Table 4). The results demonstrate that

the Stacked GRU Ensemble significantly reduces the predictive error from 0.143 to 0.116 (an 18.9% reduction). This comparison confirms that the ensemble framework effectively mitigates the stochastic instability inherent in individual training runs, providing a more robust and reproducible decision boundary for identifying high-risk patients.

Table 4. Comparison of architecture metrics for a single model and an assembled model.

Architecture	MSE (Single Model)	MSE (Full Ensemble)	Error Reduction (%)
Stacked GRU	0.1433	0.1157	19.20%
GRU	0.1825	0.1274	30.21%
TCN	0.1577	0.1450	8.06%
Transformer	0.1802	0.1707	5.27%

4.2. Global Feature Importance

SHAP explanations were computed on the model's final encoded feature representations after temporal processing, rather than on raw time-step inputs. Thus, the interpretability analysis reflects the contribution of temporally learned feature embeddings to the prediction output.

To understand the clinical drivers behind the predictions, we employed SHapley Additive exPlanations (SHAP) [27,28]. Figure 6 displays the ranking of the top features based on their mean absolute SHAP values, reflecting the global importance of each variable in the model's decision-making process.

Length of Stay emerged as the most influential predictor, followed by Age and MELD Score (Model for End-Stage Liver Disease). Hematological and inflammatory markers such as Monocytes and White Blood Cells, along with specific neurological indicators like Hydrocephalus and GCS Eye response, also played significant roles in the risk assessment.

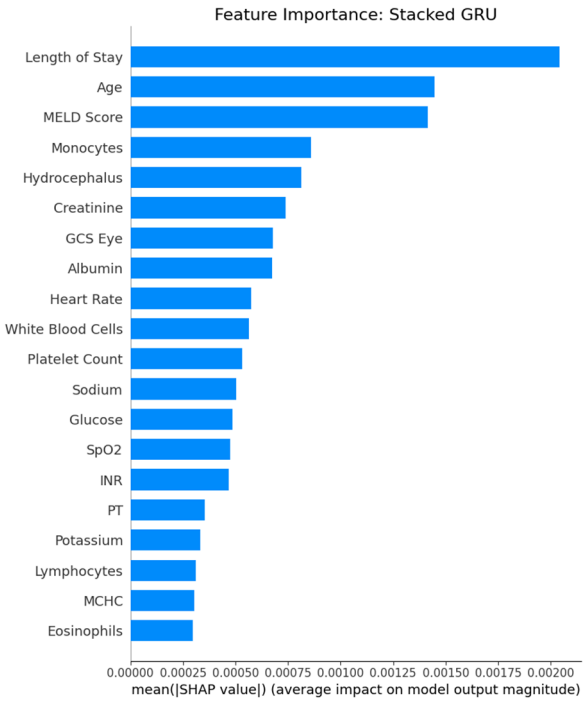

Figure 6. Global Feature Importance plot for the Stacked GRU model based on mean absolute SHAP values.

Figure 7 presents the SHAP summary plot, which provides a granular view of how the value of each feature impacts the prediction output. Each dot represents a patient

instance; red dots indicate high feature values, while blue dots indicate low values. Key observations include:

