ISSN 2288-4866 (Print)
ISSN 2288-4882 (Online)
https://www.jiisonline.org
J Intell Inform Syst 2024 June: 30(2): 25~59
http://dx.doi.org/10.13088/jiis.2024.30.2.025

Integrating Digital Twin Technology with Dynamic Ensemble Learning for Sepsis Prediction in Intensive Care Units*

Amirhossein Danesh
College of Computing and Informatics, Sungkyunkwan University
(amir.danesh@g.skku.edu)

Firuz Juraev
College of Computing and Informatics, Sungkyunkwan University
(fjuraev@g.skku.edu)

Shaker El-Sappagh
College of Computing and Informatics, Sungkyunkwan University
(Shaker@skku.edu)

Tamer Abuhmed
College of Computing and Informatics, Sungkyunkwan University
(tamer@skku.edu)

Sepsis remains a complex, life-threatening condition characterized by an overwhelming immune response to infection, leading to high mortality rates in hospital settings. Rapid and precise diagnosis is crucial to improving survival rates, but current practices lack personalized, predictive tools. The emergence of electronic health records has spurred the development of automated clinical decision systems, yet the efficacy of such tools can be significantly enhanced through the use of Machine Learning (ML) and Digital Twin (DT) technologies. This study bridges a vital gap in sepsis management by introducing a novel, layered ML model that incorporates DT technology to analyze time-series patient data within Intensive Care Units (ICUs). Leveraging a robust cohort from the MIMIC-IV dataset, we implemented and optimized an array of ML models, including classical, static ensemble, and dynamic ensemble, to predict sepsis outcomes. Our findings indicate the KNOP model, paired with classical ML classifiers, outperforms existing methodologies, offering a medically intuitive and trustworthy approach to sepsis prediction. This pioneering study is the first to apply DT technology and dynamic ensembles to ICU sepsis prediction, providing a foundation for future advancements in patient-specific healthcare.

Keywords : Sepsis Prediction, Digital Twin, Machine Learning, Intensive Care Unit, Time-Series Analysis

Received : March 26, 2024 Revised : April 16, 2024 Accepted : April 17, 2024
Publication Type : Fast Track Corresponding Author : Tamer Abuhmed

1. Introduction

Sepsis poses a critical threat to life as it results from an imbalanced host reaction to infection, leading to organ dysfunction (Singer et al., 2016). This worldwide health concern is acknowledged as

* This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ICT Creative Consilience Program (IITP-2021-2020-0-01821) supervised by the IITP (Institute for Information & communications Technology Planning & Evaluation), and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2021R1A2C1011198).

a significant contributor to inpatient mortality, with increased probabilities of cognitive decline and lasting impairment of organ function among those who survive (Rudd et al., 2020) and (El-Rashidy et al., 2022). Physicians face a significant challenge in diagnosing sepsis, given its multifactorial nature. Currently, there is a lack of personalized tools for sepsis treatment beyond general guidelines like the Surviving Sepsis Campaign, hindering real-time, patient-specific decision-making (Byrne & Van Haren, 2017) and (Marik, 2015). The escalating mortality risk is linked to each hour of delayed antimicrobial intervention, highlighting the urgent necessity for timely recognition and initiation of treatment (Ferrer et al., 2014) and (Pruinelli et al., 2018). Distinguishing sepsis from conditions with analogous clinical signs (e.g., inflammation) poses a significant challenge in early identification, given shared symptoms (e.g., fever) and molecular expressions (Lever & Mackenzie, 2007). Moreover, sepsis carries an elevated risk of cardiac arrest (Javan et al., 2019). Recognizing this, early diagnosis and swift treatment initiation emerge as pivotal elements in sepsis management. A wealth of studies highlights that prompt detection and treatment contribute significantly to reduced mortality rates and medical expenses (He et al., 2020). The pervasive adoption of electronic health records (EHRs) across healthcare facilities has enhanced the feasibility of automated systems for clinical decision-making and prediction, thereby facilitating improved monitoring and management of complex health syndromes. Such systems generate alerts and advisories by converting patient medical records into clinically actionable insights, aiming to elevate the standard of patient care (Sittig et al., 2008). In clinical environments, commonly applied rules-based scoring mechanisms for sepsis include the Systemic Inflammatory Response Syndrome (SIRS) criteria (Jaimes et al., 2003) and the Sequential Organ Failure Assessment (SOFA) score (Vincent et al., 1996). While these methodologies exhibit high sensitivity, their specificity is suboptimal, and they are not specifically engineered for predicting sepsis development. The limitations of these scoring systems or models are multifaceted: (1) they may yield inaccuracies for certain patient subgroups characterized by distinct risk factors at the time of admission, (2) they require regular revisions to incorporate medical advancements, (3) they depend on a limited set of initial characteristics recorded within the first 24 hours of ICU admission, neglecting the dynamic nature of a patient's clinical condition during their ICU stay, and (4) they often employ simplistic analytical models, such as logistic regression, which assume a linear and additive relationship between predictive variables and the outcome, potentially oversimplifying the complexity of clinical data (Javed et al., 2023) and (Ding et al., 2018). However, machine learning-based prediction (big time-series data) tools hold the potential to offer advanced notice of sepsis risk, increased specificity, and broader generalizability. This capability enables clinicians to intervene earlier while reducing the incidence of false alarms (Krishnan & Kamath, 2019). Limited research has focused on sepsis patients utilizing conventional machine learning, deep learning, or dynamic ensemble methods to enhance sepsis predictions based on raw time-series data (Kefi et al., 2019) and (Mansouri et al., 2020).

Accurate and timely prediction of sepsis is pivotal for early identification and intervention, which are critical for improving patient outcomes and survival rates. The urgency stems from sepsis's rapid progression to severe stages and septic shock within hours, significantly escalating mortality risks. Early diagnosis and treatment, particularly within the initial hours of symptom emergence, are vital for halting its advancement and reducing mortality. The concept of a 'golden hour' in sepsis treatment underscores the importance of timely interventions, with each hour's delay post-onset increasing mortality rates. Therefore, swift prediction and diagnosis are crucial for leveraging this critical intervention window to enhance survival prospects. Sepsis's dynamic nature demands real-time prediction and monitoring to adapt to rapid clinical changes, ensuring timely and appropriate care. Moreover, quick and accurate prediction models facilitate efficient healthcare resource allocation, prioritizing high-risk patients, especially in resource-constrained environments. Hence, in sepsis management, the speed of prediction is as crucial as accuracy. Models that provide real-time, actionable insights to inform immediate clinical decisions can markedly improve patient care, underscoring the necessity for technologies that balance promptness with precision in predicting sepsis.

Recently, the concept of DT has gained traction in various industries such as manufacturing and aerospace. DT adoption in healthcare, particularly for complex medical conditions like sepsis, is still in its infancy. Developing robust DT models necessitates comprehensive data integration, including patient-specific physiological data and clinical parameters, which can be challenging to acquire and standardize across healthcare systems. Additionally, the implementation of DT technology requires significant computational resources and expertise, posing practical barriers to its widespread application in clinical settings. Moreover, there is a lack of established protocols and regulatory frameworks governing the use of DT in healthcare, further hindering its adoption for sepsis detection. Overall, while the potential benefits of applying DT in sepsis detection are evident, the complexity and novelty of this approach contribute to its current underutilization in clinical practice.

Digital Twin technology enables real-time monitoring and simulation of assets, enhancing predictive maintenance and streamlining production. In automotive industries, DT expedites vehicle design and testing, ensuring quicker time-to-market and safer products. In healthcare, DT offers personalized diagnostics and treatment planning by integrating patient data, revolutionizing medical practice and improving patient care. Overall, Digital Twins offer versatile solutions to contemporary challenges, driving progress in diverse industries. DT technology promises to transform healthcare delivery and enhance patient outcomes. By creating virtual replicas of patients, it enables precise diagnostics, treatment planning, and monitoring. Integrating data from various sources like electronic health records and wearable devices, DTs offer clinicians a comprehensive view of each patient's health, improving accuracy and efficiency. This personalized approach enables early disease detection, tailored treatments, and proactive interventions. Additionally, DTs support medical research by providing virtual testbeds for exploring hypotheses and advancing precision medicine. Overall, DT technology represents a significant shift in healthcare, offering new avenues for improving patient care and driving innovation. In the following subsection we introduce DT in healthcare domain.

1.1. Digital twins in healthcare

Digital Twin technology, which generates digital representations of physical objects, has been widely adopted in multiple fields. In manufacturing, it refines design and maintenance (MIT Technology Review Insights, 2022). In the health sector, it is crucial for customizing patient care and advancing medical services (Venkatesh et al., 2022). In the medical field, DTs craft virtual counterparts of patient profiles, medical instruments, and entire healthcare infrastructures. These simulations are crucial for persistent, real-time surveillance and yield predictive insights that markedly enhance therapeutic techniques (Shengli, 2021). This progress is derived from the synthesis of varied data sets, including real-time physiological parameters and comprehensive historical medical data (Saracco, 2022). The incorporation of advanced technologies, including the Internet of Things (IoT), blockchain, and artificial intelligence (AI), into DTs markedly augments their functional potential.

DT technology offers a revolutionary approach in the Intensive Care Unit (ICU) for monitoring and predicting sepsis, serving as a dynamic, virtual replica that integrates real-time physiological data, patient history, and laboratory results into a unified model.

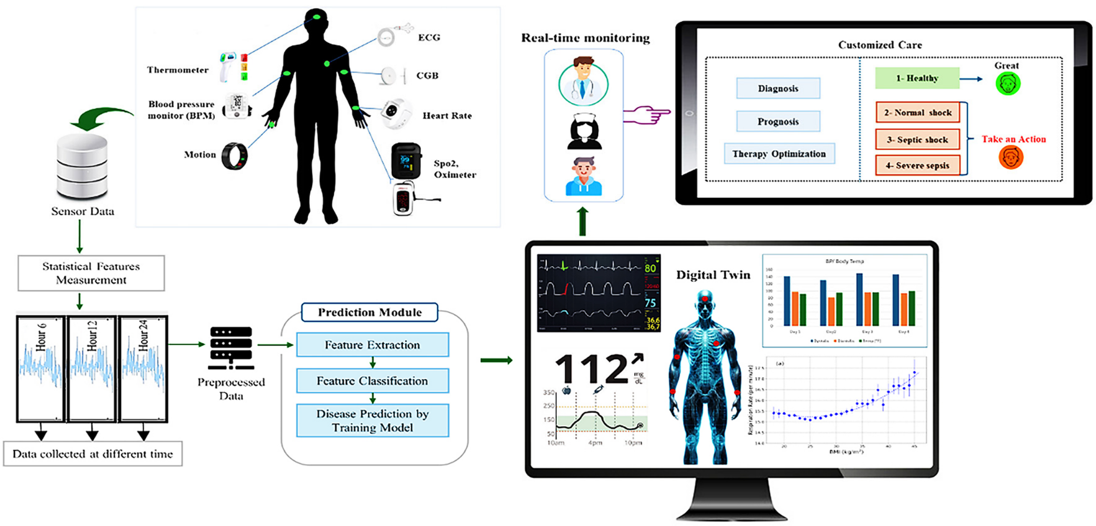
Key to the DT's efficacy is its interaction with dynamic ensemble learning models, which combine multiple predictive algorithms to improve accuracy. This system adapts over time, refining its predictions based on new data and clinical feedback, ensuring continuous improvement in sepsis detection. Data processing within the DT involves cleansing, normalizing, and extracting relevant features from patient data, which are then analyzed for sepsis indicators. Temporal analysis further enhances accuracy by evaluating the progression of physiological parameters, distinguishing between normal fluctuations and patterns signaling sepsis.

Machine learning algorithms harness historical data and patterns to forecast the future behavior or performance of assets depicted in the Digital Twin. This predictive capability facilitates the implementation of proactive maintenance and optimization strategies, thereby enhancing operational efficiency. Moreover, ML algorithms contribute to anomaly detection within Digital Twin models, pinpointing deviations from typical behavior or performance that may indicate potential issues or failures. Additionally, ML techniques enable regression analysis, empowering DTs to simulate and assess the repercussions of diverse variables or scenarios on system behavior. Overall, ML equips Digital Twins with the ability to leverage data-driven insights for predictive maintenance, anomaly detection, regression analysis, and other critical functions, thereby augmenting their efficacy across various domains. In the following subsection, we will review the importance and applications of ML in DT.

