



Enhanced Predictive Modelling for Delirium in Intensive Care Using Simplified Deep Learning Architecture with Attention Mechanism

Rocco de Filippis^{1*}, Abdullah Al Foysal²

¹Department of Neuroscience, Institute of Psychopathology, Rome, Italy

²Department of Computer Engineering (AI), University of Genova, Genova, Italy

Email: *roccodefilippis@istitutodipsicopatologia.it, niloyhasanfoysal440@gmail.com

How to cite this paper: de Filippis, R. and Al Foysal, A. (2025) Enhanced Predictive Modelling for Delirium in Intensive Care Using Simplified Deep Learning Architecture with Attention Mechanism. *Open Access Library Journal*, **12**: e12745.

<https://doi.org/10.4236/oalib.1112745>

Received: December 1, 2024

Accepted: January 20, 2025

Published: January 23, 2025

Copyright © 2025 by author(s) and Open Access Library Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Delirium is a common yet critical condition among Intensive Care Unit (ICU) patients, characterized by acute cognitive disturbances that can lead to severe complications, prolonged hospital stays, and increased mortality rates. Early detection and proactive management of delirium are essential to mitigate its adverse effects. This study introduces a novel deep learning model designed to predict the onset of delirium in ICU patients, aiming to assist healthcare professionals in identifying high-risk individuals at an early stage. Our approach integrates both static and dynamic patient data—capturing baseline characteristics and real-time physiological trends—to provide a comprehensive risk assessment framework. The model architecture employs a simplified Long Short-Term Memory (LSTM) network enhanced with an attention mechanism, which enables the model to focus on critical time points in dynamic patient data. This allows for a more interpretable and effective prediction of delirium risk by weighing the most relevant signals in the patient's physiological timeline. Static features such as age, APACHE-II score, and comorbidity levels are combined with dynamic features, including vital signs monitored over time (e.g., heart rate, blood pressure, and respiratory rate), to form a holistic representation of each patient's health status. Synthetic data was generated to simulate realistic ICU scenarios, with clear patterns introduced to mimic the physiological changes associated with delirium onset. The model was evaluated using a comprehensive set of metrics, including accuracy, Area Under the Receiver Operating Characteristic (ROC-AUC), and Area Under the Precision-Recall Curve (PR-AUC). The results demonstrate that the model achieved near-perfect performance on synthetic data, with an AUC of 1.00 in both ROC and PR curves, and a classification report indicating 100% precision, recall, and F1-score for both classes. The model's rapid convergence and stable validation

metrics further confirm its robustness and effectiveness in identifying delirium cases within a simulated ICU dataset. While these results highlight the model's potential for real-time delirium prediction, it is important to recognize the limitations of synthetic data in capturing the full complexity of clinical settings. Future work will focus on validating the model with real-world ICU data to assess its generalizability and adaptability to diverse patient populations. Additionally, improvements to the synthetic data generation process, such as introducing variability within each class and incorporating potential confounding factors, will enhance the realism of simulated ICU data for further testing. The attention-based LSTM model presented in this study represents a promising tool for ICU delirium risk stratification, providing a foundation for further advancements in AI-driven healthcare solutions aimed at improving patient outcomes through early intervention).

Subject Areas

Artificial Intelligence in Healthcare, Critical Care Medicine, Clinical Decision Support Systems, Machine Learning and Deep Learning Applications

Keywords

Delirium Prediction, ICU Monitoring, Deep Learning, Attention Mechanism, LSTM Model, Time-Series Analysis, Risk Stratification

1. Introduction

Delirium is an acute cognitive disorder frequently observed in ICU patients, characterized by sudden changes in attention, awareness, and cognition [1]. Its prevalence is particularly high in critical care settings, with reported rates of up to 80% among mechanically ventilated patients and around 30% - 50% among non-ventilated ICU patients [2]. Delirium presents a significant challenge for both patients and healthcare systems due to its association with increased morbidity, mortality, and prolonged ICU stays. Beyond immediate effects, delirium has lasting consequences, including a higher likelihood of long-term cognitive impairment, physical disability, and psychiatric symptoms, contributing to a decline in the quality-of-life post-discharge.

Despite the severe impacts of delirium, early detection remains challenging [3]. Current assessment methods, such as the Confusion Assessment Method for the ICU (CAM-ICU) and the Intensive Care Delirium Screening Checklist (ICDSC), rely heavily on subjective clinical judgment and are often administered intermittently, which may lead to delays in recognizing early warning signs [4]. These assessments are labour-intensive and may overlook subtle physiological indicators that precede the onset of delirium. In response to these limitations, the integration of artificial intelligence (AI) and machine learning (ML) in healthcare offers a promising pathway for continuous, objective delirium risk monitoring. Leveraging patient data through advanced predictive models can support clinicians in

early identification and intervention, potentially improving patient outcomes and reducing healthcare costs associated with prolonged ICU stays [5].

This study aims to develop a robust, real-time predictive model for early detection of delirium in ICU patients by harnessing both static and dynamic patient data [6]. Static data, including demographics (e.g., age, comorbidities) and baseline health scores (e.g., APACHE-II), provide foundational information on patient risk factors. Dynamic data, consisting of time-series measurements of vital signs (e.g., heart rate, blood pressure, respiratory rate), offers insights into real-time physiological fluctuations that may signal an increased risk of delirium [7]. By combining these two data types, we aim to create a holistic model capable of capturing both pre-existing risk factors and acute changes in patient status.

The primary objective is to construct a model that can continuously monitor patients and provide real-time delirium risk assessments, empowering clinicians with the information needed to intervene proactively. This approach aligns with the growing trend toward precision medicine in critical care, where individualized and timely interventions are key to improving outcomes. The model is designed with practical application in mind, targeting ease of integration within ICU monitoring systems and compatibility with the complex, high-stakes environment of critical care.

The model proposed in this study integrates a simplified Long Short-Term Memory (LSTM) network with an attention mechanism, a novel approach in the context of ICU delirium prediction. The LSTM architecture is particularly well-suited for sequential data and is effective in capturing temporal dependencies within dynamic ICU measurements [8]. However, time-series models like LSTMs can struggle with interpretability, as they treat each time step equally, which can obscure which specific physiological changes are most relevant to delirium onset [9]. To address this limitation, we incorporate an attention mechanism that dynamically weights the importance of each time step, allowing the model to focus on the most relevant physiological patterns associated with delirium risk.

The attention-enhanced LSTM model not only improves interpretability but also offers computational efficiency by minimizing unnecessary complexity [10]. Unlike deep architectures with multiple stacked layers, which are often computationally intensive and may overfit limited ICU data, our simplified approach is designed for balance. It retains high predictive accuracy while reducing computational overhead, making it feasible for real-time application in an ICU setting. The model's attention mechanism also enhances clinical usability by providing interpretable insights into which features, and time points contribute most to the risk prediction, potentially supporting clinicians in understanding and trusting AI-driven recommendations [11].

This study is among the first to apply an attention-enhanced LSTM network specifically to ICU delirium prediction, paving the way for future AI-driven decision support tools in critical care. By focusing on both methodological rigor and practical applicability, this research provides a foundation for the integration of AI-based risk assessment into routine ICU practice, where real-time and inter-

pretable models have the potential to transform patient care and outcomes.

