A Survey of Scoring Systems for Mortality Prediction and Risk Assessment in Critically Ill Patients

www.surveyx.cn

Abstract

This survey paper examines the pivotal role of scoring systems such as APSIII, SOFA, LODS, and OASIS in enhancing clinical outcomes and decision-making in Intensive Care Units (ICUs). These systems are integral for assessing patient risk and predicting mortality, facilitating timely interventions, and optimizing resource allocation. The integration of advanced machine learning models has significantly improved the predictive accuracy and interpretability of these scoring systems, as demonstrated by innovative models like the attention-augmented convolutional model, which provides interpretable insights crucial for clinical trust. The application of methods such as the dTIC model highlights the importance of personalized approaches in critical care by identifying distinct patient phenotypes with unique clinical outcomes. The incorporation of multimodal data sources has further enhanced the predictive accuracy of models for specific conditions, surpassing traditional methods. The potential of artificial intelligence to improve clinical decision-making and patient outcomes is evident, with ongoing research aimed at optimizing models with larger datasets and applying them to diverse clinical tasks. Future research should focus on validating frameworks across a broader range of predictive tools and enhancing their usability for end users. The continuous evolution of scoring systems, powered by advancements in AI and data integration, holds significant promise for improving patient care in ICUs, addressing current challenges, and exploring new research avenues to further enhance their impact on clinical outcomes and decision-making processes in critical care environments.

1 Introduction

1.1 Importance of Mortality Prediction and Risk Assessment

Mortality prediction and risk assessment are critical for managing critically ill patients in Intensive Care Units (ICUs), where timely identification of patients at risk for clinical deterioration can significantly improve outcomes. Utilizing multimodal data from ICU Electronic Health Records (EHRs) enhances predictive performance, facilitating early interventions and optimizing resource allocation. The COVID-19 pandemic underscored the importance of effective mortality prediction models, which are essential for efficient patient triage based on disease severity [1].

Integrating precise estimates of adverse outcomes, such as those associated with COVID-19, into clinical workflows can lead to better resource allocation and targeted preventive measures [2]. Inadequate predictions in ICU settings can lead to poor clinical outcomes, highlighting the need for advanced models that can process diverse data types [3]. Mortality prediction is also vital for assessing risks of ICU readmission and various mortality outcomes, thus enhancing overall patient care [4].

Specific medical conditions, such as traumatic brain injury and acute kidney injury, further illustrate the necessity of effective outcome prediction and targeted management strategies due to their significant impact on long-term health. Early predictive strategies for conditions like stroke are essential for

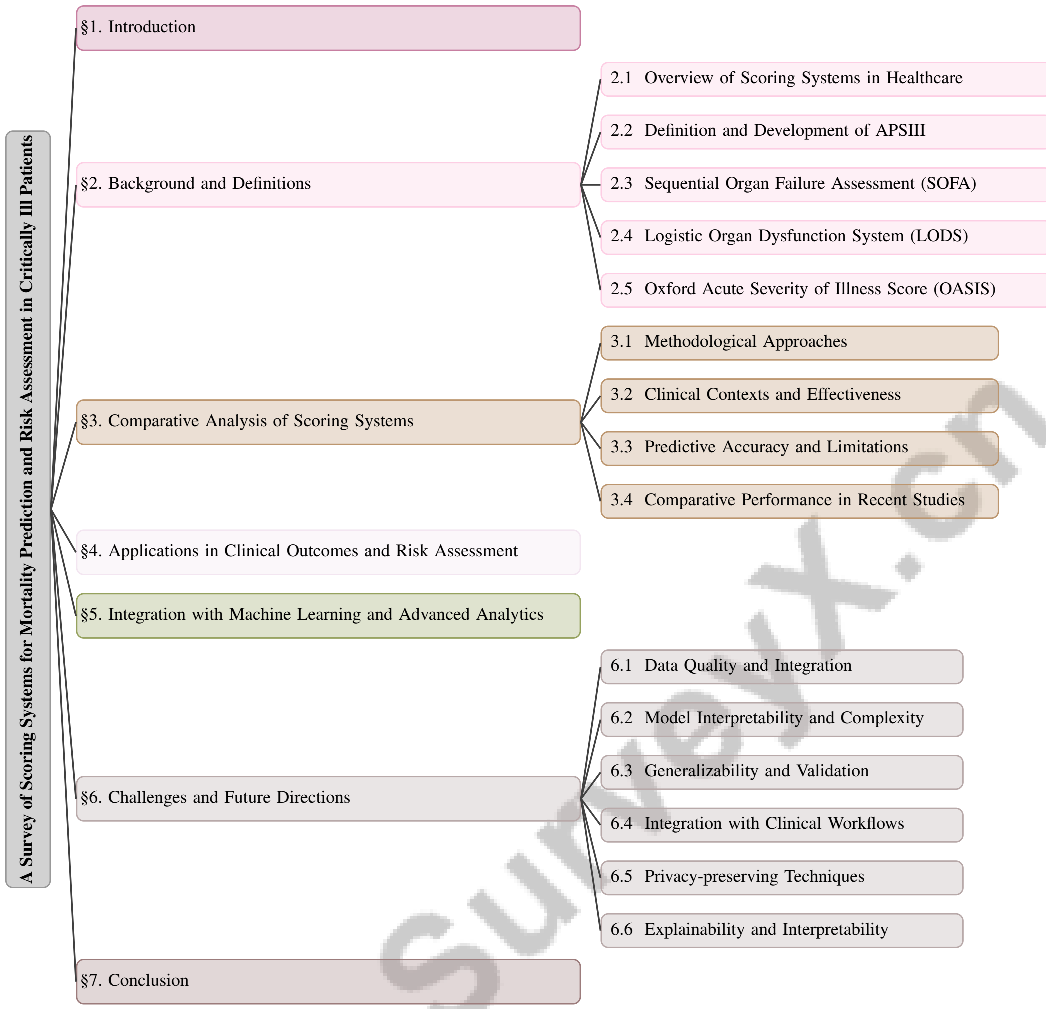


Figure 1: chapter structure

mitigating risks and improving patient prognosis [5]. Consequently, integrating mortality prediction into clinical practice not only enhances patient monitoring and care prioritization but also reduces complications, ultimately improving the quality of care in critical settings [6].

1.2 Role of Scoring Systems in Clinical Decision-Making

Scoring systems are integral to clinical decision-making in ICUs, providing a structured method for assessing patient risk, which is crucial for optimizing resource allocation and enhancing patient management [7]. Systems such as APSIII, SOFA, LODS, and OASIS convert complex clinical data into actionable insights, enabling timely interventions and improved patient outcomes [8]. The incorporation of machine learning algorithms into these scoring systems enhances predictions regarding ICU length of stay and mortality, utilizing vital signs and demographic features to support clinical decisions [7].

Traditional acuity scores often depend on manual assessments, which can be burdensome for healthcare providers [9]. Innovations in machine learning, like the eXplainable Multimodal Mortality Predictor (XMMP), address these challenges by integrating various data modalities to enhance prediction accuracy and interpretability [10]. This approach not only improves mortality risk prediction but also clarifies factors influencing patient outcomes.

Despite their advantages, existing scoring systems sometimes lack the individual prediction accuracy necessary for optimal clinical decision-making [11]. Novel methodologies, such as combining Long Short-Term Memory (LSTM) networks with latent topic models, have been proposed to interpret clinical notes and improve decision-making capabilities [12]. Additionally, real-time mortality risk assessments, exemplified by systems like BoXHED, adapt to patient health changes, providing dynamic support for clinical decisions [13].

The potential of large language models (LLMs) to structure unstructured EHR data further enhances survival predictions [14]. However, existing methods like tf-idf embeddings and latent topic distributions have proven inadequate in capturing clinical language nuances, indicating a need for improved solutions [15]. Continuous refinement of scoring systems is crucial for clinician acceptance, and as machine learning approaches enhance predictive performance, efforts must focus on improving transparency and trust in healthcare settings [10]. By integrating both structured and unstructured data and enhancing interpretability, scoring systems significantly advance clinical decision-making and optimize resource allocation in critical care environments [6].

1.3 Structure of the Survey

This survey is meticulously structured to provide a comprehensive analysis of scoring systems for mortality prediction and risk assessment in critically ill patients. It begins with an introduction that emphasizes the significance of mortality prediction and the essential role of scoring systems such as APSIII, SOFA, LODS, and OASIS in clinical decision-making and resource management. The background section elucidates key concepts and definitions, offering an overview of these scoring systems and their development.

The core of the survey presents a comparative analysis that examines the methodologies, clinical contexts, predictive accuracies, and limitations of each scoring system. Subsequent sections explore the practical applications of these systems in predicting clinical outcomes and assessing risks across various disease contexts, highlighting the integration of dynamic modeling and real-time assessments.

The survey further investigates the emerging field of machine learning and advanced analytics, detailing how these technologies enhance traditional scoring systems. This includes discussions on predictive accuracy, interpretability, and the use of multimodal data. Challenges and future directions are addressed in the penultimate section, focusing on data quality, model complexity, generalizability, and integration within clinical workflows. The survey concludes by summarizing key insights and proposing avenues for future research and development, underscoring the importance of these systems in improving outcomes in intensive care.The following sections are organized as shown in Figure 1.

2 Background and Definitions

2.1 Overview of Scoring Systems in Healthcare

In healthcare, particularly within Intensive Care Units (ICUs), scoring systems are pivotal for assessing patient risk and predicting outcomes. Traditional models like the Acute Physiology and Chronic Health Evaluation (APACHE), Simplified Acute Physiology Score (SAPS), and Mortality Prediction Models (MPM) have been fundamental due to their structured methods for evaluating illness severity. However, these models often rely on static data, which limits their effectiveness in the dynamic ICU setting, and their dependence on statistical methods can introduce bias, especially with incomplete datasets [16].