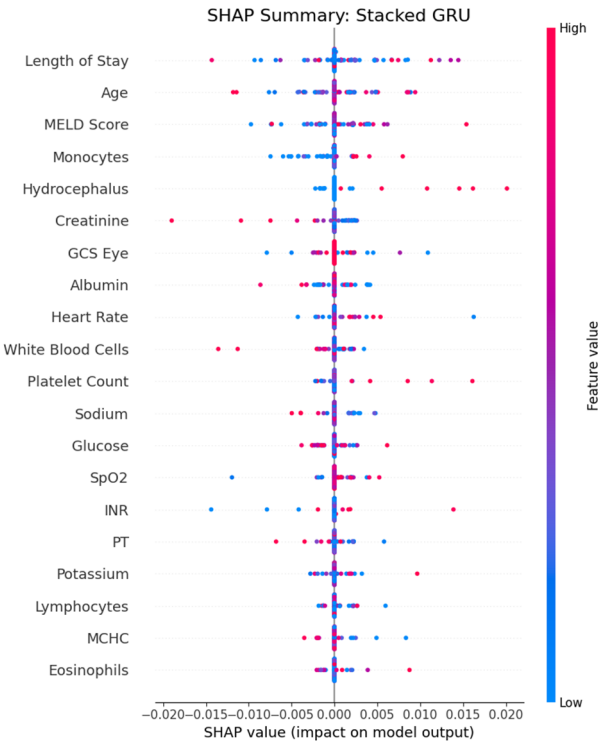

Figure 7. SHAP Summary plot for the Stacked GRU model. The x-axis represents the SHAP value (impact on model output), where positive values indicate a higher risk of ICU readmission. The color scale represents the feature value (Red = High, Blue = Low).

- Length of Stay & MELD Score: A clear positive correlation is observed; higher values (red dots) are associated with positive SHAP values, indicating an increased risk of readmission.
- GCS Eye: Lower scores (blue dots, indicating reduced consciousness) correlate with positive SHAP values, linking neurological impairment to higher readmission probability.
- Inflammatory Markers: Elevated levels of Monocytes and White Blood Cells tend to push the model towards predicting a higher risk, aligning with the physiological understanding of inflammation as a driver of clinical deterioration.

4.3. Clinical Analysis of Key Predictors

To further validate the physiological coherence of the model, we analyzed the functional relationship between the top predictors and the predicted risk. Figure 8 illustrates the SHAP dependence plots for the most significant variables. Unlike summary statistics, these plots reveal the specific trajectories of risk. For instance, they visualize how the probability of readmission evolves non-linearly with Length of Stay, identifying critical time windows where discharge decisions are most precarious. Similarly, they demonstrate the progressive increase in risk associated with elevated MELD scores and inflammatory markers, providing clinicians with interpretable thresholds that link systemic dysfunction to patient instability.