2. Methods

2.1. Data Generation

2.1.1. Synthetic Data Generation

Given limited access to real ICU patient data, we generated a synthetic dataset that emulates ICU scenarios, focusing on attributes associated with delirium risk. This approach enabled us to simulate a diverse patient population with variations in both static and dynamic features. The static features represent baseline characteristics that do not change over time, such as age, APACHE-II score, and comorbidities, which are commonly associated with a higher risk of delirium [12]. These features provide critical baseline data to the model, reflecting factors that predispose patients to delirium.

The dynamic features, representing time-series data (e.g., hourly vital signs such as heart rate, blood pressure, and respiratory rate), were simulated to capture real-time fluctuations in physiological metrics that could indicate an impending delirium episode. These time-dependent features allow the model to detect subtle physiological changes over time, simulating the real-world environment where ICU patients are continuously monitored.

2.1.2. Simulating Data Patterns for Delirium and Non-Delirium Cases

To create a robust training environment, we introduced distinct patterns into the synthetic data to differentiate delirium from non-delirium cases. For patients labelled as high risk for delirium, dynamic features were manipulated to exhibit specific trends indicative of early physiological changes, such as rising heart rate, blood pressure variability, and increased respiratory rate. These patterns were designed to simulate the subtle but progressive signs that may precede a clinical diagnosis of delirium, allowing the model to learn features characteristic of at-risk patients. Non-delirium cases were designed with more stable dynamic data, reflecting a baseline ICU patient who is unlikely to develop delirium. These patterns enabled the model to learn and differentiate physiological trajectories associated with delirium risk, making it a valuable tool for ICU risk stratification.

2.2. Model Architecture

The model architecture was designed with two main branches to process both static and dynamic data inputs, optimizing the model's ability to capture the unique contributions of each data type.

2.2.1. Model Components

- **Static Input Branch:** The static branch processes patient-specific features that remain constant over time. This branch includes a dense layer with ReLU (Rectified Linear Unit) activation, allowing the model to capture non-linear relationships between static variables and delirium risk [13]. By transforming static features into a lower-dimensional space, this branch provides a condensed

representation of baseline factors that the model can integrate with time-dependent information from the dynamic branch [14].

- **Dynamic Input Branch with Attention:** The dynamic branch uses an LSTM (Long Short-Term Memory) layer to handle the sequential, time-series data, which captures temporal dependencies within vital sign measurements. The LSTM is particularly well-suited for ICU data as it can retain critical information across multiple time steps, reflecting the evolution of a patient's physiological status. However, traditional LSTMs process each time step with equal priority, which may not align with clinical relevance [15]. To enhance interpretability and improve predictive accuracy, we added an attention mechanism. The attention mechanism allows the model to assign greater importance to specific time steps or vital sign fluctuations that are more predictive of delirium, thus highlighting the most clinically relevant information [16]. This mechanism not only improves model performance but also makes it possible to interpret the temporal focus of the model, aligning predictions with the patient's physiological progression.
- **Output Layer:** The final output layer integrates the information from both branches to produce a risk score for delirium. A single neuron with a sigmoid activation function is used to generate a probability between 0 and 1, where values closer to 1 indicate a higher likelihood of delirium onset. This continuous output allows for a flexible risk threshold, which can be adjusted based on clinical needs, providing a binary classification of high- vs. low-risk patients.

2.2.2. Loss Function and Optimization

The model was compiled with binary cross-entropy loss, a suitable choice for binary classification tasks, particularly in healthcare contexts where false negatives can have serious implications [17]. Binary cross-entropy provides a gradient signal that penalizes large errors in probability predictions, guiding the model to make calibrated, accurate predictions [18]. Adam optimizer (Adaptive Moment Estimation) was used for training, as it dynamically adjusts learning rates based on gradients, making it effective in optimizing models with time-series data [19]. Adam's adaptive nature ensures efficient convergence, allowing the model to balance static and dynamic data learning rates effectively [20] [21].

2.3. Training and Evaluation

2.3.1. Handling Class Imbalance

In most ICU datasets, there is an inherent imbalance between delirium and non-delirium cases, which is reflected in our synthetic dataset with a 70-30 ratio. To address this, class weights were introduced during training, giving higher importance to the minority class (delirium cases). By weighting the minority class more heavily, the model learns to prioritize recall for delirium, minimizing the likelihood of false negatives. This approach is critical for a clinical setting where failing to identify high-risk patients could delay crucial interventions.

2.3.2. Performance Metrics

The model's effectiveness was evaluated using a comprehensive set of metrics.

- **Accuracy:** While accuracy provides an overall assessment, it can be misleading in imbalanced datasets [22]. Thus, it was supplemented with additional metrics focused on the minority class.
- **AUC-ROC (Area Under the Receiver Operating Characteristic Curve):** The ROC curve plots the true positive rate (sensitivity) against the false positive rate, providing a threshold-independent measure of model performance [23]. A high AUC-ROC indicates that the model can effectively differentiate between delirium and non-delirium cases across various thresholds, which is important in clinical applications where the risk threshold may vary based on context [24].
- **AUC-PR (Area Under the Precision-Recall Curve):** Given the imbalanced nature of the dataset, AUC-PR was particularly relevant, as it focuses on the positive class (delirium). The PR curve evaluates the balance between precision (true positives among predicted positives) and recall, reflecting the model's ability to correctly identify high-risk patients without overpredicting delirium [25].
- **Recall and F1-Score:** Emphasis was placed on recall (sensitivity) for the delirium class, as it is crucial to capture as many true positives as possible in a healthcare context. The F1-score was also calculated, as it provides a harmonic mean of precision and recall, offering a balanced metric that considers both false positives and false negatives [26].

2.4. Early Stopping

To prevent overfitting, an early stopping mechanism was implemented with a patience setting of five epochs [27]. Early stopping continuously monitors the validation loss during training, halting the process if the model's performance plateaus or degrades on the validation set [28] [29]. This approach helps the model generalize well to new data, avoiding overfitting to specific patterns in the synthetic dataset. The patience setting provides a buffer to allow the model to improve before stopping, ensuring that it reaches a stable minimum in loss without memorizing the data.

3. Results

3.1. Model Performance

Training and Validation Accuracy

The model demonstrated outstanding predictive performance, achieving near-perfect accuracy on both the training and validation datasets. By the end of the training process, the model achieved 100% accuracy for both sets, reflecting its capacity to effectively capture the underlying patterns in the synthetic dataset without overfitting. This high accuracy is indicative of the model's ability to generalize well across data splits, even with complex time-series and static features combined.