To address these limitations, scoring systems have evolved to incorporate sophisticated, data-driven approaches that integrate multiple data sources, enhancing predictive accuracy and patient outcomes. Recent advancements highlight the constraints of traditional methods that primarily use labeled data for training deep learning models. There is an increasing need for methodologies that combine clinical measurements with narrative notes to improve predictive performance, given the non-stationary and heterogeneous nature of ICU data [4].

Innovative strategies, such as domain adaptation, have emerged to enhance mortality prediction across various ICU settings. Integrating multimodal data, including structured electronic health records (EHRs) and imaging data, has been proposed to improve prediction accuracy. Advanced machine learning techniques using initial imaging data have shown promise in predicting 30-day mortality, underscoring the potential of diverse data sources [14]. Automated systems like EventScore, which utilize logistic regression with lasso regularization, demonstrate the potential of real-time assessment tools in ICUs [9].

Benchmarks have been established to evaluate machine learning models in critical care tasks, including mortality prediction and length of stay estimation [1]. These benchmarks provide insights into the effectiveness of different approaches in real-world settings. However, existing clinical criteria and prediction algorithms often require extensive patient data and manual labeling, which can be resource-intensive [16].

The integration of advanced predictive models enhances mortality prediction precision and supports more personalized patient care strategies. By incorporating dynamic patient data, these models address the limitations of current Early Warning Systems (EWS) in predicting patient deterioration [17]. This shift towards dynamic and comprehensive scoring systems is crucial for improving clinical outcomes and optimizing resource allocation in critical care environments. Datasets from EHRs, such as those involving approximately 80,000 patients at a cancer center, provide a robust foundation for developing and evaluating these advanced models [18].

2.2 Definition and Development of APSIII

The Acute Physiology Score III (APSIII) is a refined scoring system designed to assess illness severity among critically ill patients in ICUs. Building upon predecessors such as APACHE II, APSIII incorporates an expanded set of physiological variables and patient-specific data, enhancing its predictive accuracy for in-hospital mortality [19]. This comprehensive approach facilitates timely clinical interventions and optimizes resource allocation within ICU settings [20].

APSIII was developed to address the limitations of earlier scoring systems that relied on static datasets and a narrow range of physiological parameters. By integrating dynamic patient data, APSIII enables more precise and real-time mortality risk predictions, essential in the rapidly changing ICU environment [21]. This dynamic integration is critical for accurately assessing patient conditions, as demonstrated in studies on mortality prediction for complex cases like hemorrhagic stroke and acute kidney injury (AKI).

Methodologically, APSIII employs logistic regression to analyze clinical data and identify significant mortality predictors, which are then used to formulate a scoring system reflecting illness severity [22]. Advanced techniques, such as the MixEHR-SurG method, enhance mortality predictions by combining patient-specific topic distributions with the Cox proportional hazards model [23]. Real-time mortality risk assessments facilitated by systems like BoXHED, which incorporates time-dependent covariates, further augment APSIII's predictive capabilities [13].

In practice, APSIII has shown superior performance in predicting in-hospital mortality, especially in complex clinical scenarios involving hematological malignancies and respiratory failure. Comparative benchmarks consistently highlight APSIII's enhanced predictive capabilities, underscoring its critical role in evaluating and improving predictive models in clinical contexts [24]. The integration of unstructured data from EHRs, as explored in post-radiotherapy mortality predictions, further exemplifies APSIII's potential to improve mortality risk assessments [14].

2.3 Sequential Organ Failure Assessment (SOFA)

The Sequential Organ Failure Assessment (SOFA) score is a vital tool in ICUs for evaluating patient severity by assessing the function or dysfunction of six major organ systems: respiratory, cardiovascular, hepatic, coagulation, renal, and neurological [25]. Initially developed to quantify organ dysfunction in sepsis, the SOFA score has been extensively adopted due to its straightforward application and robust predictive capabilities in assessing patient morbidity and treatment efficacy [26].

Despite its widespread use, traditional SOFA scoring faces challenges, including variability in measurement across different assessors and clinical settings, leading to inconsistent outcomes in clinical trials [26]. Additionally, the traditional SOFA score has been critiqued for inadequately predicting mortality in ICU patients, particularly when gastrointestinal dysfunction is not accounted for [25]. This limitation underscores the need for integrating additional data sources and sophisticated models to improve predictive accuracy [3].

Recent advancements have sought to address these limitations by integrating machine learning techniques, such as Long Short-Term Memory (LSTM) models, which have demonstrated superior performance in mortality prediction compared to traditional SOFA methods [17]. Models like XGBoost, achieving a mortality prediction accuracy of 88

The application of the SOFA score in intensive care is crucial for guiding clinical decisions and interventions, providing a comprehensive view of a patient's physiological status [27]. Ongoing refinement of the SOFA score through technological innovations and research continues to enhance its predictive accuracy and clinical utility, ensuring its relevance in supporting effective patient management and improving outcomes in ICU settings [26].

2.4 Logistic Organ Dysfunction System (LODS)

The Logistic Organ Dysfunction System (LODS) is a scoring system designed to assess organ dysfunction in critically ill patients, particularly in ICUs. LODS quantifies organ dysfunction by evaluating physiological performance across multiple organ systems, providing a comprehensive overview of a patient's clinical status. This system is key in predicting patient outcomes, including mortality, by integrating various clinical parameters into a logistic regression model that calculates the likelihood of organ dysfunction [28].

LODS is particularly relevant in complex ICU environments where patients often present with multi-organ dysfunction. Its capacity to incorporate diverse data types, including physiological measurements and radiological findings, enhances predictive accuracy and clinical utility. Recent advancements have further improved LODS's performance through deep learning models that combine structured time-series data with unstructured clinical notes and expert summaries generated by large language models (LLMs) [29]. This multi-representational learning framework allows for a nuanced understanding of patient conditions, thereby improving mortality predictions and facilitating timely interventions.

Innovations like the adaptive feature importance recalibration module, as seen in systems like AICare, enhance the interpretability and prediction accuracy of LODS. By dynamically adjusting the significance of various clinical features, these advancements ensure that the scoring system remains responsive to the evolving clinical picture of critically ill patients [30]. Additionally, incorporating chest X-ray image features and text representations into LODS models refines the assessment of organ dysfunction, offering a more holistic patient evaluation approach [28].

2.5 Oxford Acute Severity of Illness Score (OASIS)

The Oxford Acute Severity of Illness Score (OASIS) is a predictive tool designed to assess illness severity in critically ill patients within ICUs. OASIS distinguishes itself by effectively incorporating critical variables—such as age, heart rate, mean arterial pressure, temperature, respiratory rate, Glasgow Coma Scale, urine output, and the necessity for mechanical ventilation or vasopressors—into its mortality prediction framework. This streamlined approach enhances predictive accuracy and facilitates the integration of diverse patient data, enabling timely clinical decision-making in intensive care settings. Recent advancements in machine learning and comprehensive electronic health records (EHRs) demonstrate that models leveraging a broader array of uncurated data can outperform traditional scoring systems like OASIS, particularly in dynamic ICU environments [29, 31, 32]. This simplicity facilitates its application across diverse clinical settings, offering an accessible alternative to more complex scoring systems.

OASIS has gained widespread adoption due to its ability to provide rapid assessments without extensive data collection, making it particularly valuable in resource-constrained environments. However, recent studies indicate that advanced methodologies, such as dynamic survival prediction models, can surpass traditional scoring systems like OASIS in predicting in-hospital mortality [32]. These findings highlight the potential for integrating more sophisticated analytical techniques into existing frameworks to enhance predictive accuracy.

The development of OASIS aligns with the broader trend of incorporating multiple clinical parameters into single predictive tools, as exemplified by the MuLBSTA score, which combines factors such as multilobular infiltration and hypo-lymphocytosis [33]. This approach underscores the importance of considering a wide range of clinical indicators to improve illness severity assessment and patient management strategies.

AI-generated, for reference only.

3 Comparative Analysis of Scoring Systems

3.1 Methodological Approaches

The methodologies of scoring systems such as APSIII, SOFA, LODS, and OASIS are pivotal in mortality prediction and risk assessment for critically ill patients. Each system employs unique techniques to analyze patient data, generating risk scores that guide clinical decisions in ICUs. Some systems utilize advanced algorithms, incorporating temporal lab tests and vital signs to facilitate timely ICU admissions, while others leverage multi-modal data, including physiological measurements and radiology reports, alongside deep learning techniques to enhance mortality predictions. Interpretable machine learning models, like Bayesian neural networks, not only forecast outcomes but also emphasize the significance of clinical measurements, aiding clinicians in prioritizing care for high-risk patients. These varied methodologies highlight the need for tailored risk assessment tools to optimize patient outcomes in critical care settings [34, 28, 35].

The SOFA score assesses six major organ systems—respiratory, cardiovascular, hepatic, coagulation, renal, and neurological—to quantify a patient's condition severity [26]. It is instrumental in defining sepsis, evaluating treatment efficacy, and serving as a surrogate endpoint in clinical trials. Machine learning techniques, such as multi-channel Gated Recurrent Units (GRU), enhance its predictive capabilities by aligning feature distributions across hospitals, facilitating collaborative learning [36].

APSIII uses logistic regression to analyze a comprehensive set of physiological variables and patient-specific data, improving predictive accuracy for in-hospital mortality [22]. By integrating dynamic patient information, APSIII adapts to rapidly changing ICU conditions, essential for accurate mortality predictions in complex scenarios like traumatic brain injury [37].

LODS quantifies organ dysfunction by combining various clinical parameters into a logistic regression model [28]. Recent advancements include deep learning models merging structured time-series data with unstructured notes, enhancing mortality prediction through a multi-representational learning framework [29].

OASIS, with its streamlined variable set, allows for rapid illness severity assessments, particularly valuable in resource-constrained settings. While efficient, integrating sophisticated analytical techniques, such as dynamic survival prediction models, could enhance its predictive accuracy [32]. Although its simplicity facilitates widespread adoption, ongoing refinements are necessary to maintain relevance in complex ICU environments.