1.2. Digital twin with Machine learning

Fusion of DT technology with Machine Learning Algorithms (MLAs) has sparked significant advancements across various industries, including healthcare and diagnosis. This integration has been detailed in a collection of scholarly papers, which showcase applications ranging from fire management and petrochemical production to environmental monitoring, human-robot collaboration, and notably, the healthcare sector. The use of DTs and MLAs in healthcare have a great potential to revolutionize traditional diagnostic and treatment approaches, offering more personalized medical solutions.

(Zohdi, 2020) et al. discusses integrating MLAs with DT technology to simulate fire propagation. It combines a model tracking airborne embers, a model for combustible materials, and MLAs to optimize the simulation based on real-time data. This aims to create an adaptive, accurate DT for real-time use, particularly helpful for first responders. The integration enhances DT responsiveness and decision-making in fire management. (Min et al., 2019) et al. introduces a machine learning-enhanced DT framework for optimizing petrochemical production, utilizing IoT data for real-time adjustments to improve efficiency and economic performance. The approach overcomes challenges like data complexity and timing issues, with potential for broader manufacturing applications. Specifically, the framework incorporates Random Forest, AdaBoost, XGBoost, and LightGBM models, selected for their capabilities in handling the dynamic nature of petrochemical processes. These models train on industrial IoT and historical data to refine the DT model continually, enabling optimized production control based on real-time data updates. Klaus Dröder et al. (Dröder et al., 2018) introduces a novel DT and ML approach to enhance safety in human-robot collaboration (HRC). It details a system where robots dynamically adjust their paths around humans and obstacles using the HIRIT simulation platform. The method combines ML techniques like nearest neighbor path planning, clustering for obstacle segmentation, and ANNs for detection and 3D localization. A distinctive feature is a local path planner that creates a virtual safety envelope around humans, compliant with ISO/TS 15066 standards, through ML algorithms, notably ANNs. This allows for real-time adjustment of the robot's path, ensuring safety without hindering productivity. The research demonstrates the potential of ML to make industrial robots more adaptable, responsive, and intelligent, marking a significant advancement in HRC systems. R Ferdousi et al. (Ferdousi, 2021) introduces a DT framework enhanced with ML for disease diagnosis in human-robot collaboration. It features a dynamic MLA selection framework and testing method to predict diseases by analyzing heterogeneous health data. Three use cases including NCD risk, mental well-being, and COVID-19 risk predictions demonstrate the framework's efficacy, achieving high accuracies between 94.5% to 98% across various datasets. This approach highlights the potential of combining DT technology with ML to improve healthcare predictions, showcasing advancements towards more adaptable and intelligent systems. (Tancredi et al., 2022) et al. explores integrating DT technology with ML for anomaly



〈Figure 1〉 Advanced Patient Monitoring Through Digital Twin Technology.

detection in a food plant, enhancing safety and maintenance. It extends prior work by incorporating three ML algorithms into the DT setup: Multiple Linear Regression for predicting operational anomalies, an Artificial Neural Network (ANN) as a Multi-Layer Perceptron classifier for identifying plant statuses (normal, warning, failure), and K-Means Clustering for unsupervised classification of operational data into these status categories. This innovative approach demonstrates the potential of DTs combined with ML for real-time monitoring and predictive maintenance in industrial settings. (Scheuermann et al., 2020) et al. presents a Remote Health Monitoring (RHM) system that combines DT technology with machine learning, specifically using a Random Forest Regressor, to predict stress levels in extreme environments. Integrating smart textiles with sensors, it creates a DT that accurately mirrors human physiological responses and environmental conditions. This approach significantly enhances operational safety by providing real-time monitoring and stress prediction, showcasing the potential of DTs and ML in improving safety and performance for individuals in high-stress professions. These explorations into the synergy between DTs and ML across different fields underline a transformative shift towards more efficient, predictive, and adaptive systems. Figure 1 illustrates an integrated digital health monitoring system, which employs a network of medical devices, such as ECGs, thermometers, and blood pressure monitors, to gather sensor data and observe patient health metrics continuously. The collected data undergoes analysis to identify crucial statistical features, and through a prediction module, it facilitates feature extraction and classification. These processes contribute to a digital twin, a virtual representation of the patient's health, allowing doctors to visualize health metrics and trends like body temperature and blood pressure over time. By correlating data such as BMI with physiological parameters, the system enhances real-time monitoring and enables doctors to understand the patient's condition comprehensively, thus informing decisions for tailored patient care and delivering diagnostic and prognostic results more effectively.

In reviewing existing literature, we identified a gap in sepsis prediction research: the absence of dynamic ensemble models that leverage digital twin technology for real-time series data analysis in patients. While numerous studies have employed classical and ensemble machine learning algorithms for adult sepsis prediction (Wang et al., 2018) and (H.-F. Deng et al., 2022), none have integrated the innovative approach of digital twin technology. To address this, our study introduces a novel layered model that combines digital twin technology with the MIMIC dataset, marking a pioneering exploration into the joint application of these tools for enhanced sepsis prediction in patients. The study's contributions are summarized as follows:

1. Our study introduces a sepsis prediction framework for ICU patients using dynamic ensemble selection (DES) techniques, where we have innovatively preprocessed time-series data from the MIMIC-IV dataset. This preprocessing handle various time frames and instances of missing data, all within the context of digital twin technology which provides a dynamic and adaptive model of patient health.
2. We have rigorously evaluated an array of ML models in the classification layer. Our selection encompasses classical classifiers (i.e., decision tree, logistic regression, naive Bayes, k-nearest neighbors, multilayer perceptron), static ensemble models (i.e., random forest, AdaBoost, XGBoost, gradient boosting, LGBM), and Dynamic Ensemble Selection (DES) models (i.e., KNORAE, DESP, KNORAU, DESKNN, KNOP), all of which are integrated within the proposed digital twin framework to leverage its predictive capabilities.
3. We incorporate explainability tools within the digital twin technology. This enhancement allows physicians to gain insights into the decision-making process of the models, thus enabling a more transparent and effective implementation of ICU patient management systems.

The following sections of this paper are structured as follows: Section 2 offers a succinct review of related work. Section 3 outlines digital twin (DT) technologies and the dataset. Section 4 presents the proposed model, Section 5 explores the performance assessment, Section 6 discusses the experimental configuration, Section 7 covers the classification results, and Section 8 concludes the paper.

2. Related works

In the following section, we review the literature work on Sepsis prediction. In addition, we survey related works in ensemble learning that we have used in our proposed framework.

2.1. Sepsis prediction

Sepsis prediction is evolving with advancements in clinical models, biomarkers, and notably, Machine Learning (ML) and AI. While traditional models and biomarkers face limitations in specificity and interpretation, ML and AI offer more accurate, early predictions by analyzing extensive datasets. However, integrating these technologies into clinical practice and ensuring interpretability remain challenges. Predicting sepsis for individual ICU patients is notably more intricate, time-intensive, and costly, crucial for quick, effective patient care and ICU resource optimization. (Yong & Zhenzhou, 2024) et al. presented DGFSD, a new deep learning model for predicting sepsis mortality more accurately by analyzing clinical records and identifying essential risk indicators. Proven more effective than traditional approaches through tests on the MIMIC-III dataset, DGFSD represents a significant advancement in sepsis prognosis. (Rosnati & Fortuin, 2021) et al. introduced MGP-AttTCN, a machine learning model for early sepsis prediction that combines Multitask Gaussian Processes with an attention mechanism. It outperforms existing methods, achieving an AUROC of 0.660 and an AUPR of 0.483, which underscores its superior accuracy and interpretability for clinical use. (Rafiei et al., 2021) et al. SSP employs a deep learning approach for early sepsis detection in ICU patients, leveraging LSTM-CNN layers for analysis. It achieves high accuracy, with AUROCs of 0.89 and 0.92 on the PhysioNet/CinC 2019 dataset, demonstrating its potential to improve patient care (Nemati et al., 2018) et al. The study introduces the Artificial Intelligence Sepsis Expert (AISE) for early sepsis prediction in ICU patients by analyzing EMR and physiological data. Tested on over 83,000 ICU admissions, AISE demonstrated high predictive accuracy with AUROCs between 0.83 and 0.85, showing its potential to advance early sepsis detection and treatment strategies. (Mao et al., 2015) et al. The study utilized Insight, a machine learning model based on gradient tree boosting, analyzing six vital signs to predict sepsis, severe sepsis, and septic shock. Results showed high predictive accuracy, with AUROCs of 0.92 for sepsis, 0.87 for severe sepsis, and 0.96 for septic shock prediction four hours before onset, outperforming traditional scoring systems. (Guillén et al., 2015) et al. The research focuses on creating predictive models for severe sepsis in ICU patients using logistic regression, support vector machines, and logistic model trees, analyzing clinical data. The SVM model, combining lab values and vital signs, accurately predicted 65% of severe sepsis cases, aiming to improve early detection and intervention in ICU settings. (Kong et al., 2020) et al.'s research developed machine learning models to predict mortality among ICU sepsis patients using the MIMIC III dataset. Employing LR, LASSO, RF, and GBM methods, they compared their performance against the SAPS II model. Their findings indicated the GBM model had the best performance with an AUROC of 0.845, suggesting the potential of machine learning to improve sepsis outcome predictions in ICU settings. The study by (Camacho-Cogollo et al., 2022) et al. explores machine learning (ML) models for early sepsis detection using data from the

MIMIC-III database, focusing on vital signs, lab results, and demographics. The research tested models including SVM, KNN, ANN, Naïve Bayes, RF, AdaBoost, Stacking, and XGBoost. It found that XGBoost, in particular, outperformed traditional methods like SOFA and qSOFA in predicting sepsis onset, showcasing the efficacy of ML in improving sepsis detection in ICU environments. Determining the most efficient neural network structure involves a complex process that demands extensive resources and significant data. Furthermore, the transparency of deep learning (DL) models remains constrained, thus impeding clear insight into the elements driving their decision-making. Nonetheless, equivalent or even superior outcomes might be attainable through the use of simpler, quicker, and more dependable ensemble methods, which will be elaborated on in the following section.

2.2. Ensemble modeling

Enhancing model stability and performance can be achieved through the development of ensemble models, notably through methods such as boosting and random forests. Ensemble modeling represents a dynamic field within machine learning and pattern recognition, characterized by the integration of varied simple base models. This integration employs methodologies like bootstrapping and stacking, aiming to leverage the collective strengths of these base models for improved accuracy and reliability in predictions (Mansouri et al., 2020), (Kong et al., 2020), (El-Rashidy et al., 2022). Recent studies have underscored the advantages of using ensemble models over single base models, demonstrating their effectiveness through both theoretical foundations and practical evidence (Kuncheva, 2002) and (Polikar, 2006). (He et al., 2020) et al. introduced an early sepsis prediction model leveraging clinical electronic health record (EHR) data within an ensemble learning framework, merging artificial features derived from clinical knowledge (SIRS and SOFA criteria) with deep features from a long short-term memory (LSTM) neural network. Tested on the PhysioNet Challenge 2019 ICU patient data, the model showed a sensitivity of 0.641, specificity of 0.844, and a normalized utility score of 0.401 on public datasets. On a hidden test set, it achieved a utility score of 0.313, indicating its effectiveness in early sepsis detection and showcasing the advantage of ensemble learning in medical prediction tasks. (Rayan et al., 2021) et al. presented an ensemble machine learning model for early sepsis detection in ICU settings, achieving remarkable accuracies of 99% and 98% on different datasets. By combining various preprocessing steps and leveraging the Random Forest (RF) algorithm, it showcases significant potential for improving critical care. (Zabihi et al., 2019) et al. developed an ensemble of XGBoost models to predict sepsis early in ICU patients, using a novel feature set for missing clinical data. Utilizing a systematic approach for feature extraction and selection, they crafted an ensemble model that achieved significant success, achieving a third-place ranking in the 2019 PhysioNet/ Computing in Cardiology Challenge, attaining a utility score of 0.339 on previously unseen test data. This highlights the potential of ensemble models and innovative feature engineering in enhancing sepsis prediction accuracy. (El-Rashidy et al., 2020) et al. presented a stacking ensemble model for ICU mortality prediction, leveraging a dataset of 10,664 patients. This model, which incorporates several machine learning algorithms and expert-guided data modalities, achieved remarkable outcomes, including an accuracy of 94.4% and an AUROC of 93.3%. This performance surpasses existing methods, highlighting the model's efficacy in critical care settings.