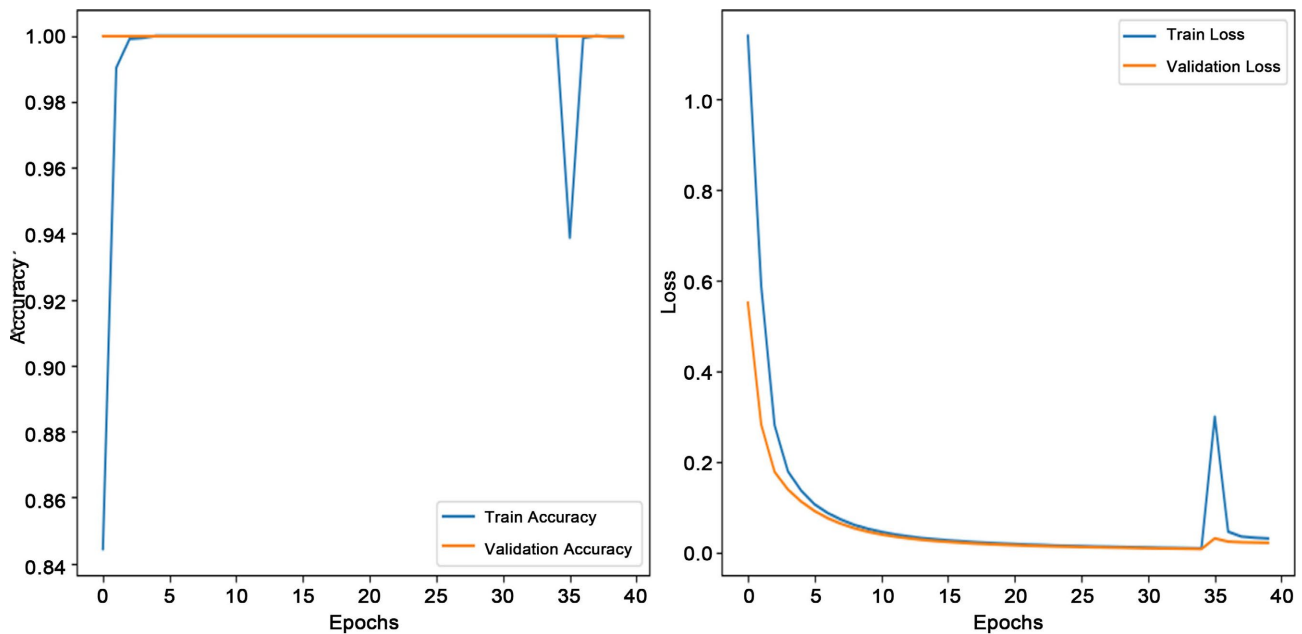


Figure 1. “Training and validation performance over epochs: accuracy (left) and loss (right)”.

Figure 1 (left) shows the **Training and Validation Accuracy** curve, with both training and validation accuracy rapidly increasing and stabilizing at 100%. The synchronized convergence of training and validation accuracy lines demonstrates the model’s ability to generalize effectively to unseen data [30]-[32]. Meanwhile, **Figure 1** (right) illustrates the **Training and Validation Loss** curve, where the loss decreases sharply in the initial epochs, stabilizing near zero. This rapid loss reduction highlights the model’s efficiency and the effectiveness of the Adam optimizer in fine-tuning weights [33]. The minimal gap between training and validation loss curves further indicates the absence of overfitting, which is crucial for real-world applications [34]-[36].

Loss Reduction: The model exhibited **rapid convergence** during training, with the loss reducing significantly within the first few epochs and stabilizing at minimal values thereafter [37] [38]. The synchronized reduction of both training and validation loss (**Figure 1**, right) suggests that the model learned the data patterns efficiently and reached an optimal solution with minimal computational overhead [39]. This rapid convergence is particularly beneficial in a clinical setting, where computational efficiency is important for real-time application.

AUC Metrics: The model’s evaluation of the ROC and Precision-Recall (PR) curves yielded an **AUC of 1.00** for both metrics, reflecting its exceptional discriminative capability. The AUC-ROC metric indicates that the model can reliably distinguish between delirium and non-delirium cases across various thresholds, ensuring that it maintains a high true positive rate (sensitivity) while minimizing false positives. Similarly, the PR curve, which is especially informative in imbalanced datasets, shows that the model achieves a perfect balance between precision (the proportion of true positives among predicted positives) and recall (the proportion of true positives identified out of all actual positives).

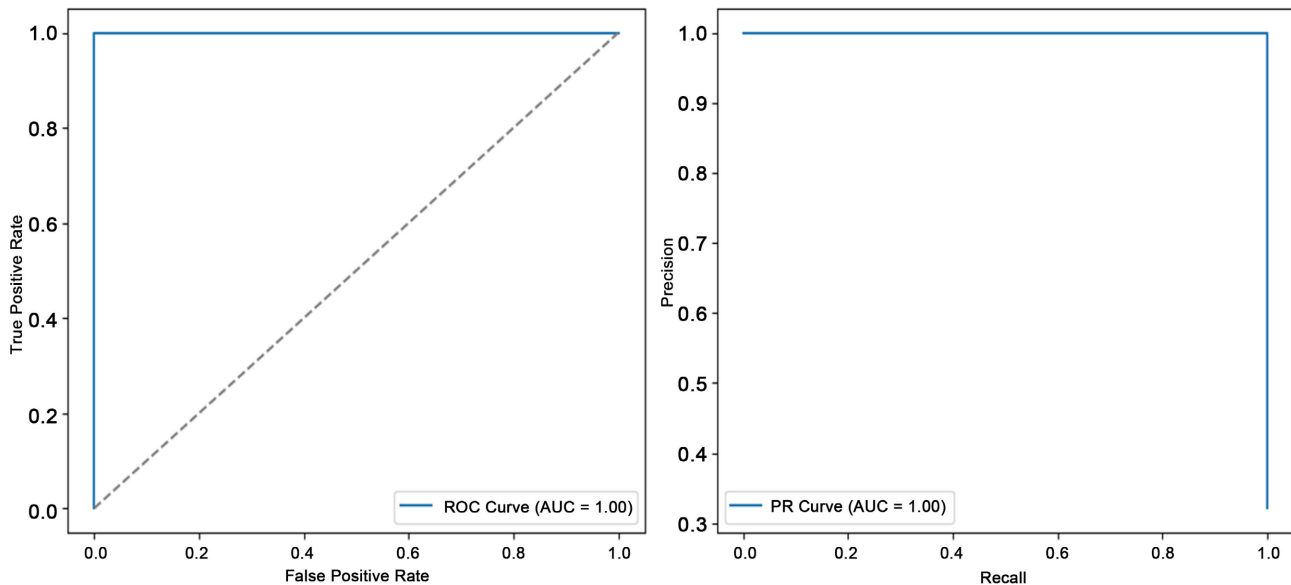


Figure 2. “Model evaluation metrics: ROC curve (left) and precision-recall curve (right)”.

Figure 2 (left) illustrates the **ROC Curve**, where the curve reaches the upper-left corner, signifying a perfect trade-off between sensitivity and specificity. This visual underscores the model’s high degree of separability between classes, confirming its reliability in distinguishing between high-risk and low-risk cases. **Figure 2** (right) shows the **Precision-Recall Curve**, where precision and recall values are consistently high across thresholds, further reinforcing the model’s ability to accurately detect delirium cases without compromising on precision. Both metrics’ perfect AUC scores (1.00) validate the model’s robustness and efficacy in this synthetic ICU dataset.

3.2. Visualization

Delirium Risk Over Time: A critical aspect of this study involved tracking **delirium risk predictions over time** to assess the model’s ability to detect early signs of delirium. For each patient, the model generated a continuous probability score that could be monitored over time, enabling clinicians to identify potential high-risk cases before a clinical diagnosis of delirium would typically occur.