Innovative methodologies complement these scoring systems, such as the Patient Trajectory Clustering and Risk Prediction Pipeline (PTCRP), which uses dynamic time warping for clustering patient vital sign trajectories, representing a methodological innovation over traditional scoring systems [38]. Benchmarks evaluating predictive models, including clinical scoring methods, traditional machine learning, and advanced EHR-specific models, provide valuable insights into the effectiveness of various methodologies in real-world settings [24].

As illustrated in Figure 2, the hierarchical categorization of methodological approaches in ICU scoring systems highlights the integration of traditional scoring systems, advanced machine learning techniques, and innovative methodologies to enhance mortality prediction and risk assessment in critical care settings. The diversity in methodologies among APSIII, SOFA, LODS, and OASIS reflects their tailored approaches to risk assessment and mortality prediction in ICUs. The integration of advanced machine learning techniques and dynamic data sources continues to enhance their predictive capabilities, supporting more personalized and effective patient care strategies [6].

3.2 Clinical Contexts and Effectiveness

The clinical application of scoring systems like APSIII, SOFA, LODS, and OASIS significantly influences their effectiveness in predicting patient outcomes and guiding clinical decisions. Each system is designed to target specific elements of patient assessment, enhancing utility in various ICU contexts. Transformer-based models have shown promise in predicting multiple clinical outcomes by leveraging unique healthcare data attributes. Deep learning approaches integrating multi-modal data, such as physiological measurements and radiology reports, further improve mortality predictions. These innovations emphasize the importance of tailored assessment tools that provide critical insights in diverse ICU scenarios, leading to better patient management and outcomes [28, 4].

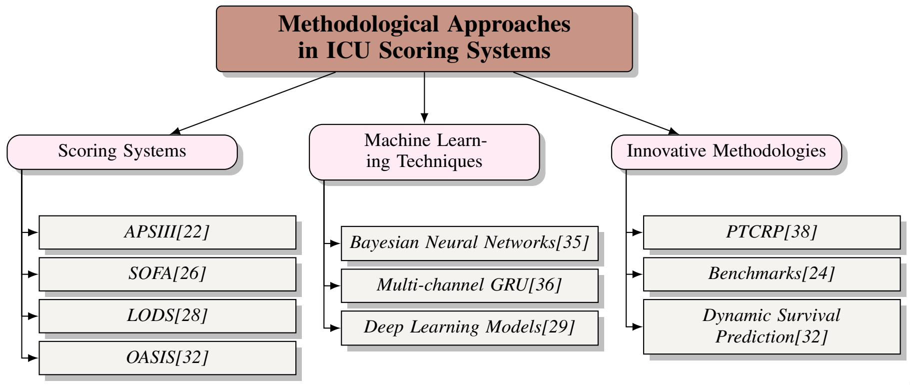


Figure 2: This figure illustrates the hierarchical categorization of methodological approaches in ICU scoring systems, highlighting the integration of traditional scoring systems, advanced machine learning techniques, and innovative methodologies to enhance mortality prediction and risk assessment in critical care settings.

APSIII excels in contexts requiring comprehensive physiological assessments for accurate in-hospital mortality predictions. Its integration of dynamic patient data allows adaptation to rapidly changing clinical conditions, proving highly effective in managing complex cases like traumatic brain injuries and acute kidney injuries. The inclusion of a wide range of physiological variables ensures detailed risk assessments, crucial for optimizing resource allocation and improving patient outcomes in diverse ICU settings [21].

The SOFA score is effective in evaluating organ dysfunction severity, particularly in patients with sepsis or multi-organ failure [26]. Its straightforward application enables quick assessments of six major organ systems, making it valuable in clinical trials and treatment efficacy evaluations, providing a reliable surrogate endpoint for patient morbidity [26]. Machine learning techniques, such as multi-channel GRU, enhance its predictive capabilities through collaborative learning across different hospital settings [36].

LODS is relevant in ICU environments where patients present with complex multi-organ dysfunction. Its ability to incorporate diverse data types, including physiological measurements and radiology findings, enhances predictive accuracy and clinical utility [28]. The integration of deep learning models combining structured time-series data with unstructured clinical notes allows LODS to provide nuanced understandings of patient conditions, improving mortality predictions and facilitating timely interventions [29].

OASIS is effective in resource-constrained environments requiring rapid illness severity assessments. Its streamlined variable set allows for quick and efficient mortality predictions, guiding clinical decision-making in diverse ICU settings [32]. However, integrating more sophisticated analytical techniques, such as dynamic survival prediction models, could enhance its predictive accuracy, ensuring continued relevance in increasingly complex clinical environments.

The effectiveness of these scoring systems is underscored by their application in specific disease contexts. For instance, the dTIC method effectively identifies distinct patient phenotypes influencing clinical outcomes, providing valuable insights into patient management strategies [39]. The use of artificial intelligence to analyze CXR images and clinical data in COVID-19 patients exemplifies the potential of these systems to predict clinical outcomes effectively [40]. The correlation between specific ECG patterns and seizure occurrences, demonstrated in machine learning-supported triage methods, underscores the importance of non-invasive monitoring in enhancing patient care [41].

3.3 Predictive Accuracy and Limitations

The predictive accuracy and limitations of scoring systems such as APSIII, SOFA, LODS, and OASIS are crucial for their application in mortality prediction and risk assessment in ICUs. Table 1 presents a detailed comparison of benchmarks used in the assessment of predictive accuracy and limitations of scoring systems in clinical informatics and other healthcare domains. Each system exhibits distinct advantages and limitations that significantly affect their practical applications in healthcare

AI-generated, for reference only.

Benchmark	Size	Domain	Task Format	Metric
LEB[18]	8,280,820	Clinical Informatics	Mortality Prediction	ROC AUC
DBN[42]	24,506	Critical Care	Mortality Prediction	AUROC, F1
MySurgeryRisk[43]	150	Surgery	Risk Assessment	AUC
HF-ICU[44]	1,177	Cardiology	Mortality Prediction	AUC-ROC, Accuracy
SOFA[45]	3,233	Cardiology	Mortality Prediction	AUC, Odds Ratio
BNN-MPM[46]	200,000	Mortality Prediction	Classification	ROC AUC
MIMIC-III[47]	58,576	Critical Care	Mortality Prediction	AUROC, AUPRC
AIFORCOVID[40]	820	Radiology	Binary Classification	Accuracy, Sensitivity

Table 1: This table provides a comprehensive overview of various benchmarks utilized in the evaluation of predictive models for mortality prediction and risk assessment across different domains. It includes information on the size, domain, task format, and performance metric of each benchmark, highlighting their relevance and applicability in clinical settings. The benchmarks are essential for understanding the capabilities and limitations of predictive tools in healthcare.

settings. For example, transformer encoders effectively process clinical texts for decision-making but face challenges related to input length limitations that may result in the loss of critical information. Various contextualized representations, such as BioBERT, have shown promise in identifying clinical outcomes, yet their effectiveness can vary depending on the task and the quality of training data. The GRASP framework provides a standardized method for assessing predictive tools, enabling clinicians to evaluate efficacy based on evidence and implementation potential. Understanding these unique characteristics is essential for optimizing the use of predictive tools in clinical practice [48, 49, 50, 51].

APSIII is renowned for its robust predictive accuracy, primarily due to its comprehensive integration of physiological variables, enhancing performance in predicting in-hospital mortality. However, its reliance on logistic regression limits adaptability to evolving clinical environments and novel data patterns. The introduction of advanced machine learning models, such as CNN-LSTM, has improved mortality predictions significantly, achieving AUC gains of 4

The SOFA score is noted for its substantial predictive accuracy in sepsis and mortality prediction, outperforming simpler models like qSOFA and SIRS, with AUROC values indicating superior diagnostic capabilities [26]. While SOFA provides reliable predictions for both in-hospital and 30-day mortality, it is sometimes outperformed by more complex models like APACHE-IV and SAPS-II in specific contexts. Its simplicity and ease of use, while advantageous, limit its ability to capture the dynamic and complex nature of patient conditions. Current studies face limitations related to inter-rater variability and the influence of clinical interventions on SOFA score assessments, affecting consistency in application [26].

LODS benefits from its ability to integrate multimodal data, including physiological measurements and radiology findings, enhancing predictive accuracy compared to traditional scoring systems [52]. However, existing models often struggle with high dimensionality and data imbalances, hindering accurate mortality predictions. Models like EventScore, achieving higher AUROC values while maintaining or improving detection times for clinical events, exemplify LODS's potential to enhance predictive performance [3].

OASIS is valued for its simplicity and rapid assessment capabilities, particularly in resource-constrained environments. However, its predictive accuracy may be limited by the exclusion of interaction terms and complex data patterns that more sophisticated models capture. Advanced models, such as the Flexible EHR Transformer, have demonstrated improved predictive accuracy, achieving a mean AUROC of 0.895 compared to traditional models [4]. Integrating year-specific data representation has improved the robustness of clinical prediction models, making them more applicable in evolving healthcare environments [17].

Traditional survival analysis models, including those underlying some scoring systems, are limited by their reliance on static covariates and the proportional hazards assumption, leading to inaccurate predictions [3]. The application of RNN-LSTM approaches leverages the sequential nature of clinical data, resulting in improved accuracy in mortality prediction compared to traditional methods [52]. The BiLSTM model has demonstrated statistically significant improvements in prediction accuracy across various tasks, highlighting the potential of advanced neural network architectures to enhance the predictive capabilities of existing scoring systems [15].

The approach in [25] shows improved predictive accuracy for ICU mortality by incorporating gastrointestinal function, often overlooked in traditional models. Models were evaluated using logistic regression classifiers trained on embeddings, with performance assessed through ROC AUC scores [18].