Dynamic Ensemble Selection (DES) has emerged as a highly effective ensemble technique, characterized by its adaptive selection of base classifiers for each instance that requires classification. This approach begins by assessing the competence level of each classifier within a predefined pool. Subsequently, for each new instance, DES selects an ensemble of the most competent classifiers based on their expertise in specific local regions of the feature space, recognizing that no single classifier excels across all possible instances. The essence of DES lies in its ability to dynamically identify and deploy the most adept classifiers for any given instance, navigating the challenge through various DES strategies like K-nearest oracle elimination (KNORAE), meta-learning for DES (mETA-DES), Dynamic ensemble selection-K nearest neighbor (DES-KNN), and others, each tailored to optimize classifier selection in different scenarios. This dynamic selection process underscores the nuanced understanding that classifiers have varying domains of expertise within the feature space, aiming to harness this diversity for improved prediction accuracy (Britto Jr et al., 2014). Dynamic Ensemble Selection (DES) has been effectively employed in addressing a range of practical challenges, encompassing applications from sepsis prediction (Zabihi et al., 2019) and handling noisy data (Xiao et al., 2010), credit risk evaluation (Hou et al., 2020), time series forecasting (Yao et al., 2019), and multi-class classification tasks (Krawczyk et al., 2018). This method stands out for its ability to dynamically select an optimal subset of classifiers for each specific instance, enhancing prediction accuracy across various domains. Through its tailored approach, DES significantly contributes to advancements in fields requiring precise and reliable predictive models, demonstrating its versatility and value in real-world applications and this research investigates the feasibility of employing sophisticated methods for deciphering intricate ICU data, including multimodal and time-series information.

3. Materials and methods

3.1. Digital twins and patients care

The application of digital twins in delivering precision healthcare has demonstrated notable effectiveness, particularly in the management of elderly care. Utilizing cloud computing and IoT technologies (Y. Deng, 2022). The ClouDTH framework demonstrates the transformative application of cloud and IoT technologies in the realm of elder care, providing a model for real-time health monitoring and management. This system, through its innovative notification feature, leverages vital sign analysis to offer timely medication reminders, illustrating the practical benefits of digital twins in healthcare. Furthermore, its capability to detect early signs of pre-frailty highlights the framework's potential in preemptively addressing health concerns, thereby significantly enhancing healthcare outcomes for the elderly (Liu et al., 2019). The continuous collection of medical data throughout the year, as opposed to solely during intermittent clinic visits, enables digital twins to facilitate the early identification of alterations in a patient's health status. This approach allows physicians to respond with greater alacrity.

3.1.1 Digital coupling & data generation

Digital Coupling enables the exchange of operational data between a physical entity and its digital twin. Designing a digital twin with a targeted function or scenario, fed by authentic data, influences the coupling's characteristics, reflecting the ongoing dialogue regarding the standardization of connectivity between the two entities (Bowman et al., 2022). Within a digital twin, a comprehensive toolkit, comprising analytical, simulation, and visualization components addresses diverse operational demands. *Analytical* tools evaluate the physical entity's conduct and efficiency based on past and present data and anticipate future scenarios. *Simulation* tools enable the exploration of potential future states under varied conditions, while *visualization* components render these analytical and simulation insights for user evaluation, thus becoming a critical element of the twin's output. In essence, analysis within a digital twin systematically processes data, juxtaposing actual conditions against simulated predictions to support strategic adjustments and informed decision-making.

Utilizing digital twins to generate synthetic datasets significantly enhances the training of machine learning algorithms for disease diagnosis and prognosis. In particular, digital twins mimicking the cardiovascular system have been instrumental in producing synthetic Photoplethysmogram (PPG) data, thereby facilitating more effective model training (Mazumder et al., 2019). Moreover, digital twins have proven crucial in improving the precision of AI-driven image segmentation, notably elevating the quality of stem cell image segmentation, detection, and tracking beyond the capabilities of traditional imaging techniques through advanced deep learning methodologies (Du et al., 2022).

We explore the application of DTs to assist individuals in managing and potentially mitigating sepsis. By perpetually monitoring vital signs, immune responses, and other critical health metrics, a DT can be developed to mirror an individual's physiological state. Through a dedicated mobile application, users can receive tailored advice on lifestyle adjustments, medication adherence, and preventive measures. Additionally, individuals have the capability to log specific health-related events or interventions, receiving instantaneous feedback on the expected impact on their risk or progression of sepsis, as predicted by their digital twin. This innovative approach aims to enhance early detection and proactive management of sepsis.

3.2. Data set

MIMIC-IV is a comprehensive and freely accessible electronic health record dataset that includes detailed information about patient care at the Beth Israel Deaconess Medical Center, covering various aspects such as patient transfers, medication, laboratory values, and billing information. It offers extensive data for research, including deidentified patient demographics, hospital admission details, and ICU stay records. The dataset supports analyses over time with its anchor year system and provides a rich source for clinical and machine learning research (Johnson et al., 2020).

Tables 1 and 2 provide statistics on the time-series data for both healthy individuals and patients with sepsis.

3.2.1. Data inclusion criteria

Figure 2 illustrates data from the MIMIC-IV database, indicating that the study included 67,738 individuals. Among them, 5,798 were identified as sepsis patients, while 8,150 were selected as healthy individuals through a random hybrid sampling method, resulting in a total of 13,948 participants in

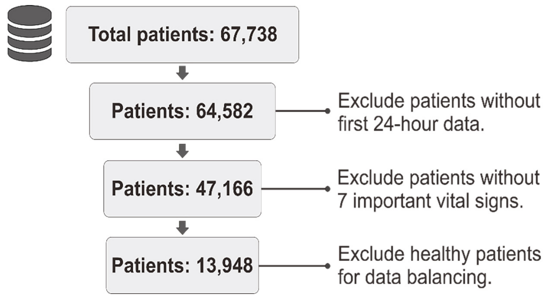
〈Table 1〉 Analytical data on time-series characteristics (Healthy N= 8,150 patients)

Feature Name	Mean	Minimum	Maximum	Standard Deviation
Heart Rate	84.75	76.59	94.20	6.39
Respiratory Rate	18.64	14.01	24.04	3.46
Temperature (°F)	97.54	96.63	98.69	0.92
BP Diastolic	67.36	64.51	73.57	5.19
BP Systolic	113.14	104.87	115.81	11.83
SpO2	97.33	97.80	99.96	1.55
Glucometer	87.29	82.12	122.11	11.16

〈Table 2〉 Analytical data on time-series characteristics (Sepsis N= 5,798 patients)

Feature Name	Mean	Minimum	Maximum	Standard Deviation
Heart Rate	94.43	85.99	117.16	6.59
Respiratory Rate	21.25	15.98	26.95	3.82
Temperature (°F)	98.05	96.85	100.75	3.22
BP Diastolic	61.76	48.95	93.17	13.82
BP Systolic	113.14	104.87	133.13	12.31
SpO2	94.24	88.00	95.18	2.13
Glucometer	150.93	115.12	233.13	6.09

the research. The MIMIC-IV database incorporates around 460 variables representing biological signals. For our analysis, we chose two data categories, specifically 3 single-valued and 7 time-series features, based on insights from scholarly research and recommendations from experts in the field (Huang et al., 2021). This study's inclusion criteria targeted ICU patients on their initial visit, aged over 18 years, prioritizing the first ICU admission in cases of multiple entries. Exclusion criteria ruled out individuals with more than one ICU stay, those younger than 18, and patients with duplicate hospital admission IDs. In the subsequent phase of selection, we removed records of patients with incomplete data or those who were outside the ICU. Our aim was to create sepsis prediction models tailored to ICU settings. Consequently, cases of patients developing sepsis after 24 hours or outside the ICU were deemed non-contributory to our study, as they fell outside the scope of our models' designed application environment. In cases of sepsis diagnosis, critical demographic and clinical information was compiled, encompassing age, gender, and principal vital parameters like heart rate, respiratory rate, blood pressure (both systolic and diastolic), body temperature, glucose levels, and pulse oximetry (SPO2). Furthermore, preliminary laboratory assessments were conducted within the initial 24 hours post-admission to the ICU. (Yang et al., 2022). We developed a model to predict sepsis outcomes based on data from the first 6, 12, and 24 hours.

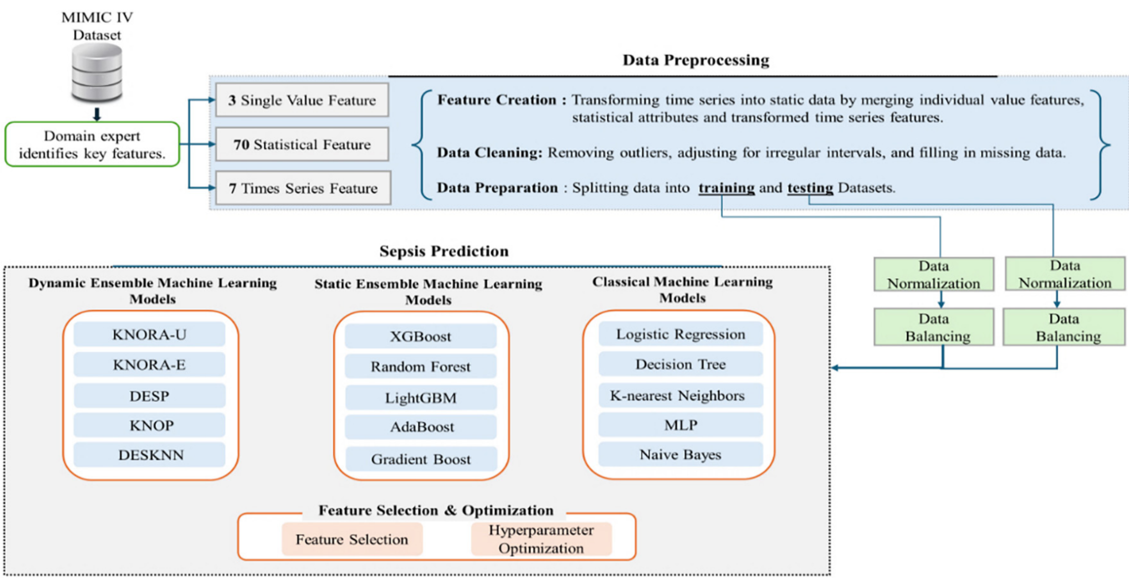


〈Figure 2〉 Inclusion criteria of patients.

4. Proposal framework

Figure 3 presents the model designed for predicting sepsis in ICU patients, focusing on accurate and interpretable outcomes. The model is structured around a key classification layer tasked with identifying sepsis cases within the ICU environment. We enhanced a suite of machine learning (ML) models, encompassing classical ML, static ensemble, and dynamic ensemble models. Post-modeling, feature selection and hyperparameter optimization are applied to enhance model performance. Their comparative effectiveness was assessed through a rigorously selected statistical test to identify any significant performance disparities among them.

The optimal model is chosen based on superior statistical performance. Machine learning specialists assess the reasoning behind its decisions, ensuring the model relies on the correct features. This allows for evaluation of the model's stability and prediction of potential inaccuracies. Additionally, a model's decision-making transparency bolsters a domain expert's confidence in its predictions. The



〈Figure 3〉 Schematic of a machine learning-driven framework for predicting sepsis in the ICU, from feature selection through model application to optimization, using the MIMIC-IV dataset.

ensuing subsections will delve into each stage of the developed model's process.

4.1. Data preparation

This section elaborates on the methodology for refining raw data into an analysis-ready state, addressing inconsistencies, and filling gaps to improve the integrity of the data, which is crucial for the accuracy of ML and statistical methodologies. In managing data irregularities, which often influence ML algorithms, the records in question are not discarded. Instead, these irregularities are initially indicated as missing via null value assignment, followed by a reconstitution of the affected entries through sophisticated imputation methods.

4.1.1. Feature extraction

Feature generation involves creating new features from existing data to unveil novel combinations and representations that may enhance ML model performance. By extracting statistical features from time-series data, a comprehensive summary of the underlying knowledge within these sequences is obtained, facilitating a deeper understanding and improved predictive capabilities (Sphinx, 2024) and (El-Sappagh et al., 2020). By focusing on seven time-series variables including heart rate, respiratory rate, systolic and diastolic blood pressure, body temperature, glucose, and pulse oxygen saturation (SPO2), we extracted essential statistical metrics like mean, minimum, maximum, variance, skewness, kurtosis, quartiles, and standard deviation. These measures offer a detailed statistical view of each feature, grounded in the extensive patient data captured by the MIMIC-IV database's 26 tables, including demographics, vital signs, and diagnostic information. Specifically, our study utilizes data from the chartEvents, outputEvents, labEvents and lab_items tables for a comprehensive analysis. The integration of data from the MetaVision and CareVue medical information systems was facilitated by employing the inputevents_cv and inputevents_mv tables. This approach ensures a rich summary of patient conditions, treatments, and outcomes, leveraging the chartEvents, ICUstay, and inputevents tables to monitor vital signs, ICU admissions, and medication administration, respectively.