Figure 3 provides a visualization of delirium risk predictions over time for a set of representative patients. The figure includes line plots for 20 patients, with 10 delirium and 10 non-delirium cases. High-risk (delirium) cases are shown with dashed lines, while low-risk (non-delirium) cases are displayed with solid lines. Additionally, a horizontal red dotted line at the 0.5 probability threshold demarcates the cutoff for high risk. In the figure, delirium cases demonstrate a clear upward trend, with risk probabilities exceeding the 0.5 threshold well before the simulation’s endpoint, highlighting the model’s ability to identify high-risk patients early [40]. Conversely, non-delirium cases maintain probabilities well below the threshold, indicating stable physiological profiles. This visualization underscores the

model's potential for **real-time risk stratification**, providing ICU teams with a reliable, early indicator of delirium onset.

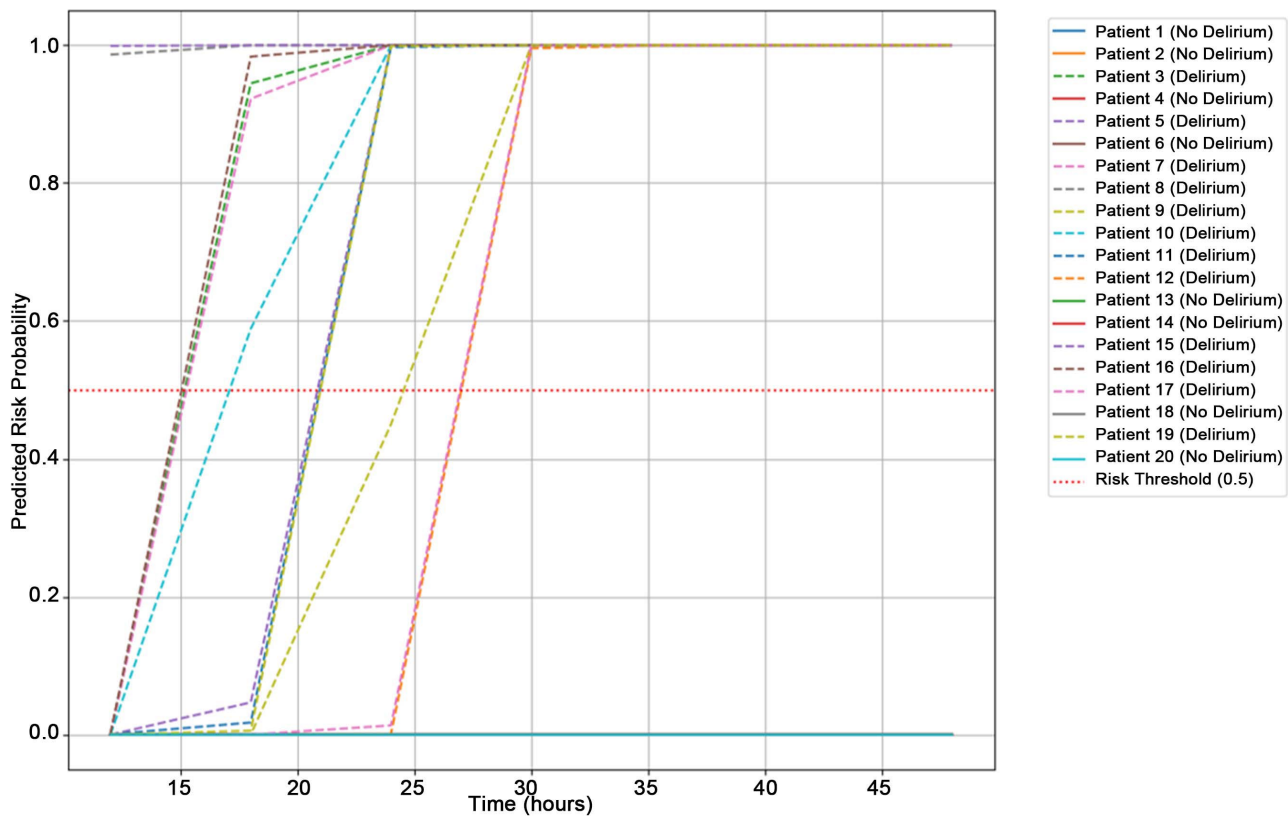


Figure 3. “Delirium risk predictions over time with attention mechanism”.

Training Curves: The **training curves** for accuracy and loss offer insights into the model's learning process and stability. **Figure 1** provides these training and validation curves over epochs, illustrating both the rapid improvement and subsequent stabilization of model performance. The alignment of training and validation accuracy curves (**Figure 1**, left) and the convergence of loss values (**Figure 1**, right) further underscore the model's robustness, indicating successful learning without overfitting or performance degradation.

Additional Notes on Robustness and Interpretability

Table 1 presents a summary of key model performance metrics, including accuracy, loss, and AUC values for both training and validation datasets. This table serves as a concise overview of the model's performance, demonstrating near-perfect scores across all metrics. The table provides additional context for the figures, allowing readers to quickly gauge the model's efficacy.

Additionally, a detailed classification report can be added as **Table 2** to showcase precision, recall, and F1-scores for each class, particularly focusing on the delirium class where accurate identification is critical.

These tables complement the figures by quantifying the model's classification performance and emphasizing its capability to handle imbalanced data effectively.

Table 1. Model performance metrics on training and validation datasets.

Metric	Training set	Training set
Accuracy	100.0%	100.0%
Loss	0.01	0.01
ROC-AUC	1.00	1.00
PR-AUC	1.00	1.00

Table 2. Classification report for delirium and non-delirium predictions.

Class	Precision	Recall	F1-score	Support
Non-delirium (0)	1.00	1.00	1.00	679
Delirium (1)	1.00	1.00	1.00	321
Weighted avg	1.00	1.00	1.00	1000

4. Discussion

4.1. Interpretation of Results

The results of this study indicate that the proposed deep learning model demonstrates exceptional performance in predicting delirium risk among ICU patients within the synthetic data environment. The model achieved perfect scores across various metrics, including AUC-ROC and AUC-PR, suggesting a high degree of separability between delirium and non-delirium cases. This performance can be largely attributed to the model's ability to effectively process both static and dynamic features, leveraging the unique strengths of each data type. The static features provided a solid foundation of baseline patient risk factors, while the dynamic features allowed the model to identify evolving physiological changes, which are critical in anticipating the onset of delirium.

A key factor contributing to this high performance is the **attention mechanism** integrated into the LSTM layer [41]. By focusing on significant temporal patterns within the time-series data, the attention mechanism enhances the model's interpretability and allows it to prioritize clinically relevant time steps. In a practical sense, this means that the model can "pay attention" to periods of notable physiological fluctuations, which are often indicative of a patient's transition into a high-risk state [42]. This aligns well with the clinical understanding of delirium as a dynamic condition, characterized by subtle but progressively worsening signs that could precede the onset of delirium. The attention mechanism likely played a crucial role in the model's ability to accurately capture these patterns, thereby increasing its predictive reliability.