3.4 Comparative Performance in Recent Studies

Recent studies provide insights into the comparative performance of scoring systems like APSIII, SOFA, LODS, and OASIS in predicting mortality and assessing risk in critically ill patients. The SOFA score has been extensively evaluated, revealing differences in methodology and findings related to patient outcomes and mortality. These analyses highlight the versatility of the SOFA score across various clinical contexts, although variability in measurement techniques can influence predictive accuracy [26].

In a comparative analysis, the Gastrointestinal-Enhanced SOFA Model (GESM) demonstrated improved performance over traditional SOFA models. By incorporating gastrointestinal function, GESM achieved higher Area Under the Curve (AUC) metrics across various machine learning models, underscoring the importance of considering additional physiological parameters to enhance predictive accuracy [25]. This enhancement indicates the potential for refining existing scoring systems by integrating more comprehensive data sources.

Evaluations of APSIII, LODS, and OASIS reveal distinct advantages in their applications. APSIII is noted for its interpretability and personalized risk assessments in acute coronary syndrome scenarios, while LODS excels in leveraging contextualized representations for detecting clinical outcomes in trial abstracts. OASIS has demonstrated competitive performance in predicting ICU readmissions and mortality through the integration of attention mechanisms in deep learning models, enhancing both accuracy and interpretability for clinicians [48, 53, 50]. APSIII's comprehensive integration of physiological variables consistently demonstrates robust predictive accuracy in various clinical scenarios, although its reliance on static logistic regression models may limit adaptability to evolving clinical environments. LODS, with its capability to integrate multimodal data, shows promise in improving mortality predictions, particularly when advanced machine learning techniques are applied.

OASIS, known for its simplicity and rapid assessment capabilities, remains valuable in resource-constrained environments. However, studies suggest that incorporating more sophisticated analytical techniques, such as dynamic survival prediction models, could further enhance its predictive accuracy. Continuous assessment of various ICU scoring systems across diverse clinical environments is essential for identifying potential enhancements, ensuring ongoing applicability, and adapting them to evolving patient care needs. This includes integrating multi-modal data, such as physiological measurements, radiological assessments, and narrative information, which collectively improve the accuracy and relevance of these scoring systems in real-world applications, as demonstrated by recent advancements in predictive models and the established utility of scores like SOFA in clinical trials [26, 28].

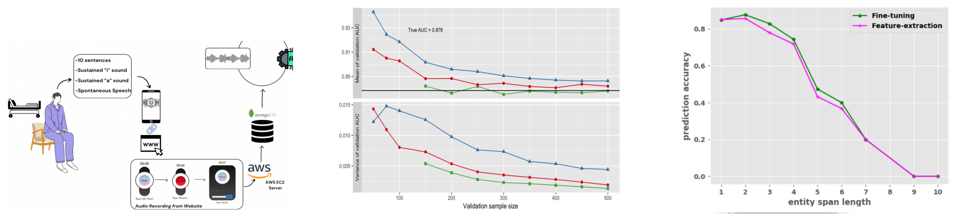
4 Applications in Clinical Outcomes and Risk Assessment

4.1 Specific Applications in Disease Contexts

Scoring systems like APSIII, SOFA, LODS, and OASIS play crucial roles in enhancing clinical outcomes and refining risk assessments in ICUs. APSIII and LODS demonstrate superior predictive capabilities for in-hospital mortality in sepsis-associated acute respiratory failure, aiding clinical decision-making and management strategies [26, 25, 28, 19]. The SOFA score is a prominent surrogate marker in clinical trials, emphasizing its relevance in both individual care and broader healthcare research. It is particularly effective for managing ADHF and sepsis, identifying high-risk patients and enhancing monitoring strategies [54]. Its integration into clinical workflows facilitates timely interventions, improving outcomes, as evidenced by its utility in cardiac surgery patients compared to models like APACHE-IV [55]. Benchmark analyses clarify sepsis criteria applicability in hematological cancer patients, enhancing decision-making [56].

In COVID-19 contexts, predictive models forecast mortality rates, aiding high-risk patient management [21]. Routine blood tests predict disease severity, offering practical management approaches [57]. The Consensus Voting Algorithm integrates multiple algorithm predictions, enhancing sepsis onset prediction and demonstrating integrated approaches' effectiveness [58]. APSIII effectively predicts in-hospital mortality for sepsis-associated ARF, critical for optimizing ICU resource allocation [19]. Its comprehensive integration of physiological variables facilitates accurate risk assessments, with dynamic ICU mortality prediction using time-series data highlighting APSIII's adaptability [59].

Machine learning models predict clinical outcomes across disease contexts. The PPMF approach offers patient-specific predictions using extensive time-series data from the initial ICU stay [11]. The model by [4] predicts ICU readmission and mortality rates, demonstrating practical applications. The LSTM approach provides timely predictions based on early physiological data, improving outcomes and resource allocation [17]. Evaluations by [60] focused on in-ICU mortality, 72-hour ICU re-admission, and 30-day ICU-free days, illustrating broad scoring systems applicability.



(a) Audio Recording System[61]

(b) Validation AUC Performance with Different Sample Sizes[62]

(c) Comparison of Prediction Accuracy for Fine-tuning and Feature-extraction Methods with Different Entity Span Lengths[48]

Figure 3: Examples of Specific Applications in Disease Contexts

Figure 4 illustrates the categorization of scoring systems and predictive models applied in various disease contexts, highlighting ICU scoring systems, COVID-19 specific applications, and machine learning models. Each category emphasizes significant methodologies and their contributions to clinical outcomes and risk assessment. The figure further demonstrates how advanced technologies and methodologies enhance clinical outcomes and risk assessment. For instance, an audio recording system analyzes vocal data for mortality prediction in hospitalized patients. The impact of sample size on validation AUC performance underscores data quantity's necessity for reliable predictive models. Comparing prediction accuracy between fine-tuning and feature-extraction methods across entity span lengths highlights efforts to refine entity detection techniques, essential for accurate disease identification and management. Collectively, these examples underscore technological advancements' importance in improving clinical decision-making and risk assessment.

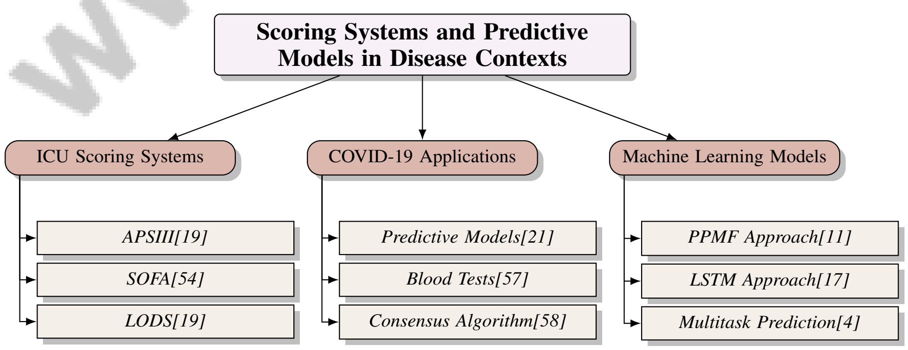


Figure 4: This figure illustrates the categorization of scoring systems and predictive models applied in various disease contexts, highlighting ICU scoring systems, COVID-19 specific applications, and machine learning models. Each category emphasizes significant methodologies and their contributions to clinical outcomes and risk assessment.

AI-generated, for reference only.

4.2 Dynamic Modeling and Real-Time Risk Assessment

Dynamic modeling and real-time risk assessment significantly advance ICUs by enabling timely and accurate patient outcome predictions. These methodologies leverage patient data's evolving nature by integrating structured physiological metrics with unstructured clinical narratives and expert summaries from LLMs. This integration allows real-time mortality risk updates, enhancing clinical decision-making processes. Multimodal learning frameworks improve predictive accuracy while providing interpretable insights, facilitating timely interventions and efficient resource allocation in critical care [15, 29, 63, 64, 50].

Dynamic modeling techniques capture temporal changes in patient conditions, enabling real-time mortality risk prediction updates [32]. This capability ensures clinical decisions are informed by the most current patient data, valuable in ICUs' rapidly changing environments. Real-time risk assessment models, like PTIM, exemplify these methodologies' practical application, providing real-time mortality predictions for timely and informed care decisions [65]. By continuously integrating updated patient data, PTIM ensures accurate and relevant risk assessments, optimizing resource allocation and improving outcomes.

The application of dynamic modeling and real-time risk assessment in ICUs underscores advanced analytical techniques' necessity to enhance predictive accuracy and support clinical decision-making. By integrating temporal data and continuously updating risk predictions, these methodologies improve patient condition understanding, accounting for clinical practices' evolving nature and individual trajectories. This approach fosters more effective and personalized care strategies, evidenced by improved predictive accuracy for adverse outcomes and resource management in dynamic healthcare environments. Models utilizing aggregated clinical features and patient clustering based on health trajectories outperform traditional static models in predicting in-hospital mortality. Incorporating clinical notes through advanced deep learning techniques enhances prediction accuracy and offers interpretable insights, facilitating informed clinical decision-making [66, 38, 50, 67].

In recent years, the application of machine learning and advanced analytics in healthcare has garnered significant attention for its potential to transform patient care and clinical outcomes. Figure 5 illustrates this integration, highlighting four main areas of focus: enhancements in predictive accuracy and interpretability, the utilization of multimodal and unstructured data, the development of novel architectures and algorithms, and the implementation of real-time and dynamic prediction models. Each category not only showcases specific methodologies and technologies but also emphasizes their contributions to improved predictive models and patient care in intensive care units (ICUs). This comprehensive overview serves to underscore the multifaceted nature of machine learning applications in healthcare, suggesting that a holistic approach is essential for maximizing the benefits of these advanced technologies.

5 Integration with Machine Learning and Advanced Analytics

5.1 Enhancements in Predictive Accuracy and Interpretability

Advancements in machine learning (ML) and analytics have markedly improved the predictive accuracy and interpretability of ICU mortality prediction models. The GRASP framework and methods for processing extensive clinical texts using transformer encoders have standardized the selection of predictive tools, ensuring critical information is preserved [51, 49]. Federated learning workflows offer a privacy-preserving alternative to traditional ML, maintaining comparable performance in ICU mortality prediction [68].