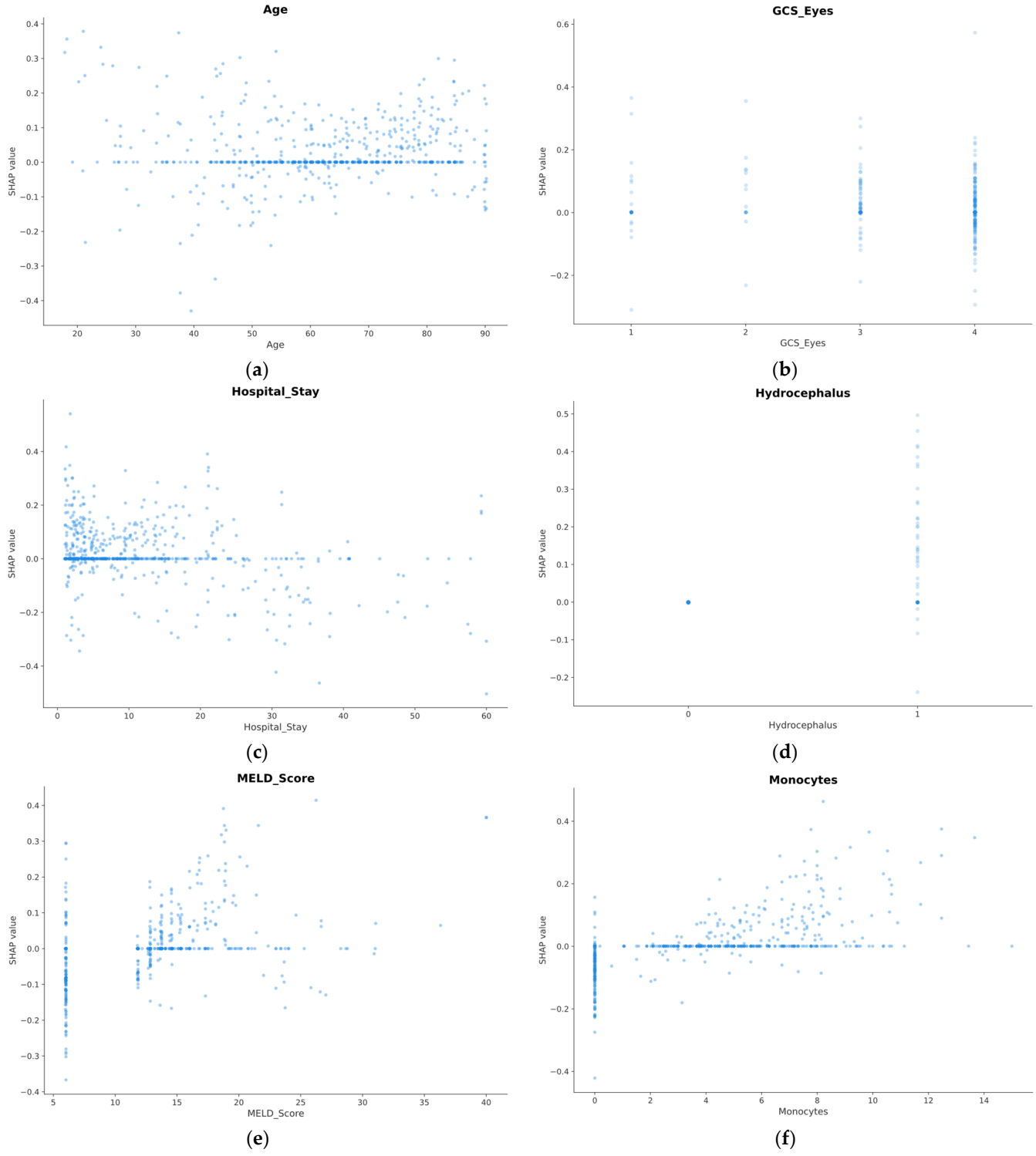

Figure 8. SHAP Dependence Plots for the top influential features: (**a**) Length of Stay, (**b**) MELD Score, (**c**) Monocytes, (**d**) Age, (**e**) MELD Score, and (**f**) Monocytes. The x-axis represents the actual value of the feature, while the y-axis represents its SHAP value (impact on model output). Positive SHAP values indicate an increased contribution to readmission risk. These visualizations highlight the non-linear dependencies captured by the model, such as the specific risk inflection points for prolonged hospitalizations and systemic organ dysfunction.

4.4. Comparison with Published Models

To validate the effectiveness of our proposed framework, we performed a quantitative comparison with the most recent relevant studies on ICU readmission predic-

tion for ICH patients [8,9]. Table 5 summarizes the methodological differences and key performance metrics.

Table 5. Comparison of the best proposed method with the published models.

Study	Model Architecture	Data Type	Balancing Technique	AUC	Accuracy	Sensitivity	Specificity
Proposed Method	Stacked GRU Ensemble	Time-Series	TimeGAN	0.912	0.845	0.881	0.807
[8]	LightGBM (Tree-based)	Static Features	SMOTEENN	0.736	0.862	0.226	0.943
[9]	ANN (Feed-forward)	Static Features	ADASYN	0.899	0.881	0.893	0.796

As shown in Table 5, the Stacked GRU Architecture Ensemble achieved the highest overall discriminative capacity with an AUC of 0.912, surpassing the LightGBM model by [8] (+17.6%) and the ANN model by [9] (+1.3%).

While [8]. achieved high specificity (0.943), their model exhibited critically low sensitivity (0.226), failing to identify the majority of readmission cases. Conversely, Ref. [9] reported high sensitivity (0.893) but lower specificity (0.796). Our proposed method demonstrates the most robust balance, achieving a superior AUC and maintaining high performance (>0.80) across both sensitivity and specificity, which is essential for minimizing both false negatives (unsafe discharges) and false positives (resource strain).

5. Discussion

Current literature addressing ICU readmissions specifically for ICH patients remains limited. Early retrospective studies [29] identified clinical correlates like dysphagia but were constrained by single-center designs and small sample sizes, limiting their generalizability. To overcome these data limitations, recent works [8,9] have leveraged the MIMIC databases, applying machine learning algorithms to larger cohorts. However, these studies primarily relied on static features extracted from the first 24 h, effectively overlooking the temporal evolution of patient stability. Consequently, despite these advances, a gap remains for a dynamic prediction framework that fully utilizes high-resolution time-series data to capture the non-linear trajectories of physiological deterioration.