4.1.2. Data division and normalization

Our dataset, significantly larger than those commonly cited in the literature, undergoes analysis through standard ML models. We meticulously divide the dataset into training (80%) and testing (20%) portions to preclude information leakage, segregating the testing set at an early phase. Upon dataset segmentation, we utilize the Min-Max normalization technique, described by the formula $Value = (x - min(x))/(max(x) - min(x))$, to achieve data standardization to a zero mean and unit standard deviation, a pivotal measure to avert data leakage and inflated outcomes.

A notable challenge with applying Min-Max normalization in time series forecasting arises from the indeterminate minimum and maximum values in out-of-sample data. To address this, we adopt the minimum (min) and maximum (max) from the in-sample dataset, mapping any out-of-sample values falling below min or above max to predetermined low and high thresholds, respectively. This approach ensures consistency in data treatment across in-sample and out-of-sample datasets, safeguarding the integrity of our predictive analysis (Al Shalabi & Shaaban, 2006) and (Sola & Sevilla, 1997).

4.1.3. Data balancing

Given the significant imbalance in our dataset, we initially employed SMOTE Tomek [60 and 61] to address this issue. We also considered under-sampling, a method that equalizes class distribution by decreasing the size of the more prevalent class, though it may lead to substantial data loss. Consequently, to preserve information integrity, we opted for oversampling techniques, applying them distinctly within training and testing sets (Megahed et al., 2021). These techniques enhance dataset equilibrium by augmenting the minority class with additional instances, generated through methods like duplication, bootstrapping, or creating synthetic samples, thereby ensuring a more balanced data representation for analysis.

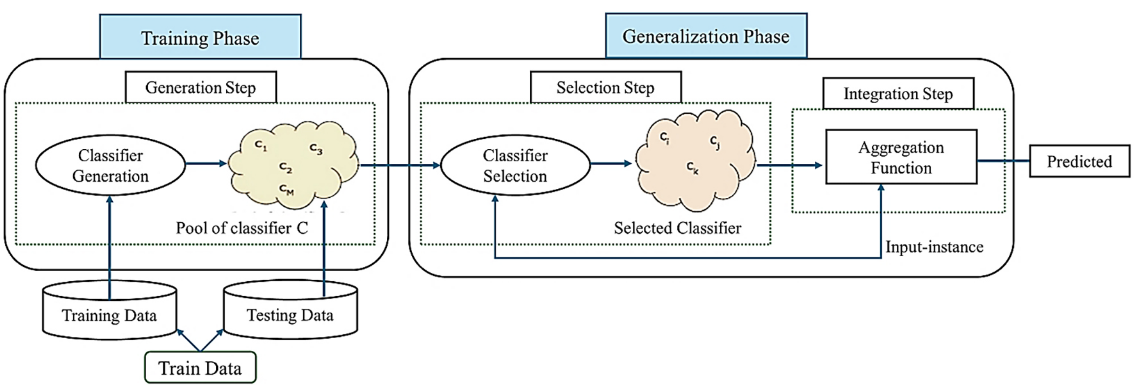
4.2. Classification layer

At this phase, we assess the efficacy of several established ML techniques, including logistic regression (LR), k-nearest neighbors (KNN), naive Bayes (NB), decision tree (DT), and multilayer perceptron (MLP). We also evaluate static ensemble methods such as LightGBM, random forest (RF), XGBoost, AdaBoost, and gradient boosting (GB) classifiers.

Additionally, cutting-edge dynamic ensemble selection (DES) approaches like KNORA-U, KNORA-E, DESP, DESKNN, and KNOP are examined for performance comparison (Ren et al., 2016) and (Cruz et al., 2018). The utilization of traditional ML techniques on the MIMIC dataset has demonstrated efficacy in addressing numerous challenges within intensive care units. Moreover, it has been observed that ensemble methods surpass the performance of their individual component algorithms (Juraev et al., 2022). The development of an ensemble model encompasses three critical phases: firstly, the generation phase, where base classifiers are developed and trained; secondly, the selection phase, which involves either static or dynamic determination of the optimal mix of classifiers for the ensemble; and thirdly, the integration phase, where the decisions from the base classifiers are amalgamated using various strategies, as depicted in Figure 4.

The aggregation of machine learning models, recognized for their improved accuracy, has prompted the development of numerous static ensemble models within academic circles. By leveraging our dataset, these models are specifically tailored to tackle diverse medical challenges encountered in intensive care settings (Awad et al., 2017) and (García et al., 2016).

Dynamic ensemble selection leverages real-time assessment to choose the most adept base classifiers from a pool, tailored for each new data instance. This approach involves estimating each classifier's competence level, thereby utilizing the most proficient classifiers to determine the label of the test instance. The capability of individual base classifiers is commonly evaluated within a localized vicinity of the feature space pertinent to the query sample. This specified area can be delineated using a variety of approaches, such as support vector machine decision boundaries or decision trees, considering that each classifier possesses specialized knowledge in certain areas of the feature space. In the current research, the authors enhanced a variety of fundamental classifiers, including neural networks, static ensembles like random forests, and Dynamic Ensemble Selection (DES) models by utilizing optimization techniques like cross-validation. The DES models demonstrated superior performance compared to other methodologies.



〈Figure 4〉 Dynamic ensemble selection (DES).

Furthermore, this research proposes a robust and medically significant process for data preparation. In this study, an array of machine learning models underwent extensive evaluation using standard and grid search-optimized hyperparameters. The models were calibrated using the complete feature set to fine-tune their performance. It was found that selecting a subset of particularly informative features significantly improved the models' effectiveness.

5. Performance assessment

In biomedical applications, classification models are evaluated with metrics like accuracy, precision, recall, and F1-score due to the critical nature of medical diagnoses and treatments. These metrics are selected due to their ability to offer a thorough evaluation of a model's performance, addressing the need for both correctness and reliability in predictions. In such settings, it's crucial not only to accurately identify positive cases (like diseases) but also to minimize false positives and false negatives, which could lead to incorrect treatments or missed diagnoses. These metrics collectively ensure that models achieve a balance between identifying true cases and avoiding misclassification, directly impacting patient care and outcomes (Behera & Kumaravelan, 2020) and (Barzekar & Yu, 2022).

In this study, the DT serves as a sophisticated simulation model designed to predict how critically ill patients with sepsis respond to specific treatments within the first 24 hours of care. It operates by mimicking the complex interactions between various organ systems in response to treatments, using a causal AI framework that incorporates directed acyclic graphs, agent-based modeling, and Bayesian networks. This virtual representation allows for real-time prediction of patient outcomes, facilitating personalized care planning. By comparing the digital twin's predictions with actual patient responses, the study verifies the model's accuracy and its potential utility in enhancing decision-making in critical care environments.

6. Experimental configuration

In our research, we undertook a series of experiments focused on identifying the optimal classifier for sepsis prediction. Our decision to choose the intervals (0h-6h, 0h-12h, 0h-24h) was based on a thorough consideration of the clinical progression of sepsis which depends on the first week of the diagnosis (León et al., 2013). According to the Surviving Sepsis Campaign (SSC) guidelines, there is a crucial emphasis on the "golden hours" concept, highlighting the significance of interventions within the first 6 hours after sepsis recognition, which significantly affects patient outcomes (Rhodes et al., 2017). Thus, our inclusion of the earliest 0h-6h interval in our analysis is aimed at capturing the window where early diagnostic markers and interventions can profoundly influence outcomes. Extending the analysis to 0h-12h and 0h-24h intervals is also based on the medical recommendation of the Sepsis diagnosis in the first 24 hours as Sepsis can develop quickly from initial infection and progress to septic shock in as little as 12 to 24 hours (Contou et al., 2016) and (Kumar et al., 2006).

These extended windows provide insights into the dynamic nature of sepsis, where early interventions may not always result in accurate diagnosis as the diagnosis by the 12 and 24 hours.

In addition, our inclusion of these time intervals facilitates a comprehensive analysis that represents the variability in clinical practice and data availability in ICU settings. The MIMIC-IV dataset, which we use in our analysis, encompasses a wide range of clinical observations, interventions, and outcomes, providing an opportunity to model sepsis prediction across different stages of the condition. Through evaluating the predictive performance of our model across these intervals, we aim to identify the most critical features and time frames for sepsis prediction, thus refining our approach to harness the most informative data points for early identification.

Moreover, analyzing multiple time windows allows us to address the inherent variability in sepsis presentation and progression. It recognizes that sepsis markers may not be uniformly present or detectable across all patients within the same time frame. This accommodation reflects the clinical reality that sepsis is a heterogeneous syndrome with variable manifestations (Seymour et al., 2016). This multi-interval analysis enhances our ability to capture the complex and diverse nature of sepsis, thereby contributing to a more comprehensive understanding of its progression and the context for accurate predictive modeling.

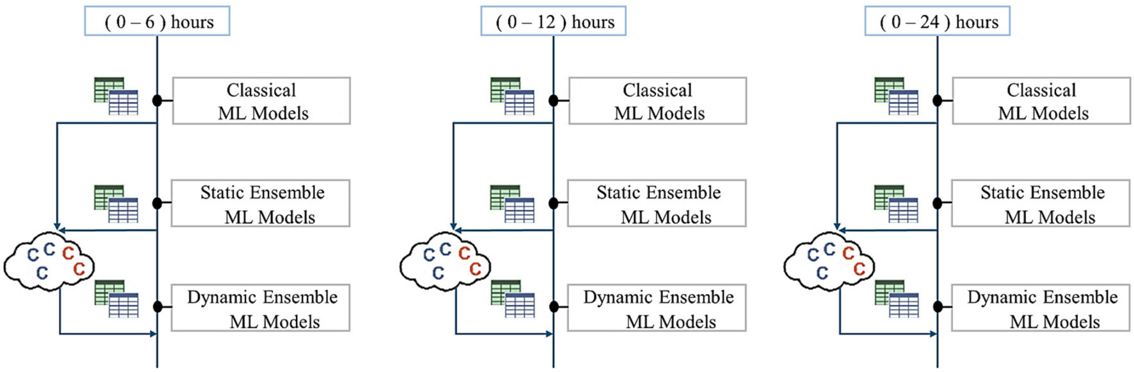
Furthermore, our methodology involved evaluating models through a comprehensive approach that included testing with default hyperparameters, optimizing feature selection and hyperparameters, and determining the ideal count of base classifiers for both dynamic and static ensemble models. The selection process for these models was grounded primarily in a statistical analysis of their performance metrics. Figure 5 depicts our stratified classification layer, designed for the prediction of sepsis at varying time intervals. This layered approach is tailored to capitalize on the temporal granularity of clinical data, ranging from 0 to 6, 0 to 12, and 0 to 24 hours. Each interval employs a unique set of ML models that progressively build upon the data accumulated over time.

Classical ML Models: At the foundational level, individual classifiers are employed within each time window. These models utilize established ML algorithms to analyze the data and produce baseline predictions concerning sepsis risk.

Static Ensemble ML Models: Advancing to the second tier, we integrate a static ensemble of the initial classifiers. Here, a predetermined method, such as weighted averaging or majority vote, combines the individual predictions. This static combination is invariant to new data inputs and relies on the robustness of aggregated judgments.

Dynamic Ensemble ML Models: The third tier introduces a dynamic aspect, wherein the ensemble adjusts the influence of individual classifiers in real-time, based on the incoming data stream. This flexibility is particularly crucial in adapting to the rapidly changing clinical markers indicative of sepsis progression.

Each model within the tiers is meticulously evaluated against performance metrics to ascertain



〈Figure 5〉 Experiments design with different patient's medical sensory data available for Sepsis prediction.

its efficacy in the context of sepsis prediction. This evaluation is crucial for understanding how early or late data acquisition impacts predictive accuracy. The juxtaposition of static and dynamic ensembles further allows us to discern the most effective strategies across the temporal spectrum of data availability, aiming to optimize the prompt and precise forecasting of sepsis and potentially improve clinical outcomes.