Recurrent Neural Networks (RNNs) enhance predictive accuracy by effectively processing complex clinical data, even in the presence of extraneous variables [60]. The RT-Surv framework exemplifies the use of unstructured clinical data to improve model accuracy, particularly in radiotherapy [14]. Integrating wearable sensors with EHRs aims to develop AI-driven acuity assessments, offering nuanced insights for timely interventions [9]. The Federated Learning for Secure Acuity (FLSA) approach further enhances model accuracy and data privacy [66].

5.2 Utilization of Multimodal and Unstructured Data

Incorporating multimodal and unstructured data, such as clinical notes and radiology images, into scoring systems has significantly improved predictive accuracy in ICUs. Enhanced models demonstrate up to 36.41

Self-explaining neural networks utilizing expert-driven concepts, such as SOFA scores, improve interpretability by aligning predictions with clinical frameworks [69]. RiskSLIM exemplifies sparse modeling techniques, optimizing feature selection and interpretability [70]. Generative sequence models capture temporal EHR structures, facilitating robust predictive models adaptable to various clinical scenarios [71].

5.3 Novel Architectures and Algorithms

Method Name	Architectural Innovations	Data Integration	Predictive Methodologies
CHF-AR-HMM[72]	Deep Learning Architectures	Ehr And Wearable	Advanced Algorithms
AI-AAS[9]	Deep Learning Models	Ehr And Accelerometers	Deep Learning Models
SET[73]	Transformer Architectures	Multimodal Ehr Data	Bayesian Neural Network
CVA[58]	Ensemble Approach	Clinical Data	Voting Models

Table 2: Comparison of AI-Driven Predictive Models in Healthcare: This table presents a detailed comparison of various AI-driven predictive models used in healthcare, focusing on their architectural innovations, data integration strategies, and predictive methodologies. The models include CHF-AR-HMM, AI-AAS, SET, and CVA, each of which incorporates unique approaches to enhance clinical predictions and patient management.

The integration of AI with scoring systems has led to novel architectures and algorithms, enhancing predictive accuracy and clinical utility in ICUs. Innovations include models learning survival parameters in real-time, reflecting changing mortality risks more accurately than static models [72]. Incorporating mobility data from wearable sensors enhances acuity assessments, offering a comprehensive view of patient health [9].

AI-driven architectures, including deep learning models, process complex EHR data, identifying significant patterns to improve scoring systems' interpretability and application. The EBM-COMET dataset enhances clinical outcome detection using fine-tuned contextualized representations like BioBERT, achieving high accuracy in outcome phrase identification. New methodologies for risk prediction in acute coronary syndrome scenarios combine traditional scoring with ML techniques, delivering personalized and reliable predictions [48, 53, 51].

Table 2 provides a comprehensive overview of the novel AI-driven architectures and algorithms that are enhancing predictive accuracy and clinical utility in Intensive Care Units (ICUs).

As shown in Figure 6, ML and advanced analytics integration is transforming healthcare. The first figure illustrates LOS distribution, enhancing understanding of patient management over time. The second highlights ML's prediction accuracy through MI versus predicted probability plots. The third figure shows CDF plots for various medical measurements, illustrating advanced analytics' role in patient health insights [47, 73, 58].

5.4 Real-time and Dynamic Prediction Models

AI integration in ICUs has significantly advanced real-time and dynamic prediction models, enabling continuous refinement of forecasts with the latest patient data. These models support evidence-based decisions by enhancing prediction accuracy through patient representation learning from EHRs and attention mechanisms in clinical notes [67, 50, 71, 64].

Real-time models process diverse data sources, including physiological signals and clinical notes, for continuous risk assessments, crucial for timely ICU interventions. RNNs and LSTMs capture temporal dependencies, enhancing real-time prediction accuracy [60]. Dynamic models benefit from multimodal data integration, providing comprehensive patient health insights [9].

Federated learning techniques further enable real-time model development, maintaining high predictive performance while preserving patient privacy. This approach allows decentralized data training, ensuring secure data handling and contributing to overall prediction accuracy [68].

6 Challenges and Future Directions

Advancing scoring systems for mortality prediction and risk assessment in ICUs involves overcoming challenges related to data quality and integration, which are crucial for robust predictive models and effective clinical decision-making. A detailed examination of these factors is vital for enhancing predictive capabilities in critical care.

6.1 Data Quality and Integration

Data quality and integration present significant challenges in deploying ICU scoring systems. Variability in EHR data quality undermines model robustness, while selective clinical descriptors and extensive preprocessing can introduce biases, impacting prediction accuracy [4]. Fragmented clinical notes and inconsistent documentation further complicate data quality, affecting prediction precision [52]. The high dimensionality and sparsity of EHR data complicate the transformation of patient records into fixed-length representations needed by conventional algorithms [71]. Multimodal data integration remains challenging, limiting predictive capabilities [3]. Incomplete parameter collection, especially in dynamic contexts like COVID-19, restricts model applicability. Addressing missing values through imputation can affect prediction accuracy and interpretability. Privacy regulations that restrict inter-hospital data sharing necessitate exploring decentralized methods for improved data integration [68]. Additionally, reliance on consensus definitions for labeling outcomes poses challenges to data quality [74]. Future research should focus on developing standardized SOFA score protocols and innovative methodologies to enhance predictive validity [26]. By refining multimodal data integration techniques and addressing overfitting and missingness, researchers can improve predictive models' accuracy and applicability in critical care.

6.2 Model Interpretability and Complexity

Balancing model interpretability and complexity is crucial for deploying ICU scoring systems. While complex models may enhance predictive accuracy, they often lack the transparency needed for clinical decision-making. The black-box nature of neural networks presents significant challenges, hindering their clinical acceptance [12]. This opacity complicates integration into clinical workflows, where understanding decision-making processes is essential. The computational demands of complex models, particularly in training and inference, remain a concern. Efficient algorithms are needed to provide accurate predictions without excessive computational overhead, ensuring timely application in critical care. Moreover, reliance on specific feature selections may introduce biases, affecting model generalizability across diverse settings [17]. Despite these challenges, advancements in model development show promise in addressing interpretability and complexity issues. Visually interpretable deep learning approaches enable clinicians to understand model outputs without sacrificing predictive performance [75]. However, generalizability across different hospital systems or patient populations can be limited by reliance on single-source data [18]. Comprehensive validation across diverse datasets is necessary to ensure predictive models' robustness and applicability. Enhancing transparency and adaptability will better support clinical decision-making and improve patient outcomes in critical care.

6.3 Generalizability and Validation

Ensuring generalizability and validation of ICU scoring models is challenging due to diverse patient populations and clinical environments. Many studies rely on single-center datasets, limiting findings' applicability across different healthcare settings [25]. Retrospective datasets may further restrict generalizability, as they may not represent all ICU populations or settings [4]. Integrating multiple data sources adds complexity to ensuring model generalizability and validation. Harmonizing diverse data types can affect model robustness, particularly when applied to varying clinical scenarios [3]. Variations in clinical practices and patient demographics across institutions also influence predictive performance. Reliance on specific model assumptions, which may not be universally applicable, complicates generalizability. The retrospective nature of many studies can introduce biases and limit findings' depth, especially when multivariate analyses are unfeasible [4]. Future research should prioritize validating scoring models across various hospital environments and patient demographics. Employing advanced patient representation techniques can improve model accuracy and ensure relevance across populations [48, 71]. Refinements in representation learning techniques, real-time prediction capabilities, improved data imputation methods, and interpretability techniques can enhance understanding of model predictions in clinical settings. By expanding datasets and refining methodologies, researchers can enhance predictive accuracy and applicability of scoring systems, fostering more effective and personalized patient care strategies in ICUs.

6.4 Integration with Clinical Workflows

Integrating scoring systems like APSIII, SOFA, LODS, and OASIS into clinical workflows is essential for optimizing decision-making processes and improving ICU patient care. These systems provide critical insights into patient risk assessment and mortality prediction, yet their seamless incorporation into clinical practice presents challenges and opportunities. Validating and deploying risk assessment services in clinical settings must address complexities to ensure effective integration [6]. Frameworks like GRASP emphasize usability testing and post-implementation evaluations, ensuring scoring systems' accuracy and user-friendliness across clinical settings [49]. Additionally, incorporating dynamic time warping and group-based trajectory modeling in analyzing SOFA scores offers a novel perspective contrasting traditional static assessments, providing a more dynamic understanding of patient trajectories [54]. Significant obstacles remain, such as data quality variability and the need for real-time analytics. Future research should focus on incorporating raw time-series clinical data into models and exploring explainability algorithms to enhance understanding of feature contributions, facilitating advanced scoring systems' integration into clinical workflows [9]. Optimizing model inference times and exploring additional features like imaging data are crucial for enhancing prediction capabilities and facilitating integration [76]. The potential for scoring systems to support real-time decision-making is underscored by the need for multicenter studies to validate findings and explore applicability in diverse patient populations, including those affected by COVID-19 [77]. Integrating comprehensive datasets and refining preprocessing techniques are critical for improving prediction capabilities [78].

6.5 Privacy-preserving Techniques

Integrating privacy-preserving techniques in scoring systems for mortality prediction and risk assessment is crucial given stringent privacy regulations governing clinical data sharing. The FedScore framework exemplifies a significant advancement by addressing privacy concerns through federated learning, allowing clinical scoring system development without centralized data storage [79]. This approach mitigates data breach risks by keeping patient data within local healthcare institutions, ensuring privacy compliance. The method proposed by [36] emphasizes protecting patient privacy during data sharing, a crucial challenge in deploying scoring systems. Utilizing cross-center learning methodologies facilitates collaborative model development while preserving patient confidentiality. The effectiveness of privacy-preserving methods is enhanced by encryption techniques to maintain data privacy, enabling collaborative learning across multiple institutions and addressing traditional centralized approaches' challenges while protecting sensitive patient data [66]. Future research should explore methods to enhance algorithm diversity and independence, as suggested by [58]. Incorporating additional regularization techniques can improve the applicability and performance of dynamic prediction models like DySurv in clinical settings [80]. Advancing privacy-preserving techniques and enhancing predictive model robustness will improve scoring systems' integration into clinical workflows while safeguarding patient privacy.