This study successfully developed and validated a robust deep learning framework for predicting ICU readmission in patients with Intracerebral Hemorrhage (ICH), utilizing the extensive MIMIC-III and MIMIC-IV databases. By implementing a Stacked GRU Architecture Ensemble balanced with TimeGAN, our approach achieved a state-of-the-art AUC of 0.912, demonstrating superior predictive capability compared to traditional machine learning models reliant on static features.

A critical finding of this research is the definitive advantage of utilizing time-series data over static snapshots. While previous studies [8,9] relied on aggregated statistics (e.g., mean or max values) from the first 24 h, our results confirm that capturing the temporal trajectory of physiological signals is essential for identifying high-risk patients. A patient with stable but abnormal values is clinically distinct from one with highly volatile fluctuations, yet static models often fail to differentiate them. By leveraging a dynamic architecture, our framework explicitly captures these non-linear trajectories, which elucidates the performance gap between our method and traditional static baselines.

The Stacked GRU architecture effectively modeled the dynamic evolution of vital signs, allowing for the detection of subtle deterioration patterns that static models inherently miss. This methodological shift not only improved overall discrimination but also significantly enhanced sensitivity (0.881), reducing the risk of unsafe premature discharges. An interesting finding of our comparative analysis was the performance disparity between

the Stacked GRU (AUC 0.912) and the Transformer model (AUC 0.827). While Transformers excel in capturing long-range dependencies in massive datasets, they lack the inductive bias for sequential processing inherent in Recurrent Neural Networks (RNNs). In clinical datasets of moderate size, such as the one used in this study (n = 2117), Transformers are prone to overfitting and may struggle to generalize effectively compared to GRUs, which are architecturally optimized to model the temporal continuity and short-to-medium term dependencies typical of ICU physiological trends.

From a clinical perspective, the practical value of this framework relies on its high sensitivity (88.1%), which effectively minimizes the risk of unsafe premature discharges. Unlike static models that may fail to detect subtle physiological deterioration, this dynamic approach captures the trajectory of patient instability. Consequently, the model serves as a robust safety tool in the discharge decision-making process, ensuring that patients with hidden risk patterns are flagged for continued observation rather than being transferred to a general ward where monitoring is less intensive.

The interpretability analysis using SHAP values further validated the clinical relevance of the model, identifying six key predictors that align with and extend current pathophysiological understanding. Length of Stay, the leading predictor, exhibits a non-linear risk pattern that corroborates the dilemma of discharge timing, suggesting that both premature stability and prolonged complexity are markers of vulnerability [30,31].

The model strongly emphasized the renal-hepatic axis, selecting the MELD Score as a top predictor. This finding underscores the critical influence of systemic organ dysfunction on ICH prognosis; elevated MELD scores reflect latent coagulopathy and impaired metabolic clearance, which can exacerbate perihematomal edema and increase the risk of re-bleeding or systemic complications after ICU discharge [32,33]. Similarly, the identification of Monocytes as a key predictor aligns with the pathophysiology of secondary brain injury. While monocytes are essential for hematoma clearance, an excessive or prolonged inflammatory surge indicates persistent physiological stress and blood–brain barrier disruption markers of instability that static evaluations often miss [34,35]. Age was confirmed as a compounding factor of frailty, aligning with established demographic risk stratifications [36,37]. Finally, the identification of Hydrocephalus validates the model's sensitivity to structural complications, mirroring seminal works and recent predictive models that establish it as a robust determinant of poor functional outcome [38,39].

6. Conclusions

This study addresses the critical challenge of anticipating ICU readmissions in patients with Intracerebral Hemorrhage (ICH) by shifting the predictive paradigm from static snapshots to dynamic temporal modeling. By successfully integrating a Stacked GRU Architecture Ensemble with TimeGAN for data augmentation, we established a robust framework that captures the non-linear trajectories of patient deterioration, achieving a superior AUC of 0.912 on the extensive MIMIC datasets.