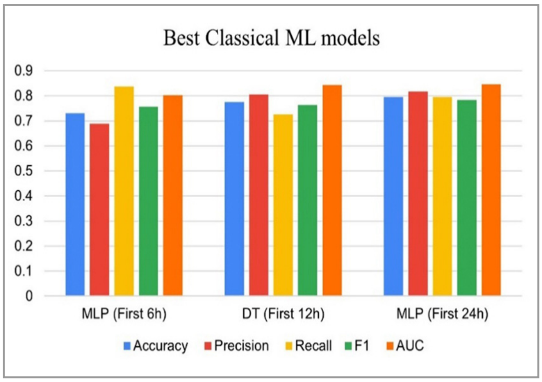
7. Results of classification

This section delineates the outcomes associated with the classification layer, articulated through three distinct subsections. Initially, the performance of classical machine learning models is examined. Subsequently, the efficacy of both static and dynamic ensemble models is scrutinized. To ensure the robustness of our findings, a 5-holdout testing technique was employed, with results reported as the mean and standard deviation. The discussion predominantly employs the F1-Score as the central performance metric due to its representativeness of broader metrics. For an extensive analysis of additional performance measures, readers are directed to consult the relevant tables provided.

7.1. Classical ML models

This investigation systematically assesses classical ML models by applying feature selection and hyperparameter optimization methodologies. The aim of this segment is to critically analyze the performance of each classical ML model across different time series datasets, ultimately determining the superior model within each dataset.

In the analysis of classical ML models over time intervals of 0 to 6, 0 to 12, and 0 to 24 hours. Table 3 reveals that the multilayer perceptron (MLP) model exhibits superior accuracy within the initial 6-hour period. Conversely, for the subsequent intervals of 12 and 24 hours, the Decision Tree (DT) and MLP models achieve the highest accuracy, respectively. These outcomes are graphically depicted in Figure 6 depicts the comparative efficacy of the models over the designated time intervals.



〈Figure 6〉 Best classical ML model performance across various patients' data intervals

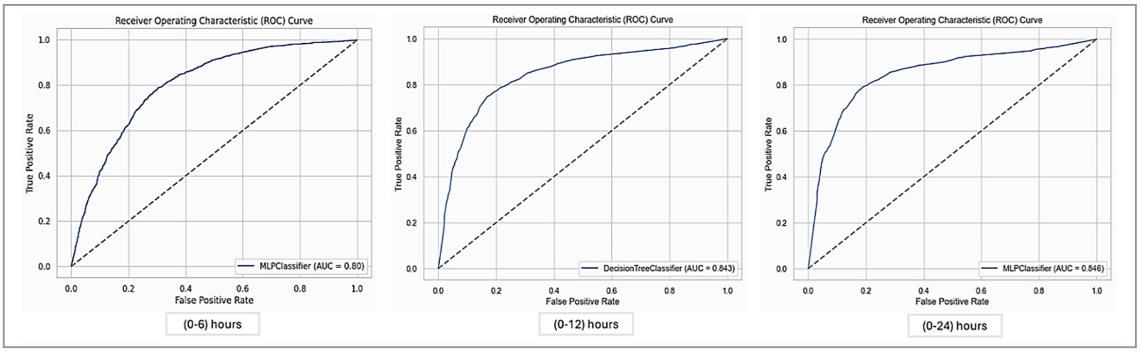
Figure 7 illustrates the ROC curves that display the performance of machine learning classifiers across three distinct time intervals. The initial interval is the shortest, with subsequent intervals extending up to a full day. The corresponding AUC values indicate that model accuracy improves as the duration of the time interval increases, signifying enhanced predictive capabilities over extended periods. Notably, the Decision Tree Classifier excels in the mid-interval, while the MLP Classifier exhibits a slight advantage in the longest interval. These curves highlight the

〈Table 3〉 ML performance for Sepsis prediction over various patients' data intervals.

Time (0 to 6) hours					
Classical ML models					
Models	Accuracy	Precision	Recall	F1	AUC
LR	0.725+0.012	0.724+0.012	0.725+0.015	0.725+0.012	0.790+0.017
DT	0.722+0.011	0.722+0.017	0.722+0.023	0.722+0.010	0.784+0.010
KNN	0.730+0.005	0.719+0.005	0.757+0.028	0.737+0.011	0.788+0.009
MLP	0.730+0.018	0.689+0.003	0.838+0.036	0.756+0.019	0.803+0.022
NB	0.672+0.010	0.671+0.003	0.676+0.021	0.674+0.011	0.672+0.012

Time (0 to 12) hours					
Classical ML models					
Models	Accuracy	Precision	Recall	F1	AUC
LR	0.722+0.003	0.721+0.007	0.724+0.011	0.723+0.004	0.794+0.003
DT	0.775+0.005	0.805+0.008	0.726+0.010	0.764+0.005	0.843+0.006
KNN	0.747+0.006	0.778+0.004	0.691+0.018	0.732+0.009	0.833+0.005
MLP	0.751+0.019	0.756+0.023	0.742+0.017	0.749+0.02	0.832+0.016
NB	0.723+0.025	0.73+0.022	0.707+0.027	0.718+0.023	0.786+0.019

Time (0 to 24) hours					
Classical ML models					
Models	Accuracy	Precision	Recall	F1	AUC
LR	0.723+0.008	0.717+0.006	0.737+0.021	0.727+0.011	0.789+0.009
DT	0.784+0.003	0.791+0.007	0.773+0.014	0.782+0.006	0.840+0.005
KNN	0.748+0.005	0.779+0.007	0.693+0.002	0.733+0.004	0.833+0.004
MLP	0.796±0.016	0.817±0.017	0.796±0.029	0.784±0.015	0.846±0.012
NB	0.739+0.019	0.752+0.015	0.713+0.022	0.732+0.027	0.810+0.024

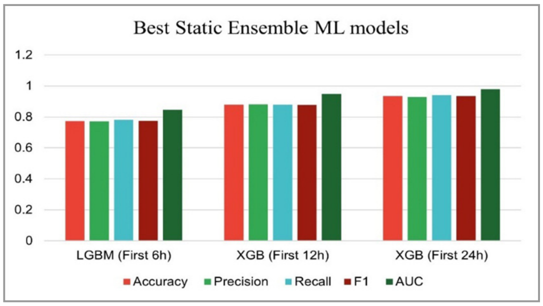


〈Figure 7〉 AUC scores of the best classical ML models enhanced by feature selection and hyperparameter optimization across various data intervals.

classifiers' ability to balance true and false positives, demonstrating their increasing diagnostic accuracy with time.

7.2. Static ensemble ML models

This section details the optimization of five prominent static ensemble ML models: Random Forest (RF), Gradient Boosting (GB), AdaBoost, Light Gradient Boosting Machine (LGBM), and Extreme Gradient Boosting (XGBoost). Each model undergoes a thorough evaluation involving feature selection and hyperparameter tuning. The objective is to appraise each model's efficacy on different time series datasets and identify the most accurate model for each observational period. Analysis presented in Table 4 indicates that LGBM outperforms others in the early window of 0 to 6 hours, while XGBoost models show the greatest accuracy for 12 and 24-hour intervals. Figure 8 visualizes this comparative analysis, showcasing model performance over the delineated time frames.



〈Figure 8〉 Best Static Ensemble ML model across various data intervals.

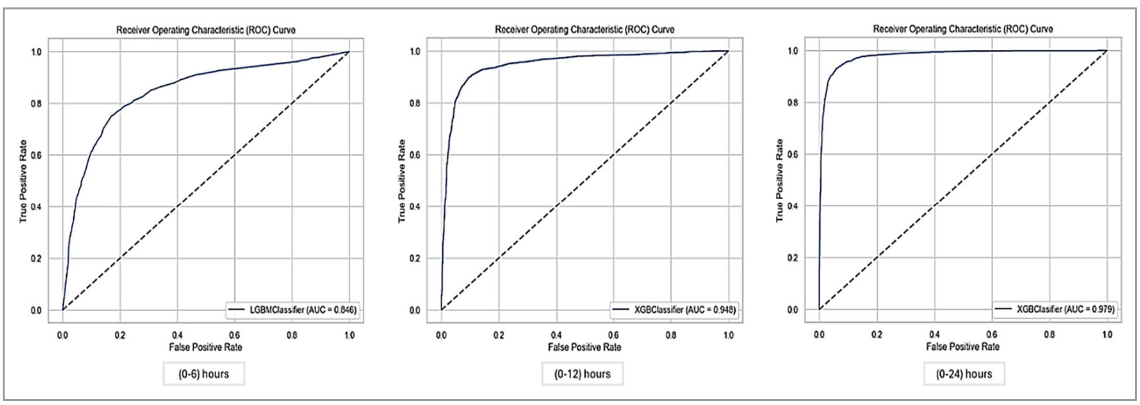
Figure 9 illustrates the ROC curves which depict the predictive performance of the LGBM and XGBoost classifiers over increasing observation windows. Initially, the LGBMClassifier shows solid predictive accuracy for the early window. As the timeframe extends, the XGBoost classifier demonstrates substantial improvements in its ability to classify, as indicated by rising AUC values. This improvement suggests the model's enhanced precision in distinguishing the positive class with longer observation periods. These AUC trends underscore the model's capacity for early detection of conditions, highlighting robust performance across the analyzed intervals.

〈Table 4〉 Static ML performance for Sepsis prediction over various patients' data intervals.

Time (0 to 6) hours					
Static Ensemble ML models					
Models	Accuracy	Precision	Recall	F1	AUC
RF	0.762+0.006	0.755+0.011	0.777+0.015	0.766+0.005	0.842+0.008
XGB	0.765+0.012	0.760+0.010	0.775+0.022	0.767+0.014	0.841+0.013
LGBM	0.774+0.007	0.771+0.005	0.781+0.022	0.776+0.010	0.846+0.011
AdaBoost	0.760+0.015	0.762+0.004	0.756+0.030	0.759+0.016	0.832+0.013
GB	0.745+0.011	0.740+0.006	0.755+0.023	0.747+0.012	0.816+0.010

Time (0 to 12) hours					
Static Ensemble ML models					
Models	Accuracy	Precision	Recall	F1	AUC
RF	0.871+0.004	0.877+0.003	0.863+0.015	0.870+0.006	0.940+0.006
XGB	0.879±0.009	0.881±0.008	0.879±0.010	0.878±0.011	0.948±0.008
LGBM	0.872+0.008	0.867+0.009	0.877+0.013	0.872+0.008	0.943+0.005
AdaBoost	0.860+0.013	0.858+0.018	0.863+0.015	0.861+0.017	0.932+0.012
GB	0.861+0.009	0.858+0.011	0.865+0.008	0.862+0.012	0.933+0.010

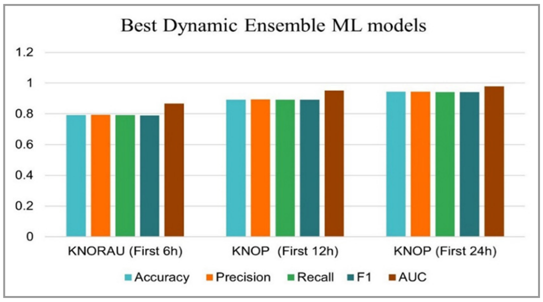
Time (0 to 24) hours					
Static Ensemble ML models					
Models	Accuracy	Precision	Recall	F1	AUC
RF	0.934+0.008	0.929+0.010	0.941+0.009	0.935+0.007	0.981+0.008
XGB	0.935+0.008	0.929+0.007	0.941+0.010	0.935+0.007	0.979+0.006
LGBM	0.903+0.007	0.895+0.011	0.913+0.005	0.904+0.006	0.961+0.007
AdaBoost	0.895+0.011	0.887+0.009	0.905+0.008	0.896+0.012	0.952+0.009
GB	0.905+0.009	0.901+0.015	0.911+0.012	0.906+0.010	0.956+0.008



〈Figure 9〉 AUC scores of the best static ensemble ML classifiers enhanced by feature selection and hyperparameter optimization across various data intervals.

7.3. Dynamic ensemble ML models

This section presents the refinement of five notable dynamic ensemble ML models: KNORA-U, KNORA-E, DESP, KNOP, and DESKNN. We meticulously assess each model, applying feature selection and fine-tuning hyperparameters to gauge their effectiveness across various time series datasets. Our aim is to determine the preeminent model for specific observation windows. As evidenced in Table 5, KNORA-U achieves superior performance within the initial (0 to 6) hour timeframe, whereas KNOP models excel in predicting over 12 and 24-hour durations. Figure 10 offers a graphical representation of this comparative study, delineating model efficacy across these timeframes.