6.6 Explainability and Interpretability

Explainability and interpretability are critical in developing and deploying advanced predictive models in healthcare, particularly in ICUs. These aspects ensure healthcare providers can trust and effectively utilize predictive models in clinical decision-making. The intricate nature of advanced predictive models, often described as "black-box" systems, presents challenges in deciphering prediction mechanisms. This lack of transparency can impede acceptance and integration into clinical workflows. Recent advancements highlight the importance of developing machine learning approaches that enhance predictive performance while prioritizing interpretability and reliability. For instance, new methodologies combining traditional risk score models with advanced machine learning techniques offer personalized predictions and clear visualizations of critical clinical information, fostering greater confidence among healthcare professionals and facilitating adoption in clinical practice [67, 53, 50]. Explainability is underscored by models providing transparent insights into prediction-influencing factors. Transparency in risk assessment methodologies is essential for clinicians, enabling comprehension of underlying rationale and informed decision-making about patient care. Integrating interpretability features into risk prediction models empowers clinicians to utilize these tools effectively, supporting optimizing clinical interventions and enhancing patient outcomes [70, 6, 43, 53, 50]. Models providing clear explanations enable healthcare providers to validate predictions against clinical knowledge and experience, enhancing trust and reliability. Interpretability focuses on presenting findings comprehensibly to clinicians, simplifying complex algorithms, and ensuring outputs align with familiar clinical concepts and terminologies. For instance, the FedScore method demonstrates privacy-preserving frameworks' potential to maintain interpretability while aggregating insights from diverse data sources, enhancing predictive models' generalizability and stability [79]. Future developments in predictive modeling should prioritize integrating explainability and interpretability features, such as visualizations and concept-based explanations, to bridge the gap between complex algorithms and clinical practice. By integrating advanced deep learning models leveraging unstructured clinical notes and multimodal data, clinicians can enhance decision-making processes with interpretable insights, leading to more accurate and timely interventions. This approach improves patient outcomes in critical care environments and optimizes resource allocation by accurately predicting risks like ICU readmission and mortality. Incorporating attention mechanisms and flexible Transformer architectures ensures clinically relevant information is highlighted, allowing for a comprehensive understanding of patient conditions and needs [51, 28, 50, 4].

7 Conclusion

Scoring systems like APSIII, SOFA, LODS, and OASIS play a crucial role in enhancing clinical outcomes and decision-making in Intensive Care Units (ICUs). These systems provide structured frameworks for assessing patient risk and predicting mortality, enabling timely interventions and efficient resource management. The integration of advanced machine learning techniques has notably improved the predictive accuracy and interpretability of these systems, as demonstrated by models that offer interpretable insights, which are essential for building clinical trust.

Innovative methodologies, such as the dTIC model, have successfully identified patient phenotypes with distinct clinical outcomes, underscoring the importance of personalized care strategies in critical settings. Additionally, the use of multimodal data, combining clinical and imaging data, has shown superior predictive capabilities for specific conditions, such as pulmonary embolism, compared to traditional scoring methods.

The potential of artificial intelligence to transform clinical decision-making and improve patient outcomes is evident. Future research should focus on optimizing these models with larger datasets and expanding their application to various clinical scenarios, including early triage and risk assessment. The observed improvements in survival rates among cancer patients in ICUs highlight the progress in clinical care facilitated by these predictive models.

Future efforts should emphasize the validation of frameworks across diverse predictive tools and enhance their accessibility for clinical users. The iterative development process, guided by a structured roadmap, is vital for creating effective risk assessment services and ensuring their ongoing improvement. Furthermore, the integration of diverse clinical data sources is crucial for advancing decision support systems in mortality prediction, thereby enhancing patient care in critical care environments.

References

[1] Yukti Makhija, Samarth Bhatia, Shalendra Singh, Sneha Kumar Jayaswal, Prabhat Singh Malik, Pallavi Gupta, Shreyas N. Samaga, Shreya Johri, Sri Krishna Venigalla, Rabi Narayan Hota, Surinder Singh Bhatia, and Ishaan Gupta. Challenges in the application of a mortality prediction model for covid-19 patients on an indian cohort, 2021.

[2] Hossein Estiri, Zachary H. Strasser, and Shawn N. Murphy. Individualized prediction of covid-19 adverse outcomes with mlho, 2020.

[3] Isaac Mativo, Yelena Yesha, Michael Grasso, Tim Oates, and Qian Zhu. Hybrid mortality prediction using multiple source systems, 2019.

[4] Benjamin Shickel, Patrick J. Tighe, Azra Bihorac, and Parisa Rashidi. Multi-task prediction of clinical outcomes in the intensive care unit using flexible multimodal transformers, 2021.

[5] Qizhang Feng, Jiayi Yuan, Forhan Bin Emdad, Karim Hanna, Xia Hu, and Zhe He. Can attention be used to explain ehr-based mortality prediction tasks: A case study on hemorrhagic stroke, 2023.

[6] Eryu Xia, Yiqin Yu, Enliang Xu, Jing Mei, and Wen Sun. From risk prediction models to risk assessment service: A formulation of development paradigm, 2019.

[7] Khalid Alghatani, Nariman Ammar, Abdelmounaam Rezgui, and Arash Shaban-Nejad. Predicting intensive care unit length of stay and mortality using patient vital signs: Machine learning model development and validation, 2021.

[8] Xiaoli Liu, Pan Hu, Zhi Mao, Po-Chih Kuo, Peiyao Li, Chao Liu, Jie Hu, Deyu Li, Desen Cao, Roger G. Mark, Leo Anthony Celi, Zhengbo Zhang, and Feihu Zhou. Interpretable machine learning model for early prediction of mortality in elderly patients with multiple organ dysfunction syndrome (mods): a multicenter retrospective study and cross validation, 2020.

[9] Jessica Sena, Mohammad Tahsin Mostafiz, Jiaqing Zhang, Andrea Davidson, Sabyasachi Bandyopadhyay, Ren Yuanfang, Tezcan Ozrazgat-Baslanti, Benjamin Shickel, Tyler Loftus, William Robson Schwartz, Azra Bihorac, and Parisa Rashidi. The potential of wearable sensors for assessing patient acuity in intensive care unit (icu), 2023.

[10] Xingqiao Li, Jindong Gu, Zhiyong Wang, Yancheng Yuan, Bo Du, and Fengxiang He. Xai for in-hospital mortality prediction via multimodal icu data, 2023.

[11] Mohammad Amin Morid, Olivia R. Liu Sheng, and Samir Abdelrahman. Ppmf: A patient-based predictive modeling framework for early icu mortality prediction, 2017.

[12] Yohan Jo, Lisa Lee, and Shruti Palaskar. Combining lstm and latent topic modeling for mortality prediction, 2017.

[13] Zhale Nowroozilarki, Arash Pakbin, James Royalty, Donald K. K. Lee, and Bobak J. Mortazavi. Real-time mortality prediction using mimic-iv icu data via boosted nonparametric hazards, 2021.

[14] Sangjoon Park, Chan Woo Wee, Seo Hee Choi, Kyung Hwan Kim, Jee Suk Chang, Hong In Yoon, Ik Jae Lee, Yong Bae Kim, Jaeho Cho, Ki Chang Keum, Chang Geol Lee, Hwa Kyung Byun, and Woong Sub Koom. Improving mortality prediction after radiotherapy with large language model structuring of large-scale unstructured electronic health records, 2024.

[15] Paulina Grnarova, Florian Schmidt, Stephanie L. Hyland, and Carsten Eickhoff. Neural document embeddings for intensive care patient mortality prediction, 2016.

[16] Ahmed M. Alaa, Jinsung Yoon, Scott Hu, and Mihaela van der Schaar. A semi-markov switching linear gaussian model for censored physiological data, 2016.

[17] Manel Mili, Asma Kerkeni, Asma Ben Abdallah, and Mohamed Hedi Bedoui. Icu mortality prediction using long short-term memory networks, 2023.

[18] Lorenzo A. Rossi, Chad Shawber, Janet Munu, and Finly Zachariah. Evaluation of embeddings of laboratory test codes for patients at a cancer center, 2019.

[19] Shiqin Fan and Jing Ma. The value of five scoring systems in predicting the prognosis of patients with sepsis-associated acute respiratory failure. *Scientific Reports*, 14(1):4760, 2024.

[20] Wioletta Sawicka, Radosław Owczuk, Magdalena Wujtewicz, and Maria Wujtewicz. The effectiveness of the apache ii, saps ii and sofa prognostic scoring systems in patients with haematological malignancies in the intensive care unit. *Anestezjologia Intensywna Terapia*, 46(3):177–182, 2019.

[21] Xiaojing Zou, Shusheng Li, Minghao Fang, Ming Hu, Yi Bian, Jianmin Ling, Shanshan Yu, Liang Jing, Donghui Li, and Jiao Huang. Acute physiology and chronic health evaluation ii score as a predictor of hospital mortality in patients of coronavirus disease 2019. *Critical care medicine*, 48(8):e657–e665, 2020.

[22] Zirun Zhao, Anne Chen, Wei Hou, James M Graham, Haifang Li, Paul S Richman, Henry C Thode, Adam J Singer, and Tim Q Duong. Prediction model and risk scores of icu admission and mortality in covid-19. *PloS one*, 15(7):e0236618, 2020.

[23] Yixuan Li, Archer Y. Yang, Ariane Marelli, and Yue Li. Mixehr-surg: a joint proportional hazard and guided topic model for inferring mortality-associated topics from electronic health records, 2024.