Our results confirm that the temporal evolution of vital signs and laboratory markers holds far greater prognostic value than admission baselines alone. Clinically, the model's reliance on predictors such as the MELD Score, Creatinine, and Monocytes underscores that readmission is not solely a neurological failure but a systemic collapse involving renal-hepatic dysfunction and inflammatory surges. By identifying these latent risks with high sensitivity (88.1%), this framework provides a viable pathway to reduce premature discharges, potentially improving long-term outcomes and resource allocation in neurocritical care.

Future investigations will prioritize prospective validation in multi-center cohorts to confirm generalizability. Future work should assess the generalizability of the proposed

framework in geographically and demographically diverse clinical settings. Additionally, we plan to conduct detailed ablation studies to strictly quantify the individual contributions of the TimeGAN augmentation and the Stacked architecture components. We will also assess the model's robustness across restricted time windows to identify the earliest reliable point for discharge prediction, integrating real-time waveform analysis to further refine predictive granularity. Although the present evaluation focused primarily on discrimination metrics to ensure comparability with prior studies, calibration performance was not formally assessed. Future research should incorporate calibration analysis (e.g., Brier Score, calibration slope, reliability curves) to further evaluate the clinical reliability of predicted probabilities prior to deployment. Moreover, these future investigations will also address a qualitative error analysis of misclassified cases to identify specific clinical phenotypes associated with model failure.

Author Contributions: Conceptualization, S.C.-B. and C.M.T.-G.; methodology, S.C.-B.; software, S.C.-B.; validation, S.C.-B., A.P.-R. and R.H.-L.; formal analysis, S.C.-B.; investigation, S.C.-B. and A.P.-R.; resources, C.M.T.-G. and R.H.-L.; data curation, S.C.-B.; writing—original draft preparation, S.C.-B.; writing—review and editing, C.M.T.-G., A.P.-R. and R.H.-L.; visualization, S.C.-B.; supervision, C.M.T.-G.; project administration, C.M.T.-G.; funding acquisition, S.C.-B. and C.M.T.-G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was co-financed by the Canary Islands Agency of Research, Innovation, and Information Society of the Ministry of Universities, Science, and Innovation and Culture, and by the European Social Fund Plus (ESF+) Integrated Operational Programme for the Canary Islands 2021–2027, Axis 3 Priority Theme 74 (85%) under Grant FPI2024010105.

Institutional Review Board Statement: The use of the MIMIC-III and MIMIC-IV databases was approved by the Institutional Review Boards of the Beth Israel Deaconess Medical Center (BIDMC) and the Massachusetts Institute of Technology (MIT). The requirement for individual patient consent was waived as the data is anonymized in accordance with the Health Insurance Portability and Accountability Act (HIPAA) standards. The authors have completed the required ethical training (CITI Data or Specimens Only Research) and obtained certification to access and utilize these datasets (Certification Number: 67993327).

Informed Consent Statement: Patient consent was waived due to the retrospective nature of this study, which utilized de-identified data from the MIMIC-III and MIMIC-IV databases.

Data Availability Statement: Some or all of the framework features that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Qureshi, A.I.; Tuhrim, S.; Broderick, J.P.; Batjer, H.H.; Hondo, H.; Hanley, D.F. Spontaneous Intracerebral Hemorrhage. *N. Engl. J. Med.* **2001**, *344*, 1450–1460. [CrossRef] [PubMed]
2. Feigin, V.L.; Norrving, B.; Mensah, G.A. Global Burden of Stroke. *Circ. Res.* **2017**, *120*, 439–448. [CrossRef]
3. Hemphill, J.C., 3rd; Greenberg, S.M.; Anderson, C.S.; Becker, K.; Bendok, B.R.; Cushman, M.; Fung, G.L.; Goldstein, J.N.; Macdonald, R.L.; Mitchell, P.H.; et al. Guidelines for the Management of Spontaneous Intracerebral Hemorrhage: A Guideline for Healthcare Professionals from the American Heart Association/American Stroke Association. *Stroke* **2015**, *46*, 2032–2060. [CrossRef] [PubMed]
4. Fernando, S.M.; Qureshi, D.; Talarico, R.; Tanuseputro, P.; Dowlatshahi, D.; Sood, M.M.; Smith, E.E.; Hill, M.D.; McCredie, V.A.; Scales, D.C.; et al. Intracerebral Hemorrhage Incidence, Mortality, and Association With Oral Anticoagulation Use: A Population Study. *Stroke* **2021**, *52*, 1673–1681. [CrossRef]
5. Hiller, M.; Burisch, C.; Wittmann, M.; Bracht, H.; Kaltwasser, A.; Bakker, J. The current state of intensive care unit discharge practices—Results of an international survey study. *Front. Med.* **2024**, *11*, 1377902. [CrossRef]