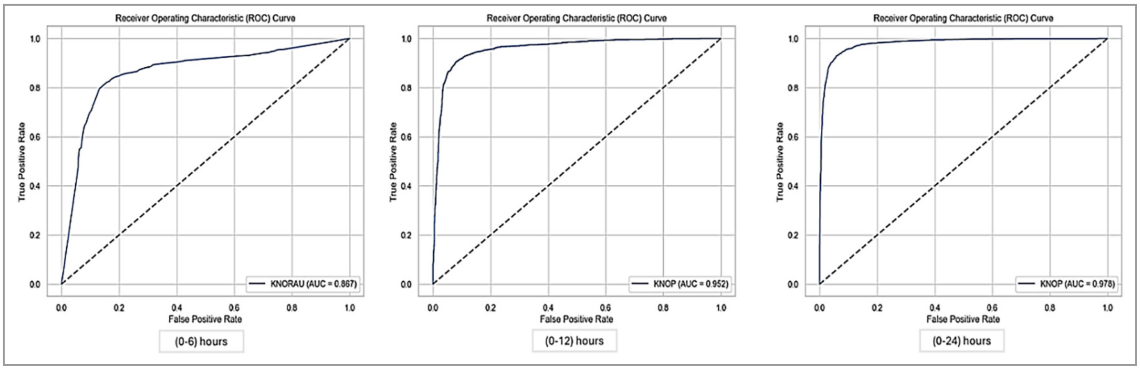
〈Figure 10〉 Best dynamic Ensemble ML model across various data intervals.

〈Table 5〉 Dynamic ML performance for Sepsis prediction over various patients' data intervals.

Time (0 to 6) hours					
Dynamic Ensemble ML models					
Models	Accuracy	Precision	Recall	F1	AUC
KNORAU	0.792±0.003	0.794±0.003	0.792±0.004	0.791±0.003	0.867±0.005
KNORAE	0.787±0.002	0.789±0.002	0.787±0.002	0.784±0.002	0.861±0.003
KNOP	0.782±0.003	0.783±0.004	0.782±0.003	0.781±0.003	0.853±0.004
DESKNN	0.776+0.017	0.767+0.019	0.794+0.017	0.780+0.018	0.849+0.016
DESP	0.775+0.011	0.767+0.013	0.791+0.014	0.779+0.012	0.849+0.010

Time (0 to 12) hours					
Dynamic Ensemble ML models					
Models	Accuracy	Precision	Recall	F1	AUC
KNORAU	0.891+0.005	0.886+0.004	0.897+0.006	0.891+0.005	0.950+0.003
KNORAE	0.888±0.002	0.890±0.002	88.84±0.002	88.84±0.002	0.948±0.001
KNOP	0.892±0.002	0.894±0.003	0.892±0.002	0.892±0.002	0.952±0.002
DESKNN	0.880±0.005	0.883±0.003	0.880±0.007	0.880±0.004	0.939±0.005
DESP	0.891+0.006	0.889+0.007	0.893+0.011	0.891+0.006	0.954+0.008

Time (0 to 24) hours					
Dynamic Ensemble ML models					
Models	Accuracy	Precision	Recall	F1	AUC
KNORAU	0.941+0.002	0.946+0.003	0.935+0.002	0.941+0.002	0.978+0.002
KNORAE	0.940+0.004	0.944+0.003	0.936+0.003	0.940+0.005	0.972+0.003
KNOP	0.944+0.008	0.945+0.012	0.942+0.010	0.943+0.008	0.978+0.007
DESKNN	0.935+0.013	0.93+0.009	0.941+0.014	0.936+0.011	0.979+0.008
DESP	0.939+0.004	0.941+0.005	0.938+0.006	0.939+0.004	0.974+0.003



〈Figure 11〉 AUC scores for optimized best dynamic ensemble ML classifiers across different intervals periods.

Figure 11 showcases ROC curves for the dynamic ensemble ML models KNORA-U and KNOP across various observation periods. The model's ability to distinguish between classes is measured using the AUC metric. In the initial monitoring phase, KNORA-U demonstrates strong predictive accuracy, indicated by a high AUC score. As the observation period expands, KNOP's accuracy shows notable enhancement, with higher AUC scores during extended monitoring phases. This increase suggests that KNOP's capacity for accurate predictions improves with a broader data timeframe, highlighting its increasing precision in classifying the positive class over time.

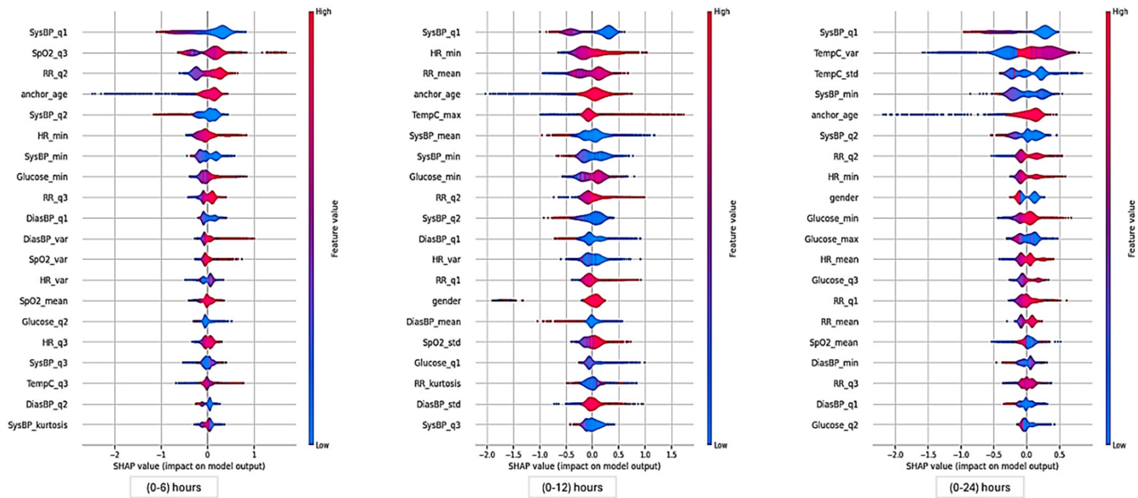
7.4. Feature impact analysis with SHAP values

SHAP (SHapley Additive exPlanations) values are a technique from game theory that explain the output of machine learning models. They assess the influence of each feature on the model's prediction, offering insight into the effects of various features on the model's outcome.

Figure 12 displays SHAP graphs for intervals of 0 to 6, 0 to 12, and 0 to 24 hours, illustrating the dynamic impact of features over time. High SHAP values, such as those for SysBP_q1 in the initial interval, are crucial to the model's predictive accuracy. As the observation window widens, the significance of certain features evolves, with some, like TempC_var, becoming more influential in extended intervals.

The breadth of SHAP value distributions across the graphs varies, reflecting the change in certainty regarding each feature's influence. These visualizations facilitate the identification of features that consistently drive predictions and those whose importance increases with the accumulation of temporal data.

The main advantage of dynamic ensemble selection models over both static ensemble models and traditional machine learning (ML) models lies in their ability to leverage the strengths of multiple diverse models for making a final prediction. These dynamic models draw upon the capabilities of both static ensembles and traditional ML models, using them as a resource pool for decision-making. This approach allows for more nuanced and adaptable predictions. Additionally, the KNOP model, a type



〈Figure 12〉 SHAP value distributions for model features, indicating variability in predictive impact across different observation periods.

of dynamic ensemble selection, considers a wide range of features with equal importance. This balanced consideration helps it to understand the problem more comprehensively, leading to better generalization across different situations. In contrast, static ensembles and traditional ML models tend to focus on a limited set of features. This narrow focus can result in overfitting, where the model performs well on the training data but poorly on new, unseen data. Consequently, these models often show less generalization ability when applied to test datasets.

8. Conclusion

In this research, we pioneered a stratified machine learning (ML) model that harnesses the synergies of Digital Twin (DT) technology and Dynamic Ensemble Learning (DEL) to predict sepsis in Intensive Care Units (ICUs), marking a distinctive approach in the utilization of time-series patient data across three critical intervals: 0 to 6, 0 to 12, and 0 to 24 hours. Employing a comprehensive evaluation framework, our methodology entailed a rigorous assessment of a wide spectrum of ML models, including classical, static ensemble, and dynamic ensemble types, optimized through hyperparameters and feature selection. This optimization, leveraging the extensive MIMIC-IV dataset comprising a patient cohort of 13,948, showcased the KNOP model in tandem with classical ML classifiers as significantly superior in the sepsis prediction task. It achieved an unparalleled accuracy of 0.944 ± 0.008, precision of 0.945 ± 0.012, recall of 0.942 ± 0.010, and F1-score of 0.943 ± 0.008, thus rendering it a more medically insightful and reliable solution than other models documented in the literature. The objective of our study was to fill the existing gap in personalized, real-time sepsis prediction tools by introducing a sophisticated model that merges the predictive strength of DEL with the real-time monitoring and simulation capabilities of DTs. Through an exhaustive process of data preprocessing, feature selection, and the refinement of various machine learning models, our study explored the collaborative potential of DT and DEL in advancing sepsis prediction methodologies. The findings from our investigation not only underscored the enhanced predictive accuracy and reliability of the KNOP model across all examined time windows but also emphasized the advantage of integrating DT with DEL, thus providing a robust framework for immediate, patient-centric sepsis prediction. From an academic perspective, this research blazes a new trail in the application of DT and DEL within the sepsis prediction domain, potentially revolutionizing ICU patient care. However, the reliance on data from a single institution, as contained in the MIMIC-IV dataset, and the computational demands of implementing DT and DEL in real-world clinical scenarios are challenges that warrant further exploration. Moreover, our model exploited a large set of statistical features that can be optimized to a short list of significantly important features. In future studies, investigating the role of each feature on the model performance supported by statistical analysis tests to narrow down the features used in the prediction model. This leads to the implementation of more accurate and cost-effective models. The study also did not consider the integration of other medically relevant patient’s data for Sepsis diagnosis that can be explored in future studies. As for the predictive models of the study, all models used aggregated statistics of the patient time series sensory data and avoided using the raw time-series data with more convenient DL models like LSTM or CNN.

9. Future work

Building on our study’s contributions to sepsis prediction in ICU settings, we identify several promising areas for further research. Integrating multi-institutional datasets from diverse geographical and clinical environments could enhance the robustness and applicability of our predictive models, moving beyond the insights provided by the MIMIC-IV dataset. Additionally, including additional physiological and biochemical markers could refine the models’ predictive accuracy, with future efforts focusing on emerging biomarkers identified in recent research for their potential in early sepsis detection. Implementing real-time, dynamic models using Digital Twin and Dynamic Ensemble Learning technologies in clinical settings represents a critical advancement, offering continuous risk assessments and timely interventions. Exploring cutting-edge machine learning techniques such as deep learning and reinforcement learning could address the complexities of sepsis progression, potentially uncovering new insights and enhancing model performance. Investigating how different treatment strategies affect model accuracy could further personalize sepsis management, aligning interventions with predictive insights. Tailoring models to individual patient profiles, including historical health data and demographic information, could significantly improve prediction personalization and accuracy. Assessing the economic and operational impact of these technologies within ICU settings would illuminate their cost-effectiveness and influence on clinical workflows. Continued attention to the ethical use and regulatory compliance of advanced predictive technologies is essential to ensure patient privacy and data security. These directions aim to advance the utility of predictive models in sepsis management, ultimately enhancing patient care and outcomes.