[24] Junyi Gao, Yinghao Zhu, Wenqing Wang, Yasha Wang, Wen Tang, Ewen M. Harrison, and Liantao Ma. A comprehensive benchmark for covid-19 predictive modeling using electronic health records in intensive care, 2024.

[25] Yehudit Aperstein, Lidor Cohen, Itai Bendavid, Jonathan Cohen, Elad Grozovsky, Tammy Rotem, and Pierre Singer. Improved icu mortality prediction based on sofa scores and gastrointestinal parameters. *PLoS One*, 14(9):e0222599, 2019.

[26] Simon Lambden, Pierre Francois Laterre, Mitchell M Levy, and Bruno Francois. The sofa score—development, utility and challenges of accurate assessment in clinical trials. *Critical Care*, 23:1–9, 2019.

[27] Eleni Karakike, Evdoxia Kyriazopoulou, Iraklis Tsangaris, Christina Routsi, Jean-Louis Vincent, and Evangelos J Giamarellos-Bourboulis. The early change of sofa score as a prognostic marker of 28-day sepsis mortality: analysis through a derivation and a validation cohort. *Critical care*, 23:1–8, 2019.

[28] Mingquan Lin, Song Wang, Ying Ding, Lihui Zhao, Fei Wang, and Yifan Peng. An empirical study of using radiology reports and images to improve icu mortality prediction, 2023.

[29] Harshavardhan Battula, Jiacheng Liu, and Jaideep Srivastava. Enhancing in-hospital mortality prediction using multi-representational learning with llm-generated expert summaries, 2024.

[30] Liantao Ma, Chaohe Zhang, Junyi Gao, Xianfeng Jiao, Zhihao Yu, Xinyu Ma, Yasha Wang, Wen Tang, Xinju Zhao, Wenjie Ruan, and Tao Wang. Mortality prediction with adaptive feature importance recalibration for peritoneal dialysis patients: a deep-learning-based study on a real-world longitudinal follow-up dataset, 2023.

[31] Hoda Memarzadeh, Nasser Ghadiri, and Maryam Lotfi Shahreza. Assessing mortality prediction through different representation models based on concepts extracted from clinical notes, 2022.

[32] Jacob Deasy, Pietro Liò, and Ari Ercole. Dynamic survival prediction in intensive care units from heterogeneous time series without the need for variable selection or pre-processing, 2019.

[33] Lingxi Guo, Dong Wei, Xinxin Zhang, Yurong Wu, Qingyun Li, Min Zhou, and Jieming Qu. Clinical features predicting mortality risk in patients with viral pneumonia: the mulbsta score. *Frontiers in microbiology*, 10:2752, 2019.

[34] Ahmed M. Alaa, Jinsung Yoon, Scott Hu, and Mihaela van der Schaar. Personalized risk scoring for critical care prognosis using mixtures of gaussian processes, 2016.

[35] Hiske Overweg, Anna-Lena Popkes, Ari Ercole, Yingzhen Li, José Miguel Hernández-Lobato, Yordan Zaykov, and Cheng Zhang. Interpretable outcome prediction with sparse bayesian neural networks in intensive care, 2019.

[36] Ruiqing Ding, Fangjie Rong, Xiao Han, and Leye Wang. Cross-center early sepsis recognition by medical knowledge guided collaborative learning for data-scarce hospitals, 2023.

[37] Lucas Steinmetz, Shivam Maheshwari, Garik Kazanjian, Abigail Loyson, Tyler Alexander, Venkat Margapuri, and C. Nataraj. Predicting mortality and functional status scores of traumatic brain injury patients using supervised machine learning, 2024.

[38] Thea Barnes, Enrico Werner, Jeffrey N. Clark, and Raul Santos-Rodriguez. Towards personalised patient risk prediction using temporal hospital data trajectories, 2024.

[39] Yuanfang Ren, Yanjun Li, Tyler J. Loftus, Jeremy Balch, Kenneth L. Abbott, Shounak Datta, Matthew M. Ruppert, Ziyuan Guan, Benjamin Shickel, Parisa Rashidi, Tezcan Ozrazgat-Baslanti, and Azra Bihorac. Identifying acute illness phenotypes via deep temporal interpolation and clustering network on physiologic signatures, 2023.

[40] Paolo Soda, Natascha Claudia D'Amico, Jacopo Tessadori, Giovanni Valbusa, Valerio Guarrasi, Chandra Bortolotto, Muhammad Usman Akbar, Rosa Sicilia, Ermanno Cordelli, Deborah Fazzini, Michaela Cellina, Giancarlo Oliva, Giovanni Callea, Silvia Panella, Maurizio Cariati, Diletta Cozzi, Vittorio Miele, Elvira Stellato, Gian Paolo Carrafiello, Giulia Castorani, Annalisa Simeone, Lorenzo Preda, Giulio Iannello, Alessio Del Bue, Fabio Tedoldi, Marco Alì, Diego Sona, and Sergio Papa. Aiforcovid: predicting the clinical outcomes in patients with covid-19 applying ai to chest-x-rays. an italian multicentre study, 2020.

[41] Raphael Azriel, Cecil D. Hahn, Thomas De Cooman, Sabine Van Huffel, Eric T. Payne, Kristin L. McBain, Danny Eytan, and Joachim A. Behar. Machine learning to support triage of children at risk for epileptic seizures in the pediatric intensive care unit, 2022.

[42] Tony Wang, Tom Velez, Emilia Apostolova, Tim Tschampel, Thuy L. Ngo, and Joy Hardison. Semantically enhanced dynamic bayesian network for detecting sepsis mortality risk in icu patients with infection, 2018.

[43] Meghan Brennan, Sahil Puri, Tezcan Ozrazgat-Baslanti, Rajendra Bhat, Zheng Feng, Petar Momcilovic, Xiaolin Li, Daisy Zhe Wang, and Azra Bihorac. Comparing clinical judgment with mysurgeryrisk algorithm for preoperative risk assessment: A pilot study, 2018.

[44] Negin Ashrafi, Armin Abdollahi, Jiahong Zhang, and Maryam Pishgar. Optimizing mortality prediction for icu heart failure patients: Leveraging xgboost and advanced machine learning with the mimic-iii database, 2024.

[45] Adi Elias, Reham Agbarieh, Walid Saliba, Johad Khoury, Fadel Bahouth, Jeries Nashashibi, and Zaher S Azzam. Sofa score and short-term mortality in acute decompensated heart failure. *Scientific reports*, 10(1):20802, 2020.

[46] Jacob R. Epifano, Stephen Glass, Ravi P. Ramachandran, Sharad Patel, Aaron J. Masino, and Ghulam Rasool. Deployment of a robust and explainable mortality prediction model: The covid-19 pandemic and beyond, 2023.

[47] Sanjay Purushotham, Chuizheng Meng, Zhengping Che, and Yan Liu. Benchmark of deep learning models on large healthcare mimic datasets, 2017.

[48] Micheal Abaho, Danushka Bollegala, Paula R Williamson, and Susanna Dodd. Assessment of contextualised representations in detecting outcome phrases in clinical trials, 2022.

[49] Mohamed Khalifa, Farah Magrabi, and Blanca Gallego. Developing an evidence-based framework for grading and assessment of predictive tools for clinical decision support, 2019.

[50] Justin R. Lovelace, Nathan C. Hurley, Adrian D. Haimovich, and Bobak J. Mortazavi. Explainable prediction of adverse outcomes using clinical notes, 2019.

[51] Mohammad Junayed Hasan, Suhra Noor, and Mohammad Ashrafuzzaman Khan. On preserving the knowledge of long clinical texts, 2025.

[52] Satya Narayan Shukla and Benjamin M. Marlin. Integrating physiological time series and clinical notes with deep learning for improved icu mortality prediction, 2021.

[53] Francisco Valente, Jorge Henriques, Simão Paredes, Teresa Rocha, Paulo de Carvalho, and João Morais. A new approach for interpretability and reliability in clinical risk prediction: Acute coronary syndrome scenario, 2021.

[54] Yuhe Ke, Matilda Swee Sun Tang, Celestine Jia Ling Loh, Hairil Rizal Abdullah, and Nicholas Brian Shannon. Cluster trajectory of sofa score in predicting mortality in sepsis, 2023.

[55] Abraham Schoe, Ferishta Bakhshi-Raiez, Nicolette de Keizer, Jaap T van Dissel, and Evert de Jonge. Mortality prediction by sofa score in icu-patients after cardiac surgery; comparison with traditional prognostic–models. *BMC anesthesiology*, 20:1–8, 2020.

[56] Lucie Probst, Enrico Schalk, Tobias Liebregts, Vanja Zeremski, Asterios Tzalavras, Michael von Bergwelt-Baildon, Nina Hesse, Johanna Prinz, Jörg Janne Vehreschild, Alexander Shimabukuro-Vornhagen, et al. Prognostic accuracy of sofa, qsofa and sirs criteria in hematological cancer patients: a retrospective multicenter study. *Journal of intensive care*, 7:1–10, 2019.

[57] Sakifa Aktar, Md. Martuza Ahamad, Md. Rashed-Al-Mahfuz, AKM Azad, Shahadat Uddin, A H M Kamal, Salem A. Alyami, Ping-I Lin, Sheikh Mohammed Shariful Islam, Julian M. W. Quinn, Valsamma Eapen, and Mohammad Ali Moni. Predicting patient covid-19 disease severity by means of statistical and machine learning analysis of blood cell transcriptome data, 2020.

[58] Matthew A. Reyna and Gari D. Clifford. Voting of predictive models for clinical outcomes: consensus of algorithms for the early prediction of sepsis from clinical data and an analysis of the physionet/computing in cardiology challenge 2019, 2020.

[59] Tiago Alves, Alberto Laender, Adriano Veloso, and Nivio Ziviani. Dynamic prediction of icu mortality risk using domain adaptation, 2019.