6. Papadakis, E.; Proklou, A.; Kokkini, S.; Papakitsou, I.; Konstantinou, I.; Konstantinidi, A.; Prinianakis, G.; Intzes, S.; Symeonidou, M.; Kondili, E. ICU Readmission and In-Hospital Mortality Rates for Patients Discharged from the ICU—Risk Factors and Validation of a New Predictive Model: The Worse Outcome Score (WOScore). *J. Pers. Med.* **2025**, *15*, 479. [CrossRef]
7. Yau, F.-F.F.; Chiu, I.-M.; Wu, K.-H.; Cheng, C.-Y.; Lee, W.-C.; Chen, H.-C.; Cheng, C.-I.; Chen, T.-Y. Machine learning-based prediction of coronary care unit readmission: A multihospital validation study. *Digit. Health* **2024**, *10*, 20552076241277030. [CrossRef]
8. Miao, J.; Zuo, C.; Cao, H.; Gu, Z.; Huang, Y.; Song, Y.; Wang, F. Predicting ICU readmission risks in intracerebral hemorrhage patients: Insights from machine learning models using MIMIC databases. *J. Neurol. Sci.* **2024**, *456*, 122849. [CrossRef] [PubMed]
9. Chen, S.; Fan, J.; Abdollahi, A.; Ashrafi, N.; Alaei, K.; Placencia, G.; Pishgar, M. Machine Learning-Based Prediction of ICU Readmissions in Intracerebral Hemorrhage Patients: Insights from the MIMIC Databases. In *Lecture Notes in Networks and Systems*; Springer: Cham, Switzerland, 2025; pp. 454–470. [CrossRef]
10. Kim, J.; Kim, H.; Kim, H.G.; Lee, D.; Yoon, S. A comprehensive survey of deep learning for time series forecasting: Architectural diversity and open challenges. *Artif. Intell. Rev.* **2025**, *58*, 216. [CrossRef]
11. Rasheed, N.; Khaliq, U.; Ahmed, A.; Bermak, A. Predicting Patient ICU Readmission Using Recurrent Neural Networks With Long Short-Term Memory. *IEEE Access* **2025**, *13*, 153375–153385. [CrossRef]
12. Tungal, A.; Singh, K.; Singh, P.; Rehman, A.U.; Sood, S.; Kant, V.; Kumar, A.; Hussen, S.; Hamam, H. Recurrent neural network-based automated early detection of pandemic-prone diseases through symptoms analysis. *Discov. Soc. Sci. Health* **2025**, *5*, 69. [CrossRef]
13. Bednarski, B.P.; Singh, A.D.; Zhang, W.; Jones, W.M.; Naeim, A.; Ramezani, R. Temporal convolutional networks and data rebalancing for clinical length of stay and mortality prediction. *Sci. Rep.* **2022**, *12*, 21247. [CrossRef]
14. Apalak, M.; Kiasaleh, K. Advancing Early Detection of Sepsis With Temporal Convolutional Networks Using ECG Signals. *IEEE Access* **2024**, *12*, 3417–3427. [CrossRef]
15. Rong, R.; Gu, Z.; Lai, H.; Nelson, T.L.; Keller, T.; Walker, C.; Jin, K.W.; Chen, C.; Navar, A.M.; Velasco, F.; et al. A deep learning model for clinical outcome prediction using longitudinal inpatient electronic health records. *JAMIA Open* **2025**, *8*, ooaf026. [CrossRef]
16. Chang, P.; Li, H.; Quan, S.F.; Lu, S.; Wung, S.-F.; Roveda, J.; Li, A. A transformer-based diffusion probabilistic model for heart rate and blood pressure forecasting in Intensive Care Unit. *Comput. Methods Programs Biomed.* **2024**, *246*, 108060. [CrossRef] [PubMed]