References

Al Shalabi, L., & Shaaban, Z. (2006). Normalization as a preprocessing engine for data mining and the approach of preference matrix. *2006 International Conference on Dependability of Computer Systems*, 207–214. https://doi.org/10.1109/DEPCOS-RELCOMEX.2006.38

Ali, A., Shamsuddin, S. M., & Ralescu, A. L. (2013). Classification with class imbalance problem. *Int. J. Advance Soft Compu. Appl*, *5*(3), 176–204. https://www.researchgate.net/publication/288228469_Classification_with_class_imbalance_problem_A_review

Arsiwala, A., Elghaish, F., & Zoher, M. (2023). Digital twin with Machine learning for predictive monitoring of CO2 equivalent from existing buildings. *Energy and Buildings*, *284*, 112851. https://doi.org/10.1016/j.enbuild.2023.112851

Awad, A., Bader-El-Den, M., McNicholas, J., & Briggs, J. (2017). Early hospital mortality prediction of intensive care unit patients using an ensemble learning approach. *International Journal of Medical Informatics*, *108*, 185–195. https://doi.org/10.1016/j.ijmedinf.2017.10.002

Barzekar, H., & Yu, Z. (2022). C-Net: A reliable convolutional neural network for biomedical image classification. *Expert Systems with Applications*, *187*, 116003. https://doi.org/10.1016/j.eswa.2021.116003

Behera, B., & Kumaravelan, G. (2020). Performance evaluation of machine learning algorithms in biomedical document classification. *Performance Evaluation*, *29*(5), 5704–5716. http://sersc.org/journals/index.php/IJAST/article/view/15054

Bowman, D., Dwyer, L., Levers, A., Patterson, E. A., Purdie, S., & Vikhorev, K. (2022). A unified approach to digital twin architecture—proof-of-concept activity in the nuclear sector. *Ieee Access*, *10*, 44691–44709. https://doi.org/10.1109/ACCESS.2022.3161626

Britto Jr, A. S., Sabourin, R., & Oliveira, L. E. S. (2014). Dynamic selection of classifiers—a comprehensive review. *Pattern Recognition*, *47*(11), 3665–3680. https://doi.org/10.1016/j.patcog.2014.05.003

Byrne, L., & Van Haren, F. (2017). Fluid resuscitation in human sepsis: time to rewrite history? *Annals of Intensive Care*, *7*, 1–8. https://doi.org/10.1186%2Fs13613-016-0231-8

Camacho-Cogollo, J. E., Bonet, I., Gil, B., & Iadanza, E. (2022). Machine learning models for early prediction of sepsis on large healthcare datasets. *Electronics*, *11*(9), 1507. https://doi.org/10.3390/electronics11091507

Contou, D., Roux, D., Jochmans, S., Coudroy, R., Guérot, E., Grimaldi, D., Ricome, S., Maury, E., Plantefeve, G., & Mayaux, J. (2016). Septic shock with no diagnosis at 24 hours:

a pragmatic multicenter prospective cohort study. *Critical Care*, *20*, 1–10. https://doi.org/10.1186/s13054-016-1537-5

Cruz, R. M. O., Sabourin, R., & Cavalcanti, G. D. C. (2018). Dynamic classifier selection: Recent advances and perspectives. *Information Fusion*, *41*, 195–216. https://doi.org/10.1016/j.inffus.2017.09.010

Deng, H.-F., Sun, M.-W., Wang, Y., Zeng, J., Yuan, T., Li, T., Li, D.-H., Chen, W., Zhou, P., & Wang, Q. (2022). Evaluating machine learning models for sepsis prediction: A systematic review of methodologies. *Iscience*, *25*(1). https://doi.org/10.1016/j.isci.2021.103651

Deng, Y. (2022). Digital twin-based modeling of complex systems for smart aging. *Discrete Dynamics in Nature and Society*, *2022*. https://doi.org/10.1155/2022/7365223

Ding, Y., Wang, Y., & Zhou, D. (2018). Mortality prediction for ICU patients combining just-in-time learning and extreme learning machine. *Neurocomputing*, *281*, 12–19. https://doi.org/10.1016/j.neucom.2017.10.044

Dröder, K., Bobka, P., Germann, T., Gabriel, F., & Dietrich, F. (2018). A machine learning-enhanced digital twin approach for human-robot-collaboration. *Procedia Cirp*, *76*, 187–192. https://doi.org/10.1016/j.procir.2018.02.010

Du, X., Liu, M., & Sun, Y. (2022). Segmentation, detection, and tracking of stem cell image by digital twins and lightweight deep learning. *Computational Intelligence and Neuroscience*, *2022*. https://doi.org/10.1155/2022/6003293

El-Rashidy, N., Abuhmed, T., Alarabi, L., El-Bakry, H. M., Abdelrazek, S., Ali, F., & El-Sappagh, S. (2022). Sepsis prediction in intensive care unit based on genetic feature optimization and stacked deep ensemble learning. *Neural Computing and Applications*, 1–30. https://doi.org/10.1007/s00521-021-06631-1

El-Rashidy, N., El-Sappagh, S., Abuhmed, T., Abdelrazek, S., & El-Bakry, H. M. (2020). Intensive care unit mortality prediction: An improved patient-specific stacking ensemble model. *IEEE Access*, *8*, 133541–133564. http://dx.doi.org/10.1109/ACCESS.2020.3010556

El-Sappagh, S., Abuhmed, T., Islam, S. M. R., & Kwak, K. S. (2020). Multimodal multitask deep learning model for Alzheimer's disease progression detection based on time series data. *Neurocomputing*, *412*, 197–215. https://doi.org/10.1016/j.neucom.2020.05.087

Ferdousi, R. (2021). *Digital Twin Disease Diagnosis Using Machine Learning*. Université d'Ottawa/University of Ottawa. http://dx.doi.org/10.20381/ruor-26990

Ferdousi, R., Laamarti, F., Hossain, M. A., Yang, C., & El Saddik, A. (2022). Digital twins for well-being: an overview. *Digital Twin*, *1*, 7. https://doi.org/10.12688/digitaltwin.17475.2

Ferrer, R., Martin-Loeches, I., Phillips, G., Osborn, T. M., Townsend, S., Dellinger, R. P., Artigas, A., Schorr, C., & Levy, M. M. (2014). Empiric antibiotic treatment reduces mortality in severe sepsis and septic shock from the first hour: results from a guideline-based performance improvement program. *Critical Care Medicine*, *42*(8), 1749–1755. https://doi.org/10.1097/CCM.0000000000000330

García, M. N. M., Herráez, J. C. B., Barba, M. S., & Hernández, F. S. (2016). Random forest based ensemble classifiers for predicting healthcare-associated infections in intensive care units. *Distributed Computing and Artificial Intelligence, 13th International Conference*, 303–311. https://doi.org/10.1007/978-3-319-40162-1_33

Guillén, J., Liu, J., Furr, M., Wang, T., Strong, S., Moore, C. C., Flower, A., & Barnes, L. E. (2015). Predictive models for severe sepsis in adult ICU patients. *2015 Systems and Information Engineering Design Symposium*, 182–187. https://doi.org/10.1109/SIEDS.2015.7116970

He, Z., Du, L., Zhang, P., Zhao, R., Chen, X., & Fang, Z. (2020). Early sepsis prediction using ensemble learning with deep features and artificial features extracted from clinical electronic health records. *Critical Care Medicine*, *48*(12), e1337–e1342. https://doi.org/10.1097/CCM.0000000000004644

Hou, W., Wang, X., Zhang, H., Wang, J., & Li, L. (2020). A novel dynamic ensemble selection classifier for an imbalanced data set: An application for credit risk assessment. *Knowledge-Based Systems*, *208*, 106462. https://doi.org/10.1016/j.knosys.2020.106462

Huang, X., Liang, Z., Li, T., Lingna, Y., Zhu, W., & Li, H. (2021). A nomogram to predict in-hospital mortality of neonates admitted to the intensive care unit. *International Health*, *13*(6), 633–639. https://doi.org/10.1093/inthealth/ihab012

Jaimes, F., Garcés, J., Cuervo, J., Ramírez, F., Ramírez, J., Vargas, A., Quintero, C., Ochoa, J., Tandioy, F., & Zapata, L. (2003). The systemic inflammatory response syndrome (SIRS) to identify infected patients in the emergency room. *Intensive Care Medicine*, *29*, 1368–1371. https://doi.org/10.1007/s00134-003-1874-0

Javan, S. L., Sepehri, M. M., Javan, M. L., & Khatibi, T. (2019). An intelligent warning model for early prediction of cardiac arrest in sepsis patients. *Computer Methods and Programs in Biomedicine*, *178*, 47–58. https://doi.org/10.1016/j.cmpb.2019.06.010

Javed, H., Muqeet, H. A., Javed, T., Rehman, A. U., & Sadiq, R. (2023). Ethical Frameworks for Machine Learning in Sensitive Healthcare Applications. *IEEE Access*. https://doi.org/10.1109/ACCESS.2023.3340884

Johnson, A., Bulgarelli, L., Pollard, T., Horng, S., Celi, L. A., & Mark, R. (2020). Mimic-iv. *PhysioNet. Available Online at: Https://Physionet.Org/Content/Mimiciv/1.0/(Accessed August 23, 2021)*, 49–55. https://doi.org/10.13026/s6n6-xd98

Jonathan, B., Putra, P. H., & Ruldeviyani, Y. (2020). Observation imbalanced data text to predict users selling products on female daily with smote, tomek, and smote-tomek. *2020 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)*, 81–85. https://doi.org/10.1109/IAICT50021.2020.9172033

Juraev, F., El-Sappagh, S., Abdukhamidov, E., Ali, F., & Abuhmed, T. (2022). Multilayer dynamic ensemble model for intensive care unit mortality prediction of neonate patients. *Journal of Biomedical Informatics*, *135*, 104216. https://doi.org/10.1016/j.jbi.2022.104216

Kefi, Z., Aloui, K., & Naceur, M. S. (2019). New approach based on machine learning for short-term mortality prediction in neonatal intensive care unit. *International Journal of Advanced Computer Science and Applications*, *10*(7). http://dx.doi.org/10.14569/IJACSA.2019.0100778

Kong, G., Lin, K., & Hu, Y. (2020). Using machine learning methods to predict in-hospital mortality of sepsis patients in the ICU. *BMC Medical*

Informatics and Decision Making, *20*, 1–10. https://doi.org/10.1186/s12911-020-01271-2

Krawczyk, B., Galar, M., Woźniak, M., Bustince, H., & Herrera, F. (2018). Dynamic ensemble selection for multi-class classification with one-class classifiers. *Pattern Recognition*, *83*, 34–51. https://doi.org/10.1016/j.patcog.2018.05.015

Krishnan, G. S., & Kamath, S. (2019). A novel GA-ELM model for patient-specific mortality prediction over large-scale lab event data. *Applied Soft Computing*, *80*, 525–533. http://dx.doi.org/10.1016/j.asoc.2019.04.019

Kumar, A., Roberts, D., Wood, K. E., Light, B., Parrillo, J. E., Sharma, S., Suppes, R., Feinstein, D., Zanotti, S., & Taiberg, L. (2006). Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. *Critical Care Medicine*, *34*(6), 1589–1596. https://doi.org/10.1097/01.ccm.0000217961.75225.e9

Kuncheva, L. I. (2002). A theoretical study on six classifier fusion strategies. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*(2), 281–286. https://doi.org/10.1109/34.982906

León, A. L., Hoyos, N. A., Barrera, L. I., De La Rosa, G., Dennis, R., Dueñas, C., Granados, M., Londoño, D., Rodríguez, F. A., & Molina, F. J. (2013). Clinical course of sepsis, severe sepsis, and septic shock in a cohort of infected patients from ten Colombian hospitals. *BMC Infectious Diseases*, *13*, 1–9. https://doi.org/10.1186/1471-2334-13-345

Lever, A., & Mackenzie, I. (2007). Sepsis: definition, epidemiology, and diagnosis. *Bmj*, *335*(7625), 879–883. https://doi.org/10.1136%2Fbmj.39346.495880.AE

Liu, Y., Zhang, L., Yang, Y., Zhou, L., Ren, L., Wang, F., Liu, R., Pang, Z., & Deen, M. J. (2019). A novel cloud-based framework for the elderly healthcare services using digital twin. *IEEE Access*, *7*, 49088–49101. http://dx.doi.org/10.1109/ACCESS.2019.2909828

Mansouri, A., Noei, M., & Abadeh, M. S. (2020). Predicting hospital length of stay of neonates admitted to the NICU using data mining techniques. *2020 10th International Conference on Computer and Knowledge Engineering (ICCKE)*, 629–635. http://dx.doi.org/10.1109/ICCKE50421.2020.9303666

Mao, Q., Jay, M., Hoffman, J. L., Calvert, J., Barton, C., Shimabukuro, D., Shieh, L., Chettipally, U., Fletcher, G., & Kerem, Y. (2015). Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU. BMJ Open. 2018; 8 (1): e017833. *Shock*, *44*(5), 390–396. https://doi.org/10.1136/bmjopen-2017-017833

Marik, P. E. (2015). The demise of early goal-directed therapy for severe sepsis and septic shock. *Acta Anaesthesiologica Scandinavica*, *59*(5), 561–567. https://doi.org/10.1111/aas.12479

Mazumder, O., Roy, D., Bhattacharya, S., Sinha, A., & Pal, A. (2019). Synthetic ppg generation from haemodynamic model with baroreflex autoregulation: a digital twin of cardiovascular system. *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 5024–5029. https://doi.org/10.1109/embc.2019.8856691

Megahed, F. M., Chen, Y.-J., Megahed, A., Ong, Y., Altman, N., & Krzywinski, M. (2021). The class imbalance problem. *Nat Methods*, *18*(11), 1270–1277. https://doi.org/10.1038/s41592-021-01302-4