[60] Eugene Laksana, Melissa Aczon, Long Ho, Cameron Carlin, David Ledbetter, and Randall Wetzel. The impact of extraneous variables on the performance of recurrent neural network models in clinical tasks, 2019.

[61] Nihat Ahmadli, Mehmet Ali Sarsil, Berk Mizrak, Kurtulus Karauzum, Ata Shaker, Erol Tulumen, Didar Mirzamidinov, Dilek Ural, and Onur Ergen. Voice-driven mortality prediction in hospitalized heart failure patients: A machine learning approach enhanced with diagnostic biomarkers, 2024.

[62] W. Katherine Tan and Patrick J. Heagerty. Surrogate-guided sampling designs for classification of rare outcomes from electronic medical records data, 2020.

[63] Mengqi Jin, Mohammad Taha Bahadori, Aaron Colak, Parminder Bhatia, Busra Celikkaya, Ram Bhakta, Selvan Senthivel, Mohammed Khalilia, Daniel Navarro, Borui Zhang, Tiberiu Doman, Arun Ravi, Matthieu Liger, and Taha Kass-hout. Improving hospital mortality prediction with medical named entities and multimodal learning, 2018.

[64] Chenxi Sun, Hongyan Li, Moxian Song, Derun Cai, Baofeng Zhang, and Shenda Hong. Continuous diagnosis and prognosis by controlling the update process of deep neural networks, 2022.

[65] Adam J. Starr, Manjula Julka, Arun Nethi, John D. Watkins, Ryan W. Fairchild, Michael W. Cripps, Dustin Rinehart, and Hayden N. Box. Parkland trauma index of mortality (ptim): Real-time predictive model for polytrauma patients, 2020.

[66] Zichang Wang, Haoran Li, Luchen Liu, Haoxian Wu, and Ming Zhang. Predictive multi-level patient representations from electronic health records, 2019.

[67] Bret Nestor, Matthew B. A. McDermott, Geeticka Chauhan, Tristan Naumann, Michael C. Hughes, Anna Goldenberg, and Marzyeh Ghassemi. Rethinking clinical prediction: Why machine learning must consider year of care and feature aggregation, 2018.

[68] Lena Mondrejevski, Ioanna Miliou, Annaclaudia Montanino, David Pitts, Jaakko Hollmén, and Panagiotis Papapetrou. Flicu: A federated learning workflow for intensive care unit mortality prediction, 2022.

[69] Sayantan Kumar, Sean C. Yu, Thomas Kannampallil, Zachary Abrams, Andrew Michelson, and Philip R. O. Payne. Self-explaining neural network with concept-based explanations for icu mortality prediction, 2022.

[70] Hans-Jürgen Profitlich and Daniel Sonntag. Interactivity and transparency in medical risk assessment with supersparse linear integer models, 2019.

[71] Ethan Steinberg, Ken Jung, Jason A. Fries, Conor K. Corbin, Stephen R. Pfohl, and Nigam H. Shah. Language models are an effective patient representation learning technique for electronic health record data, 2020.

[72] Yilin Yin and Chun-An Chou. Early icu mortality prediction and survival analysis for respiratory failure, 2021.

[73] Adrian Lindenmeyer, Malte Blattmann, Stefan Franke, Thomas Neumuth, and Daniel Schneider. Inadequacy of common stochastic neural networks for reliable clinical decision support, 2024.

[74] Michael Staniek, Marius Fracarolli, Michael Hagmann, and Stefan Riezler. Early prediction of causes (not effects) in healthcare by long-term clinical time series forecasting, 2024.

[75] William Caicedo-Torres and Jairo Gutierrez. Iseeu: Visually interpretable deep learning for mortality prediction inside the icu, 2019.

[76] Mohammad Dehghani, Mobin Mohammadi, and Diyana Tehrany Dehkordy. Deep neural decision forest: A novel approach for predicting recovery or decease of patients, 2024.

[77] Sijia Liu, Ni Yao, Yanru Qiu, and Chengqi He. Predictive performance of sofa and qsofa for in-hospital mortality in severe novel coronavirus disease. *The American Journal of Emergency Medicine*, 38(10):2074–2080, 2020.

[78] Huachuan Wang and Yuanfei Bi. Building deep learning models to predict mortality in icu patients, 2020.

[79] Siqi Li, Yilin Ning, Marcus Eng Hock Ong, Bibhas Chakraborty, Chuan Hong, Feng Xie, Han Yuan, Mingxuan Liu, Daniel M. Buckland, Yong Chen, and Nan Liu. Fedscore: A privacy-preserving framework for federated scoring system development, 2023.

[80] Munib Mesinovic, Peter Watkinson, and Tingting Zhu. Dysurv: dynamic deep learning model for survival analysis with conditional variational inference, 2024.

Disclaimer:

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.

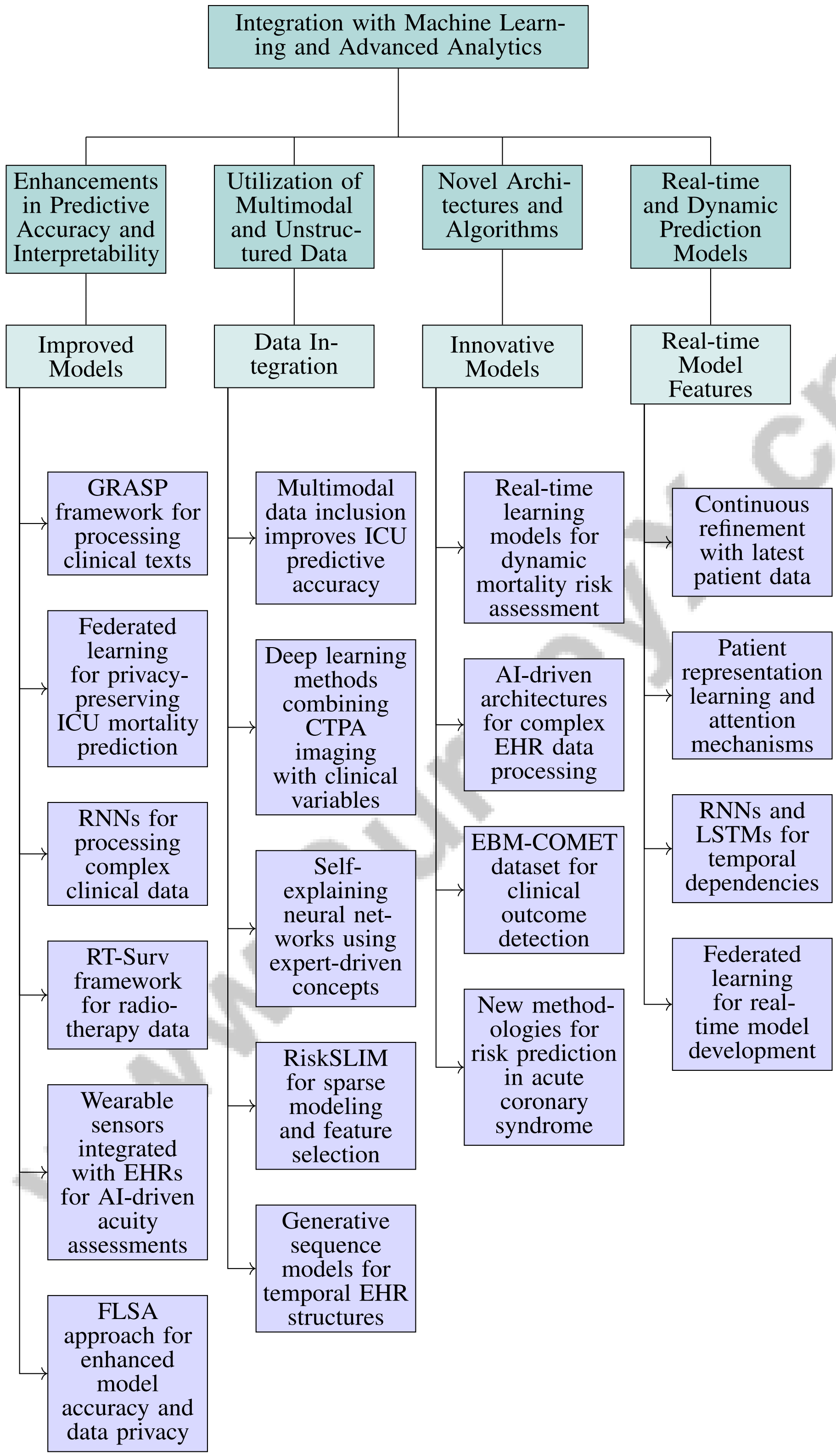
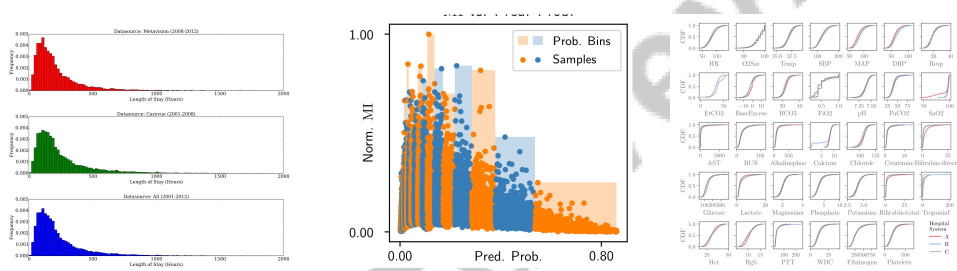


Figure 5: This figure illustrates the integration of machine learning and advanced analytics in healthcare, highlighting four main areas: enhancements in predictive accuracy and interpretability, utilization of multimodal and unstructured data, novel architectures and algorithms, and real-time and dynamic prediction models. Each category showcases specific methodologies and technologies contributing to improved predictive models and patient care in ICUs.



(a) Length of Stay Distribution in Healthcare Data[47]

(b) MI vs. Predicted Prob.[73]

(c) CDF plots for various medical measurements across different hospitals[58]

Figure 6: Examples of Novel Architectures and Algorithms