17. Saleh, H.; McCann, M.; El-Sappagh, S.; Breslin, J.G. TransformerFusionNet: A Real-Time Multimodal Framework for ICU Heart Failure Mortality Prediction Using Big Data Streaming. In Proceedings of the 2024 International Conference on Computer and Applications, ICCA 2024, Cairo, Egypt, 17–19 December 2024. [CrossRef]
18. Johnson, A.E.W.; Pollard, T.J.; Shen, L.; Lehman, L.-W.H.; Feng, M.; Ghassemi, M.; Moody, B.; Szolovits, P.; Celi, L.A.; Mark, R.G. MIMIC-III, a freely accessible critical care database. *Sci. Data* **2016**, *3*, 160035. [CrossRef]
19. Johnson, A.E.W.; Bulgarelli, L.; Shen, L.; Gayles, A.; Shammout, A.; Horng, S.; Pollard, T.J.; Hao, S.; Moody, B.; Gow, B.; et al. MIMIC-IV, a freely accessible electronic health record dataset. *Sci. Data* **2023**, *10*, 1. [CrossRef]
20. Wang, H.L.; Hsu, W.Y.; Lee, M.H.; Weng, H.H.; Chang, S.W.; Yang, J.T.; Tsai, Y.H. Automatic machine-learning-based outcome prediction in patients with primary intracerebral hemorrhage. *Front. Neurol.* **2019**, *10*, 476100. [CrossRef]
21. Trevisi, G.; Caccavella, V.M.; Scerrati, A.; Signorelli, F.; Salamone, G.G.; Orsini, K.; Fasciani, C.; D'aRrigo, S.; Auricchio, A.M.; D'oNofrio, G.; et al. Machine learning model prediction of 6-month functional outcome in elderly patients with intracerebral hemorrhage. *Neurosurg. Rev.* **2022**, *45*, 2857–2867. [CrossRef]
22. Yan, C.; Zhang, Z.; Nyemba, S.; Li, Z. Generating Synthetic Electronic Health Record Data Using Generative Adversarial Networks: Tutorial. *JMIR AI* **2024**, *3*, e52615. [CrossRef]
23. Celada-Bernal, S.; Pérez-Acosta, G.; Travieso-González, C.M.; Blanco-López, J.; Santana-Cabrera, L. Applying Neural Networks to Recover Values of Monitoring Parameters for COVID-19 Patients in the ICU. *Mathematics* **2023**, *11*, 3332. [CrossRef]
24. Liu, Y.; Song, Z.; Xu, X.; Rafique, W.; Zhang, X.; Shen, J.; Khosravi, M.R.; Qi, L. Bidirectional GRU networks-based next POI category prediction for healthcare. *Int. J. Intell. Syst.* **2022**, *37*, 4020–4040. [CrossRef]
25. Xia, M.; Shao, H.; Ma, X.; De Silva, C.W. A Stacked GRU-RNN-Based Approach for Predicting Renewable Energy and Electricity Load for Smart Grid Operation. *IEEE Trans. Industr. Inform.* **2021**, *17*, 7050–7059. [CrossRef]
26. Biffi, C.; Roffo, G.; Salvagnini, P.; Cherubini, A. A temporal convolutional network-based approach and a benchmark dataset for colonoscopy video temporal segmentation. *Comput. Methods Programs Biomed.* **2025**, *270*, 108782. [CrossRef]
27. Yuan, Z.-N.; Xue, Y.-J.; Wang, H.-J.; Qu, S.-N.; Huang, C.-L.; Wang, H.; Zhang, H.; Zhang, M.-Z.; Xing, X.-Z. A predictive model for hospital death in cancer patients with acute pulmonary embolism using XGBoost machine learning and SHAP interpretation. *Sci. Rep.* **2025**, *15*, 18268. [CrossRef]
28. Qi, X.; Wang, S.; Fang, C.; Jia, J.; Lin, L.; Yuan, T. Machine learning and SHAP value interpretation for predicting comorbidity of cardiovascular disease and cancer with dietary antioxidants. *Redox Biol.* **2025**, *79*, 103470. [CrossRef]