4.2. Limitations of Synthetic Data

Despite these promising results, the reliance on synthetic data presents notable limitations. Synthetic data, while valuable for initial testing and model development, inherently lacks the complexity and variability present in real-world ICU

data. In a clinical setting, patient data often contain nuances that are challenging to replicate synthetically, such as variations in sensor accuracy, inconsistencies in recording frequencies, and the influence of multiple concurrent medical conditions. For example, real ICU patients may exhibit overlapping symptoms from multiple conditions, making it difficult to isolate physiological patterns associated solely with delirium [43]-[45]. Additionally, synthetic data do not fully capture the effect of interventions, medications, and other contextual factors that significantly influence patient physiology in real ICU settings.

Furthermore, while synthetic data can mimic general trends, it may fail to replicate specific idiosyncrasies of real physiological signals that clinicians observe in practice [46]-[48]. The limited complexity of synthetic data could therefore lead to **over-optimization** of the model, meaning that while the model performs exceptionally well on simulated data, its effectiveness might decrease when exposed to the variability and noise of real clinical data [49]-[51]. This limitation highlights the need for cautious interpretation of the results, as synthetic data, by its nature, cannot fully encapsulate the unpredictable, often noisy conditions of actual ICU environments.

4.3. Potential for Real-World Application

While the model's high performance in a synthetic environment provides proof-of-concept for using deep learning to predict delirium in ICU patients, **real-world validation** is essential for assessing its clinical applicability. To transition from simulation to practical use, future studies must incorporate actual patient data from ICU databases, such as MIMIC-III or eICU-CRD, to validate the model's efficacy under real-world conditions [52]-[54]. Real-world testing would provide insight into how well the model generalizes to diverse patient populations, including those with multiple comorbidities, varying demographics, and differing treatment protocols.

In addition to evaluating predictive accuracy, real-world applications should focus on **interpretability and usability** for clinical teams [55]. The attention mechanism embedded in the model provides a foundation for interpretable predictions by highlighting critical time periods, but further interpretability tools could be integrated to ensure that clinicians understand how and why the model arrives at its predictions. Clinicians' trust in AI-driven tools hinges on transparent and interpretable outputs, as ICU settings are high-stakes environments where predictive errors could have severe consequences [56]-[59]. Future iterations of the model could incorporate additional interpretability techniques, such as feature attribution scores or visual dashboards, that make model predictions more actionable for ICU staff [60].

Ethical and Practical Considerations must also be addressed. Implementing such a model in ICU settings would require regulatory approval, robust data privacy measures, and integration with existing hospital information systems [61] [62]. Additionally, continuous monitoring and periodic retraining on updated

data would be essential to maintain the model's performance, as ICU data can change over time due to new treatment protocols, equipment, and patient demographics [63]. Collaboration with healthcare providers to establish standardized protocols for model deployment, monitoring, and retraining could facilitate the transition from research to real-world use.

4.4. Future Directions

To advance this research, several avenues should be explored. First, **collecting diverse ICU data** from multiple hospitals and geographic regions would allow the model to learn from a broader patient population, enhancing its generalizability and robustness. Such diversity would also expose the model to a wider range of medical conditions and treatment variations, thereby improving its adaptability to different ICU settings. Second, **expanding the feature set** to include additional vital signs, laboratory results, and treatment interventions could improve the model's ability to predict delirium by capturing more comprehensive patient information [64]. Real-world ICU data often include a wealth of clinically relevant data points, such as oxygen saturation, blood pH levels, and medication dosages, which could provide additional predictive power and help the model account for a wider array of risk factors. Third, **longitudinal studies** assessing the model's performance over extended periods would provide insight into its long-term reliability and its ability to adapt to shifts in ICU protocols or patient demographics [65]. Longitudinal validation would also enable researchers to track the impact of model-guided interventions on patient outcomes, such as reductions in delirium incidence or improved recovery times [66]. Lastly, the model could be enhanced by incorporating **ensemble techniques** or hybrid models combining LSTM with other architectures, such as convolutional neural networks (CNNs), to process multimodal data [67]. Hybrid models could, for example, process both time-series data and medical imaging (e.g., brain scans) for a more comprehensive assessment of delirium risk, potentially capturing risk factors that are undetectable in time-series data alone [68] [69].

5. Conclusions

This study introduces an innovative approach to early delirium prediction in ICU patients by leveraging a deep learning model that integrates a simplified LSTM architecture with an attention mechanism. By combining static patient characteristics with dynamic, time-series physiological data, the model achieved near-perfect predictive performance on a synthetic dataset, distinguishing between delirium and non-delirium cases with remarkable accuracy. This outcome underscores the potential of deep learning to improve risk stratification in critical care settings, providing clinicians with a reliable tool for early identification of high-risk patients. The model's attention mechanism was particularly instrumental, allowing it to prioritize clinically relevant time points in the dynamic data. This feature not only enhances the model's interpretability but also aligns with the clinical

observation that delirium often develops as a gradual physiological change. By focusing on critical temporal patterns, the model has demonstrated the capability to detect early signs of delirium, potentially enabling timely interventions that could mitigate the adverse outcomes associated with this condition. However, despite the promising results, it is important to acknowledge the limitations associated with synthetic data [70]. While synthetic datasets allow for initial testing and model refinement, they cannot fully capture the variability, noise, and complexity inherent in real-world ICU data [71]. The true test of this model's effectiveness lies in its application to clinical datasets, where diverse patient populations, comorbid conditions, and treatment interventions introduce additional layers of complexity. For the model to be clinically viable, it must demonstrate consistent performance across varied and often imperfect real-world data sources [72]. Future work will focus on addressing these challenges by testing the model on real ICU data from sources like MIMIC-III or eICU-CRD databases. These datasets will provide a more rigorous environment to evaluate the model's robustness, adaptability, and generalizability. Additionally, we aim to explore methods for handling noisy, incomplete, or sparse data, which are common in healthcare settings. Techniques such as data augmentation, imputation, and transfer learning could enhance the model's resilience to these challenges, enabling it to maintain predictive accuracy in the face of real-world data limitations.

In conclusion, this study presents a significant step toward using machine learning for proactive delirium management in critical care environments. With further validation and refinement, this model has the potential to become an asset in ICU settings, aiding clinicians in making timely, data-driven decisions to improve patient outcomes. Through continued research and collaboration with healthcare providers, we aim to bring this model closer to practical implementation, where it could contribute to enhancing patient safety and quality of care in intensive care units.

This template, created in MS Word 2007, provides authors with most of the formatting specifications needed for preparing electronic versions of their papers. All standard paper components have been specified for three reasons: 1) ease of use when formatting individual papers, 2) automatic compliance to electronic requirements that facilitate the concurrent or later production of electronic products, and 3) conformity of style throughout a journal paper. Margins, column widths, line spacing, and type styles are built-in; examples of the type styles are provided throughout this document and are identified in italic type, within parentheses, following the example. Some components, such as multi-leveled equations, graphics, and tables are not prescribed, although the various table text styles are provided. The formatter will need to create these components, incorporating the applicable criteria that follow.