Min, Q., Lu, Y., Liu, Z., Su, C., & Wang, B. (2019). Machine learning based digital twin framework for production optimization in petrochemical industry. *International Journal of Information Management*, *49*, 502–519. https://doi.org/10.1016/j.ijinfomgt.2019.05.020

MIT Technology Review Insights. (2022). *Digital twins improve real-life manufacturing*. https://www.technologyreview.com/2022/01/05/1042981/digital-twins-improve-real-life-manufacturing/

Nemati, S., Holder, A., Razmi, F., Stanley, M. D., Clifford, G. D., & Buchman, T. G. (2018). An interpretable machine learning model for accurate prediction of sepsis in the ICU. *Critical Care Medicine*, *46*(4), 547–553. https://doi.org/10.1097/ccm.0000000000002936

Polikar, R. (2006). Ensemble based systems in decision making. *IEEE Circuits and Systems Magazine*, *6*(3), 21–45. https://doi.org/10.1109/MCAS.2006.1688199

Pruinelli, L., Westra, B. L., Yadav, P., Hoff, A., Steinbach, M., Kumar, V., Delaney, C. W., & Simon, G. (2018). Delay within the 3-hour surviving sepsis campaign guideline on mortality for patients with severe sepsis and septic shock. *Critical Care Medicine*, *46*(4), 500–505. https://doi.org/10.1097/ccm.0000000000002949

Rafiei, A., Rezaee, A., Hajati, F., Gheisari, S., & Golzan, M. (2021). SSP: Early prediction of sepsis using fully connected LSTM-CNN model. *Computers in Biology and Medicine*, *128*, 104110. https://doi.org/10.1016/j.compbiomed.2020.104110

Rayan, Z., Alfonse, M., & Salem, A.-B. M. (2021). An Ensemble Model for Early Sepsis Prediction using Clinical Records from the Intensive Care Unit (ICU). *Egyptian Computer Science Journal*, *45*(1). https://www.researchgate.net/publication/348805257_An_Ensemble_Model_for_Early_Sepsis_Prediction_using_Clinical_Records_from_the_Intensive_Care_Unit_ICU

Ren, Y., Zhang, L., & Suganthan, P. N. (2016). Ensemble classification and regression-recent developments, applications and future directions. *IEEE Computational Intelligence Magazine*, *11*(1), 41–53. https://doi.org/10.1109/MCI.2015.2471235

Rhodes, A., Evans, L. E., Alhazzani, W., Levy, M. M., Antonelli, M., Ferrer, R., Kumar, A., Sevransky, J. E., Sprung, C. L., & Nunnally, M. E. (2017). Surviving sepsis campaign: international guidelines for management of sepsis and septic shock: 2016. *Intensive Care Medicine*, *43*, 304–377. https://doi.org/10.1007/s00134-017-4683-6

Rosnati, M., & Fortuin, V. (2021). MGP-AttTCN: An interpretable machine learning model for the prediction of sepsis. *Plos One*, *16*(5), e0251248. https://doi.org/10.1371/journal.pone.0251248

Rudd, K. E., Johnson, S. C., Agesa, K. M., Shackelford, K. A., Tsoi, D., Kievlan, D. R., Colombara, D. V, Ikuta, K. S., Kissoon, N., & Finfer, S. (2020). Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the Global Burden of Disease Study. *The Lancet*, *395*(10219), 200–211. https://doi.org/10.1016/s0140-6736(19)32989-7

Saracco, R. (2022). Digital Twins for Healthcare. IEEE Future Direction. https://cmte.ieee.org/futuredirections/2022/09/09/digital-twins-for-healthcare-ii/

Scheuermann, C., Binderberger, T., Von Frankenberg, N., & Werner, A. (2020). Digital twin: A machine learning approach to predict individual stress levels in extreme environments. *Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers*, 657–664. http://dx.doi.org/10.1145/3410530.3414316

Seymour, C. W., Liu, V. X., Iwashyna, T. J., Brunkhorst, F. M., Rea, T. D., Scherag, A., Rubenfeld, G., Kahn, J. M., Shankar-Hari, M., & Singer, M. (2016). Assessment of clinical criteria for sepsis: for the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *Jama*, *315*(8), 762–774. https://doi.org/10.1001/jama.2016.0288

Shengli, W. (2021). Is human digital twin possible? *Computer Methods and Programs in Biomedicine Update*, *1*, 100014. https://doi.org/10.1016/j.cmpbup.2021.100014

Singer, M., Deutschman, C. S., Seymour, C. W., Shankar-Hari, M., Annane, D., Bauer, M., Bellomo, R., Bernard, G. R., Chiche, J.-D., & Coopersmith, C. M. (2016). The third international consensus definitions for sepsis and septic shock (Sepsis-3). *Jama*, *315*(8), 801–810. https://doi.org/10.1001/jama.2016.0287

Sittig, D. F., Wright, A., Osheroff, J. A., Middleton, B., Teich, J. M., Ash, J. S., Campbell, E., & Bates, D. W. (2008). Grand challenges in clinical decision support. *Journal of Biomedical Informatics*, *41*(2), 387–392. https://doi.org/10.1016/j.jbi.2007.09.003

Sola, J., & Sevilla, J. (1997). Importance of input data normalization for the application of neural networks to complex industrial problems. *IEEE Transactions on Nuclear Science*, *44*(3), 1464–1468. https://doi.org/10.1109/23.589532

Sphinx. (2024). *Feature generation & reduction*. Dataiku. https://knowledge.dataiku.com/latest/ml-analytics/model-design/concept-feature-generation-reduction.html

Tancredi, G. P., Vignali, G., & Bottani, E. (2022). Integration of digital twin, machine-learning and industry 4.0 tools for anomaly detection: An application to a food plant. *Sensors*, *22*(11), 4143. https://doi.org/10.3390/s22114143

van Dinter, R., Tekinerdogan, B., & Catal, C. (2022). Predictive maintenance using digital twins: A systematic literature review. *Information and Software Technology*, *151*, 107008. https://doi.org/10.1016/j.infsof.2022.107008

Venkatesh, K. P., Raza, M. M., & Kvedar, J. C. (2022). Health digital twins as tools for precision medicine: Considerations for computation, implementation, and regulation. *NPJ Digital Medicine*, *5*(1), 150. https://doi.org/10.1038/s41746-022-00694-7

Vincent, J.-L., Moreno, R., Takala, J., Willatts, S., De Mendonça, A., Bruining, H., Reinhart, C. K., Suter, P., & Thijs, L. G. (1996). *The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure: On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine (see contributors to the project in the appendix)*. Springer-Verlag. https://doi.org/10.1007/bf01709751

Wang, X., Wang, Z., Weng, J., Wen, C., Chen, H., & Wang, X. (2018). A new effective machine learning framework for sepsis diagnosis. *IEEE Access*, *6*, 48300–48310. http://dx.doi.org/10.1109/ACCESS.2018.2867728

Xiao, J., He, C., Jiang, X., & Liu, D. (2010). A dynamic classifier ensemble selection approach for noise data. *Information Sciences*, *180*(18), 3402–3421. https://doi.org/10.1016/j.ins.2010.05.021

Yang, R., Han, D., Zhang, L., Huang, T., Xu, F., Zheng, S., Yin, H., & Lyu, J. (2022). Analysis of the correlation between the longitudinal trajectory of SOFA scores and prognosis in patients with sepsis at 72 hour after admission based on group trajectory modeling. *Journal of Intensive Medicine*, *2*(01), 39–49. https://doi.org/10.1016/j.jointm.2021.11.001

Yao, C., Dai, Q., & Song, G. (2019). Several novel dynamic ensemble selection algorithms for time series prediction. *Neural Processing Letters*, *50*, 1789–1829. https://doi.org/10.1007/s11063-018-9957-7

Yong, L., & Zhenzhou, L. (2024). Deep learning-based prediction of in-hospital mortality for sepsis. *Scientific Reports*, *14*(1), 372. https://doi.org/10.1038/s41598-023-49890-9

Zabihi, M., Kiranyaz, S., & Gabbouj, M. (2019). Sepsis prediction in intensive care unit using ensemble of XGboost models. *2019 Computing in Cardiology (CinC)*, Page-1. http://dx.doi.org/10.22489/CinC.2019.238

Zohdi, T. I. (2020). A machine-learning framework for rapid adaptive digital-twin based fire-propagation simulation in complex environments. *Computer Methods in Applied Mechanics and Engineering*, *363*, 112907. https://doi.org/10.22489/CinC.2019.238

국문요약

집중 치료실에서 패혈증 예측을 위한 디지털 트윈 기술과 동적 앙상블 학습의 통합

아미르호세인 더네시* · 피루즈 주라예프* · 샤케르 엘사파그* · 타메르 아부흐메드**

패혈증은 감염에 대한 과도한 면역 반응이 특징인 복잡하고 생명을 위협하는 상태로, 병원에서 높은 사망률을 초래한다. 빠르고 정확한 진단은 생존율을 향상시키는 데 필수적이지만, 현재의 관행은 개인화된 예측 도구가 부족하다. 전자 의무 기록의 등장은 자동화된 임상 의사 결정 시스템의 개발을 촉진시켰지만, 이러한 도구의 효과는 기계 학습(ML) 및 디지털 트윈(DT) 기술의 사용을 통해 크게 향상시킬 수 있다. 본 연구는 중환자실(ICU)에서 시계열 환자 데이터를 분석하기 위해 DT 기술을 통합한 새로운 계층적 ML 모델을 도입함으로써 패혈증 관리에서 중요한 격차를 메우고자 한다. MIMIC-IV 데이터 세트에서 강력한 코호트를 활용하여, 고전적, 정적 앙상블 및 동적 앙상블을 포함한 다양한 ML 모델을 구현하고 최적화한다. KNOP 모델이 기존의 방법론을 능가하며, 의학적으로 직관적이고 신뢰할 수 있는 패혈증 예측 방법을 제공한다는 것을 나타낸다. 이 선구적인 연구는 ICU 패혈증 예측에 DT 기술과 동적 앙상블을 적용한 최초의 연구로, 환자 맞춤형 건강 관리에서 미래의 발전을 위한 기초를 제공한다.

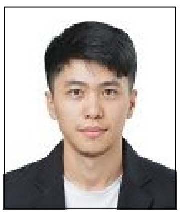
주제어 : 패혈증 예측, 디지털 트윈, 기계 학습, 중환자실, 시계열 분석

논문접수일 : 2024년 3월 26일　　논문수정일 : 2024년 4월 16일　　게재확정일 : 2024년 4월 17일
원고유형 : Fast Track　　교신저자 : Tamer Abuhmed

* 성균관대학교 소프트웨어융합대학
** 교신저자: Tamer Abuhmed
성균관대학교 소프트웨어융합대학
(16419) 경기도 수원시 장안구 서부로 2066
Tel: 031-290-7968, Fax: 031-290-7968, E-mail: tamer@skku.edu

저 자 소 개

Amirhossein Danesh
He received his bachelor's degree in advanced Electronics Engineering from Sogang University, Seoul, South Korea. After completing his undergraduate studies, he immersed himself in various research and development roles, focusing on integrating artificial intelligence and machine learning into practical applications. Amirhossein is currently pursuing further research opportunities that align with his broad interests, which also include robotics, smart systems, and IoT. His work aims to bridge the gap between theoretical research and real-world applications, making significant contributions to the field of digital technology.



Firuz Juraev
He received the B.S. from Inha University in Tashkent in 2019. He is currently pursuing the Master's and Ph.D. degrees in Computer Science and Engineering at Sungkyunkwan University, Suwon, South Korea. His main research interests are machine learning, ensemble learning optimization, and information security.

Shaker El-Sappagh
He received his Ph.D. in computer science from the Department of Information Systems, Faculty of Computers and Information, Mansoura University, Mansoura, Egypt, in 2015. Currently, he is an Associate Professor at Galala University, Egypt, and a Senior Researcher at Sungkyunkwan University, South Korea. His research focuses on machine learning, medical informatics, ontology engineering, clinical decision support systems, and cloud computing. He has also contributed to journals and is actively involved in research on disease diagnosis and treatment.

Tamer ABUHMED
He received his Ph.D. degree in information and telecommunication engineering from Inha University, in 2012. He is currently an associate professor with the College of Computing and Informatics, at Sungkyunkwan University, South Korea. His research interests include information security, trustworthy and adversarial machine learning, and expert systems for medical applications.