29. Tangonan, R.; Alvarado-Dyer, R.; Loggini, A.; El Ammar, F.; Kumbhani, R.; Lazaridis, C.; Kramer, C.; Goldenberg, F.D.; Mansour, A. Frequency, Risk Factors, and Outcomes of Unplanned Readmission to the Neurological Intensive Care Unit after Spontaneous Intracerebral Hemorrhage. *Neurocritical Care* **2022**, *37*, 390–398. [CrossRef] [PubMed]
30. Loggini, A.; Mansour, A.; El Ammar, F.; Lazaridis, C.; Kramer, C.L.; Bulwa, Z.; Velez, F.S.; McCoy, C.; Goldenberg, F.D. Early Determinants of Neurocritical Care Unit Length of Stay in Patients with Spontaneous Intracerebral Hemorrhage. *Neurocritical Care* **2020**, *34*, 485–491. [CrossRef] [PubMed]
31. Loggini, A.; Del Brutto, V.J.; Qureshi, A.I.; Hornik, J.; Wallery, S.S.; Schwertman, A.; Nomani, S.; Hornik, A.; Velez, F.G.S. Medical complications associated with prolonged length of stay in patients with nontraumatic intracerebral hemorrhage: A nationwide cohort study. *Clin. Neurol. Neurosurg.* **2025**, *254*, 108934. [CrossRef]
32. Luo, H.; Yang, X.; Chen, K.; Lan, S.; Liao, G.; Xu, J. Blood creatinine and urea nitrogen at ICU admission and the risk of in-hospital death and 1-year mortality in patients with intracranial hemorrhage. *Front. Cardiovasc. Med.* **2022**, *9*, 967614. [CrossRef] [PubMed]
33. Morbitzer, K.A.; Jordan, J.D.; Dehne, K.A.; Durr, E.A.; Olm-Shipman, C.M.; Rhoney, D.H. Enhanced Renal Clearance in Patients with Hemorrhagic Stroke. *Crit. Care Med.* **2019**, *47*, 800–808. [CrossRef]
34. Mackey, J.; Blatsioris, A.D.; Saha, C.; Moser, E.A.S.; Carter, R.J.L.; Cohen-Gadol, A.A.; Leipzig, T.J.; Williams, L.S. Higher Monocyte Count is Associated with 30-Day Case Fatality in Intracerebral Hemorrhage. *Neurocritical Care* **2020**, *34*, 456–464. [CrossRef]
35. Lattanzi, S.; Cagnetti, C.; Provinciali, L.; Silvestrini, M. Neutrophil-to-Lymphocyte Ratio Predicts the Outcome of Acute Intracerebral Hemorrhage. *Stroke* **2016**, *47*, 1654–1657. [CrossRef] [PubMed]
36. Kim, B.D.; Kurian, C.; Stein, L.K.; Tuhrim, S.; Dhamoon, M.S. Index Admission Characteristics and All-Cause Readmissions Analysis in Younger and Older Adults with Intracerebral Hemorrhage. *Cerebrovasc. Dis.* **2020**, *49*, 375–381. [CrossRef] [PubMed]
37. Foschi, M.; D'aNna, L.; Gabriele, C.; Conversi, F.; Gabriele, F.; Orlandi, B.; De Santis, F.; Ornello, R.; Sacco, S. Sex Differences in the Epidemiology of Intracerebral Hemorrhage Over 10 Years in a Population-Based Stroke Registry. *J. Am. Heart Assoc.* **2024**, *13*, 32595. [CrossRef] [PubMed]
38. Diringer, M.N.; Edwards, D.F.; Zazulia, A.R. Hydrocephalus: A previously unrecognized predictor of poor outcome from supratentorial intracerebral hemorrhage. *Stroke* **1998**, *29*, 1352–1357. [CrossRef] [PubMed]
39. Chen, Q.; Tang, J.; Tan, L.; Guo, J.; Tao, Y.; Li, L.; Chen, Y.; Liu, X.; Zhang, J.H.; Chen, Z.; et al. Intracerebral hematoma contributes to hydrocephalus after intraventricular hemorrhage via aggravating iron accumulation. *Stroke* **2015**, *46*, 2902–2908. [CrossRef]