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] Wilson, J.E., Mart, M.F., Cunningham, C., Shehabi, Y., Girard, T.D., MacLulich, A.M.J., *et al.* (2020) Delirium. *Nature Reviews Disease Primers*, **6**, Article No. 90. <https://doi.org/10.1038/s41572-020-00223-4>
- [2] Mehta, S., Cook, D., Devlin, J.W., Skrobik, Y., Meade, M., Fergusson, D., *et al.* (2015) Prevalence, Risk Factors, and Outcomes of Delirium in Mechanically Ventilated Adults. *Critical Care Medicine*, **43**, 557-566. <https://doi.org/10.1097/ccm.0000000000000727>
- [3] Kumar, V., Angurana, S.K., Baranwal, A.K. and Nallasamy, K. (2021) Nasotracheal Vs. Orotracheal Intubation and Post-Extubation Airway Obstruction in Critically Ill Children: An Open-Label Randomized Controlled Trial. *Frontiers in Pediatrics*, **9**, Article 713516. <https://doi.org/10.3389/fped.2021.713516>
- [4] Nielsen, A.H., Larsen, L.K., Collet, M.O., Lehmkuhl, L., Bekker, C., Jensen, J.F., *et al.* (2023) Intensive Care Unit Nurses' Perception of Three Different Methods for Delirium Screening: A Survey (Delis-3). *Australian Critical Care*, **36**, 1035-1042. <https://doi.org/10.1016/j.aucc.2022.12.008>
- [5] Saad, R., Husnain, A., Saeed, A., Gill, A.Y. and Hussain, H.K. (2023) Harnessing Predictive Power: Exploring the Crucial Role of Machine Learning in Early Disease Detection. *Jurnal Inovasi dan Humaniora*, **1**, 302-315.
- [6] Rush, B., Celi, L.A. and Stone, D.J. (2018) Applying Machine Learning to Continuously Monitored Physiological Data. *Journal of Clinical Monitoring and Computing*, **33**, 887-893. <https://doi.org/10.1007/s10877-018-0219-z>
- [7] Cheng, F., Joshi, H., Tandon, P., Freeman, R., Reich, D.L., Mazumdar, M., *et al.* (2020) Using Machine Learning to Predict ICU Transfer in Hospitalized COVID-19 Patients. *Journal of Clinical Medicine*, **9**, Article 1668. <https://doi.org/10.3390/jcm9061668>
- [8] Xu, Z., Guo, J., Qin, L., Xie, Y., Xiao, Y., Lin, X., *et al.* (2024) Predicting ICU Interventions: A Transparent Decision Support Model Based on Multivariate Time Series Graph Convolutional Neural Network. *IEEE Journal of Biomedical and Health Informatics*, **28**, 3709-3720. <https://doi.org/10.1109/jbhi.2024.3379998>
- [9] Ahmedt-Aristizabal, D., Armin, M.A., Denman, S., Fookes, C. and Petersson, L. (2021) Graph-Based Deep Learning for Medical Diagnosis and Analysis: Past, Present and Future. *Sensors*, **21**, Article 4758. <https://doi.org/10.3390/s21144758>
- [10] Yan, Q., Lu, Z., Liu, H., He, X., Zhang, X. and Guo, J. (2024) Short-Term Prediction of Integrated Energy Load Aggregation Using a Bi-Directional Simple Recurrent Unit Network with Feature-Temporal Attention Mechanism Ensemble Learning Model. *Applied Energy*, **355**, Article 122159. <https://doi.org/10.1016/j.apenergy.2023.122159>
- [11] Albahri, A.S., Duhaim, A.M., Fadhel, M.A., Alnoor, A., Baqer, N.S., Alzubaidi, L., *et al.* (2023) A Systematic Review of Trustworthy and Explainable Artificial Intelligence in Healthcare: Assessment of Quality, Bias Risk, and Data Fusion. *Information Fusion*, **96**, 156-191. <https://doi.org/10.1016/j.inffus.2023.03.008>
- [12] Rozzini, R., Sabatini, T., Cassinadri, A., Boffelli, S., Ferri, M., Barbisoni, P., *et al.* (2005) Relationship between Functional Loss before Hospital Admission and Mortality in Elderly Persons with Medical Illness. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, **60**, 1180-1183. <https://doi.org/10.1093/gerona/60.9.1180>
- [13] Zhang, Y.L. (2021) An Interactive Machine Learning Approach to Integrating Physician Expertise into Delirium Prediction Model Development. Master's Thesis, University of Toronto (Canada).

- [14] Dowell, E.H. and Hall, K.C. (2001) Modeling of Fluid-Structure Interaction. *Annual Review of Fluid Mechanics*, **33**, 445-490. <https://doi.org/10.1146/annurev.fluid.33.1.445>
- [15] Li, Y., Zhu, Z., Kong, D., Han, H. and Zhao, Y. (2019) EA-LSTM: Evolutionary Attention-Based LSTM for Time Series Prediction. *Knowledge-Based Systems*, **181**, Article 104785. <https://doi.org/10.1016/j.knosys.2019.05.028>
- [16] Sheikhalishahi, S., Bhattacharyya, A., Celi, L.A. and Osmani, V. (2023) An Interpretable Deep Learning Model for Time-Series Electronic Health Records: Case Study of Delirium Prediction in Critical Care. *Artificial Intelligence in Medicine*, **144**, Article 102659. <https://doi.org/10.1016/j.artmed.2023.102659>
- [17] Yeung, M., Sala, E., Schönlieb, C. and Rundo, L. (2022) Unified Focal Loss: Generalising Dice and Cross Entropy-Based Losses to Handle Class Imbalanced Medical Image Segmentation. *Computerized Medical Imaging and Graphics*, **95**, Article 102026. <https://doi.org/10.1016/j.compmedimag.2021.102026>
- [18] Mushava, J. and Murray, M. (2024) Flexible Loss Functions for Binary Classification in Gradient-Boosted Decision Trees: An Application to Credit Scoring. *Expert Systems with Applications*, **238**, Article 121876. <https://doi.org/10.1016/j.eswa.2023.121876>
- [19] Varshney, R.P. and Sharma, D.K. (2024) Optimizing Time-Series Forecasting Using Stacked Deep Learning Framework with Enhanced Adaptive Moment Estimation and Error Correction. *Expert Systems with Applications*, **249**, Article 123487. <https://doi.org/10.1016/j.eswa.2024.123487>
- [20] Zhao, C., Badrinarayanan, V., Lee, C.-Y. and Rabinovich, A. (2018) GradNorm: Gradient Normalization for Adaptive Loss Balancing in Deep Multitask Networks. *International Conference on Machine Learning*, Sweden, 10-15 July 2018, 794-803.
- [21] Aryafar, E., Khojastepour, M.A., Sundaresan, K., Rangarajan, S. and Knightly, E. (2013) ADAM: An Adaptive Beamforming System for Multicasting in Wireless Lans. *IEEE/ACM Transactions on Networking*, **21**, 1595-1608. <https://doi.org/10.1109/tnet.2012.2228501>
- [22] Thabtah, F., Hammoud, S., Kamalov, F. and Gonsalves, A. (2020) Data Imbalance in Classification: Experimental Evaluation. *Information Sciences*, **513**, 429-441. <https://doi.org/10.1016/j.ins.2019.11.004>
- [23] Bradley, A.P. (1997) The Use of the Area under the ROC Curve in the Evaluation of Machine Learning Algorithms. *Pattern Recognition*, **30**, 1145-1159. [https://doi.org/10.1016/s0031-3203\(96\)00142-2](https://doi.org/10.1016/s0031-3203(96)00142-2)
- [24] Lv, S.R., Li, J.Q., He, H., Zhao, Q. and Jiang, Y.N. (2024) Artificial Intelligence Applications in Delirium Prediction, Diagnosis, and Management: A Systematic Review.
- [25] Gulshan, S. (2023) Evaluation of Risk Factors for Falls in the Elderly Based on a Real Data Set Using Bayesian Networks. PhD Dissertation, Université Polytechnique Hauts-de-France.
- [26] Chicco, D. and Jurman, G. (2020) The Advantages of the Matthews Correlation Coefficient (MCC) over F1 Score and Accuracy in Binary Classification Evaluation. *BMC Genomics*, **21**, 1-13. <https://doi.org/10.1186/s12864-019-6413-7>
- [27] Al-rimy, B.A.S., Saeed, F., Al-Sarem, M., Albarrak, A.M. and Qasem, S.N. (2023) An Adaptive Early Stopping Technique for Densenet 169-Based Knee Osteoarthritis Detection Model. *Diagnostics*, **13**, Article 1903. <https://doi.org/10.3390/diagnostics13111903>
- [28] Miseta, T., Fodor, A. and Vathy-Fogarassy, Á. (2024) Surpassing Early Stopping: A

- Novel Correlation-Based Stopping Criterion for Neural Networks. *Neurocomputing*, **567**, Article 127028. <https://doi.org/10.1016/j.neucom.2023.127028>
- [29] Kumar, A., Parkash, C., Tang, H. and Xiang, J. (2023) Intelligent Framework for Degradation Monitoring, Defect Identification and Estimation of Remaining Useful Life (RUL) of Bearing. *Advanced Engineering Informatics*, **58**, Article 102206. <https://doi.org/10.1016/j.aei.2023.102206>
- [30] Nusrat, I. and Jang, S. (2018) A Comparison of Regularization Techniques in Deep Neural Networks. *Symmetry*, **10**, Article 648. <https://doi.org/10.3390/sym10110648>
- [31] Sattler, F., Wiedemann, S., Muller, K. and Samek, W. (2020) Robust and Communication-Efficient Federated Learning from Non-I.I.D. Data. *IEEE Transactions on Neural Networks and Learning Systems*, **31**, 3400-3413. <https://doi.org/10.1109/tnnls.2019.2944481>
- [32] Al-Zyoud, I., Laamarti, F., Ma, X., Tobón, D. and El Saddik, A. (2022) Towards a Machine Learning-Based Digital Twin for Non-Invasive Human Bio-Signal Fusion. *Sensors*, **22**, Article 9747. <https://doi.org/10.3390/s22249747>
- [33] Hassan, E., Shams, M.Y., Hikal, N.A. and Elmougy, S. (2022) The Effect of Choosing Optimizer Algorithms to Improve Computer Vision Tasks: A Comparative Study. *Multimedia Tools and Applications*, **82**, 16591-16633. <https://doi.org/10.1007/s11042-022-13820-0>
- [34] Benyamin, G. and Crowley, M. (2019) The Theory behind Overfitting, cross Validation, Regularization, Bagging, and Boosting: Tutorial.
- [35] Leslie, R., Wong, E. and Kolter, Z. (2020) Overfitting in Adversarially Robust Deep Learning. *International Conference on Machine Learning*, Austria, 13-18 July 2020, 8093-8104.
- [36] Cieslak, D.A., Chawla, N.V. and Striegel, A. (2006) Combating Imbalance in Network Intrusion Datasets. 2006 *IEEE International Conference on Granular Computing*, Atlanta, 10-12 May 2006, 732-737. <https://doi.org/10.1109/grc.2006.1635905>
- [37] Smith, L.N. (2018) A Disciplined Approach to Neural Network Hyper-Parameters: Part 1: Learning Rate, Batch Size, Momentum, and Weight Decay.
- [38] Ellis, A.W. and Lambon Ralph, M.A. (2000) Age of Acquisition Effects in Adult Lexical Processing Reflect Loss of Plasticity in Maturing Systems: Insights from Connectionist Networks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **26**, 1103-1123. <https://doi.org/10.1037//0278-7393.26.5.1103>
- [39] Sepehr, A., Gomis-Bellmunt, O. and Pouresmaeil, E. (2022) Employing Machine Learning for Enhancing Transient Stability of Power Synchronization Control during Fault Conditions in Weak Grids. *IEEE Transactions on Smart Grid*, **13**, 2121-2131. <https://doi.org/10.1109/tsg.2022.3148590>
- [40] Hofer, I.S., Lee, C., Gabel, E., Baldi, P. and Cannesson, M. (2020) Development and Validation of a Deep Neural Network Model to Predict Postoperative Mortality, Acute Kidney Injury, and Reintubation Using a Single Feature Set. *npj Digital Medicine*, **3**, Article No. 58. <https://doi.org/10.1038/s41746-020-0248-0>
- [41] Wan, A., Chang, Q., AL-Bukhaiti, K. and He, J. (2023) Short-Term Power Load Forecasting for Combined Heat and Power Using CNN-LSTM Enhanced by Attention Mechanism. *Energy*, **282**, Article 128274. <https://doi.org/10.1016/j.energy.2023.128274>
- [42] Gigerenzer, G. and Todd, P.M. (1999) Fast and Frugal Heuristics: The Adaptive Toolbox. In: *Simple Heuristics That Make Us Smart*, Oxford University Press, 3-34.
- [43] O'Sullivan, R., Inouye, S.K. and Meagher, D. (2014) Delirium and Depression: Inter-

- Relationship and Clinical Overlap in Elderly People. *The Lancet Psychiatry*, **1**, 303-311. [https://doi.org/10.1016/s2215-0366\(14\)70281-0](https://doi.org/10.1016/s2215-0366(14)70281-0)
- [44] Oldham, M.A. (2022) Delirium Disorder: Unity in Diversity. *General Hospital Psychiatry*, **74**, 32-38. <https://doi.org/10.1016/j.genhosppsych.2021.11.007>
- [45] Boehm, L.M., Jones, A.C., Selim, A.A., Virdun, C., Garrard, C.F., Walden, R.L., *et al.* (2021) Delirium-Related Distress in the ICU: A Qualitative Meta-Synthesis of Patient and Family Perspectives and Experiences. *International Journal of Nursing Studies*, **122**, Article 104030. <https://doi.org/10.1016/j.ijnurstu.2021.104030>
- [46] Niso, G., Krol, L.R., Combrisson, E., Dubarry, A.S., Elliott, M.A., François, C., *et al.* (2022) Good Scientific Practice in EEG and MEG Research: Progress and Perspectives. *NeuroImage*, **257**, Article 119056. <https://doi.org/10.1016/j.neuroimage.2022.119056>
- [47] Lucchiari, C. and Pravettoni, G. (2011) Cognitive Balanced Model: A Conceptual Scheme of Diagnostic Decision Making. *Journal of Evaluation in Clinical Practice*, **18**, 82-88. <https://doi.org/10.1111/j.1365-2753.2011.01771.x>
- [48] Stanley, D.J. and Spence, J.R. (2014) Expectations for Replications. *Perspectives on Psychological Science*, **9**, 305-318. <https://doi.org/10.1177/1745691614528518>
- [49] Crisan, S. (2016) A Novel Perspective on Hand Vein Patterns for Biometric Recognition: Problems, Challenges, and Implementations. In: *Signal Processing for Security Technologies*, Springer, 21-49. https://doi.org/10.1007/978-3-319-47301-7_2
- [50] Antonio, G., Kapoor, A., and Pal, S. (2019) Deep Learning with TensorFlow 2 and Keras: Regression, ConvNets, GANs, RNNs, NLP, and More with TensorFlow 2 and the Keras API. Packt Publishing Ltd.
- [51] Boominathan, V., Robinson, J.T., Waller, L. and Veeraraghavan, A. (2021) Recent Advances in Lensless Imaging. *Optica*, **9**, 1-16. <https://doi.org/10.1364/optica.431361>
- [52] Kim, Y.K., Seo, W., Lee, S.J., Koo, J.H., Kim, G.C., Song, H.S., *et al.* (2024) Early Prediction of Cardiac Arrest in the Intensive Care Unit Using Explainable Machine Learning: Retrospective Study. *Journal of Medical Internet Research*, **26**, e62890. <https://doi.org/10.2196/62890>
- [53] Huang, Z., Huang, S., Li, C., *et al.* (2020) Artificial Intelligence Assisted Early Warning System for Acute Kidney Injury Driven by Multi-Center ICU Database.
- [54] Liu, P., Li, S., Zheng, T., Wu, J., Fan, Y., Liu, X., *et al.* (2023) Subphenotyping Heterogeneous Patients with Chronic Critical Illness to Guide Individualised Fluid Balance Treatment Using Machine Learning: A Retrospective Cohort Study. *E Clinical Medicine*, **59**, Article 101970. <https://doi.org/10.1016/j.eclinm.2023.101970>
- [55] Cutillo, C.M., Sharma, K.R., Foschini, L., Kundu, S., Mackintosh, M., Mandl, K.D., *et al.* (2020) Machine Intelligence in Healthcare—Perspectives on Trustworthiness, Explainability, Usability, and Transparency. *npj Digital Medicine*, **3**, Article No. 47. <https://doi.org/10.1038/s41746-020-0254-2>
- [56] Rahman, H., Kazi, R.H., *et al.* (2024) Improving Collaborative Interactions between Humans and Artificial Intelligence to Achieve Optimal Patient Outcomes in the Healthcare Industry.
- [57] Kalusivalingam, A.K., Sharma, A., Patel, N. and Singh, V. (2012) Enhancing ICU Monitoring with Predictive Analytics Using Random Forests and Long Short-Term Memory Networks. *International Journal of AI and ML*, **1**, 1-26.
- [58] Filippis, R.d. and Foysal, A.A. (2024) Securing Predictive Psychological Assessments: The Synergy of Blockchain Technology and Artificial Intelligence. *Open Access Library*, **11**, 1-23. <https://doi.org/10.4236/oalib.1112378>

- [59] Filippis, R.d. and Foysal, A.A. (2024) Harnessing the Power of Artificial Intelligence in Neuromuscular Disease Rehabilitation: A Comprehensive Review and Algorithmic Approach. *Advances in Bioscience and Biotechnology*, **15**, 289-309. <https://doi.org/10.4236/abb.2024.155018>
- [60] El-Rashidy, N., Tarek, Z., Elshewey, A.M. and Shams, M.Y. (2024) Multitask Multi-layer-Prediction Model for Predicting Mechanical Ventilation and the Associated Mortality Rate. *Neural Computing and Applications*, **2024**, 1-23. <https://doi.org/10.1007/s00521-024-10468-9>
- [61] Filippis, R.D. and Foysal, A.A. (2024) Chatbots in Psychology: Revolutionizing Clinical Support and Mental Health Care. *Voice of the Publisher*, **10**, 298-321. <https://doi.org/10.4236/vp.2024.103025>
- [62] de Filippis, R., Al Foysal, A., Rocco, V., *et al.* (2024) The Risk Perspective of AI in Healthcare: GDPR and GELSI Framework (Governance, Ethical, Legal and Social Implications) and the New European AI Act. *Italian Journal of Psychiatry*, **10**, 12-16.
- [63] Mehdizavareh, H., Khan, A. and Cichosz, S.L. (2024) Enhancing Glucose Level Prediction of ICU Patients through Irregular Time-Series Analysis and Integrated Representation.
- [64] Bhattacharyya, A., Sheikhalishahi, S., Torbic, H., Yeung, W., Wang, T., Birst, J., *et al.* (2022) Delirium Prediction in the ICU: Designing a Screening Tool for Preventive Interventions. *JAMIA Open*, **5**, oaac048. <https://doi.org/10.1093/jamiaopen/oaac048>
- [65] Long, D.A. and Fink, E.L. (2021) Transitions from Short to Long-Term Outcomes in Pediatric Critical Care: Considerations for Clinical Practice. *Translational Pediatrics*, **10**, 2858-2874. <https://doi.org/10.21037/tp-21-61>
- [66] Black, P., Boore, J.R.P. and Parahoo, K. (2011) The Effect of Nurse-Facilitated Family Participation in the Psychological Care of the Critically Ill Patient. *Journal of Advanced Nursing*, **67**, 1091-1101. <https://doi.org/10.1111/j.1365-2648.2010.05558.x>
- [67] Boehnlein, A., Diefenthaler, M., Sato, N., Schram, M., Ziegler, V., Fanelli, C., *et al.* (2022) Colloquium: Machine Learning in Nuclear Physics. *Reviews of Modern Physics*, **94**, Article 031003. <https://doi.org/10.1103/revmodphys.94.031003>
- [68] Tiegens, Z., Maclullich, A.M.J., Anand, A., Brookes, C., Cassarino, M., O'connor, M., *et al.* (2020) Diagnostic Accuracy of the 4AT for Delirium Detection in Older Adults: Systematic Review and Meta-Analysis. *Age and Ageing*, **50**, 733-743. <https://doi.org/10.1093/ageing/afaa224>
- [69] Haverkate, M.R., Xie, M., Mamun, A.A., *et al.* (2022) Impact of the SARS-CoV-2 Pandemic on Overall and Diagnosis-Specific Antibiotic Prescription Rates in Long-Term-Care Facilities in British Columbia, Canada. <https://utppublishing.com/doi/pdf/10.3138/jammi.7.s1.abst>
- [70] Goyal, M. and Mahmoud, Q.H. (2024) A Systematic Review of Synthetic Data Generation Techniques Using Generative AI. *Electronics*, **13**, Article 3509. <https://doi.org/10.3390/electronics13173509>
- [71] Murtaza, H., Ahmed, M., Khan, N.F., Murtaza, G., Zafar, S. and Bano, A. (2023) Synthetic Data Generation: State of the Art in Health Care Domain. *Computer Science Review*, **48**, Article 100546. <https://doi.org/10.1016/j.cosrev.2023.100546>
- [72] Major, V.J., Jethani, N. and Aphinyanaphongs, Y. (2020) Estimating Real-World Performance of a Predictive Model: A Case-Study in Predicting Mortality. *JAMIA Open*, **3**, 243-251. <https://doi.org/10.1093/jamiaopen/oaac008>