

ML Application in ICU (MIMIC_Database)

Subjects: Critical Care Medicine | Others

Contributor: Mahanazuddin Syed

Modern Intensive Care Units (ICUs) provide continuous monitoring of critically ill patients susceptible to many complications affecting morbidity and mortality. ICU settings require a high staff-to-patient ratio and generates a sheer volume of data. For clinicians, the real-time interpretation of data and decision-making is a challenging task. Machine Learning (ML) techniques in ICUs are making headway in the early detection of high-risk events due to increased processing power and freely available datasets such as the Medical Information Mart for Intensive Care (MIMIC). We conducted a systematic literature review to evaluate the effectiveness of applying ML in the ICU settings using the MIMIC dataset. A total of 322 articles were reviewed and a quantitative descriptive analysis was performed on 61 qualified articles that applied ML techniques in ICU settings using MIMIC data. We assembled the qualified articles to provide insights into the areas of application, clinical variables used, and treatment outcomes that can pave the way for further adoption of this promising technology and possible use in routine clinical decision-making.

Keywords: intensive care unit ; critical care ; MIMIC ; machine learning ; deep learning ; systematic review ; sepsis ; acute kidney injury ; mortality ; readmission

1. Introduction

Artificial intelligence (AI) encompasses a broad-spectrum of technologies that aim to imitate cognitive functions and intelligent behavior of humans ^[1]. Machine Learning (ML) is a subfield of AI that focuses on algorithms that allow computers to define a model for complex relationships or patterns from empirical data without being explicitly programmed ^[2]. ML, powered by increasing availability of healthcare data, is being used in a variety of clinical applications ranging from diagnosis to outcome prediction ^{[1][3]}. The predictive power of ML improves as the number of samples available for learning increases ^{[4][5]}.

ML algorithms can be supervised or unsupervised based on the type of learning rule employed. In supervised learning, an algorithm is trained using well-labeled data. Thereafter, the machine predicts on unseen data by applying knowledge gained from the training data ^[6]. Most adopted supervised ML models are Random Forest (RF), Support Vector Machines (SVM), and Decision Tree algorithms ^[6]. In unsupervised learning, there is no ground truth labeling required. Instead, the machine learns from the inherent structure of the unlabeled data ^[7]. Either type of ML is an iterative process in which the algorithm tries to find the optimal combination of both model variables and variable weights with the goal of minimizing error in the predicted outcome ^{[5][8]}. If the algorithm performs with a reasonably low error rate, it can be employed for making predictions where outputs are not known. However, while developing a ML model, an optimal bias-variance tradeoff should be selected to optimize prediction error rate ^[9]. Improper selection of bias and variance results in two problems: (1) underfitting and (2) overfitting ^[9]. Finding the “sweet spot” between the bias and variance is crucial to avoid both underfitting and overfitting ^{[8][10]}.

Deep learning (DL), a subcategory of machine learning, achieves great power and flexibility compared to conventional ML models by drawing inspiration from biological neural networks to solve a wide variety of complex tasks, including the classification of medical imaging and Natural Language Processing (NLP) ^{[10][11][12][13][14]}. Most widely used DL models are variants of Artificial Neural Network (ANN) and Multi-Layer Perceptron (MLP). In general, ML models are data driven and they rely on a deep understanding of the system for prediction, thereby, empowering users to make informed decision.

To provide better patient care and facilitate translational research, healthcare institutions are increasingly leveraging clinical data captured from Electronic Health Records (EHR) systems ^[15]. Of these systems, the Intensive Care Unit (ICU) generates an immense volume of data, and requires a high staff-to-patient ratio ^{[16][17]}. To avoid adverse events and prolonged ICU stays, early detection and intervention on patients vulnerable to complications is crucial; for these reasons, the ML literature is increasingly using ICU patient data for clinical event prediction and secondary usage, such as sepsis and septic shock ^[18]. ML techniques in ICUs are making headway in the early detection of high-risk events due to

increased processing power and freely available datasets such as the Medical Information Mart for Intensive Care (MIMIC) [19]. The data available in the MIMIC database includes highly structured data from time-stamped, nurse-verified physiological measurements made at the bedside, as well as unstructured data, including free-text interpretations of imaging studies provided by the radiology department [13].

2. Discussion

The aim of this study was to provide an up-to-date and holistic view of the current ML applications in ICU settings using MIMIC data in the attempt to predict clinical outcomes. Our review revealed ML application was widely adopted in areas such as mortality, risk stratification, readmission, and infectious disease in critically ill patients using retrospective data. This review may be used to provide insights for choosing key variables and best performing models for further research.

The application of ML techniques within the ICU domain is rapidly expanding with improvement of modern computing, which has enabled the analysis of huge volumes of complex and diverse data [1]. ML expands on existing statistical techniques, utilizing methods that are not based on a priori assumptions about the distribution of the data, but deriving insights directly from the data [20][21].

With ICUs being complex settings that generate a variety of time-sensitive data, more and more ML-based studies have begun tapping the openly available, large tertiary care hospital data (MIMIC). Our screening resulted in 61 publications that utilized MIMIC data to train and test ML models enabling reproducibility. The majority of these publications focused on predicting mortality, sepsis, AKI, and readmissions.

2.1. Mortality Prediction

Mortality prediction for ICU patients is critical and crucial for assessing severity of illness and adjudicating the value of treatments, and timely interventions. ML algorithms developed for predicting mortality in ICUs focused mainly on in-hospital mortality and 30 days mortality at discharge. Studies by Marafino et al. [22], Pirracchio et al. [23], Hoogendoorn et al. [24], Awad et al. [25], Davoodi et al. [26], and Weissman et al. [27] predicted in-hospital mortality, whereas Du et al. [28] predicted 28 days mortality at discharge, and Zahid et al. [29] predicted both 30 days and in-hospital mortality. Most studies focusing on predicting in-hospital mortality looked at mortality after 24 h of ICU admission. However, one in particular, Awad et al. [25], predicted mortality within 6 h of admission. Marafino et al. [22] predicted mortality using only nursing notes from the first 24 h of ICU admission, whereas Weissman et al. [27] improved mortality prediction by combining structured and unstructured data generated within the first 48 h of the ICU stay. Davoodi et al. [26] and Hoogendoorn et al. [24] predicted after 24 h and within a median of 72 h, respectively. Studies by Tang et al. [30], Caicedo-Torres et al. [31], Sha et al. [32], and Zhang et al. [33] predicted in-hospital mortality irrespective of the admission or discharge time.

For mortality prediction, all of the studies used three main categories of clinical variables: (1) demographics, (2) vital signs, and (3) laboratory test variables. In addition to the most commonly used data elements, other clinical information such as medications, intake/output variables, risk scores, and comorbidities were also utilized. Weissman et al. [27] and Zhang et al. [33] used clinical variables from both structured and unstructured data types for mortality prediction.

Multiple studies predicted mortality on disease-specific patient cohorts. Celi et al. [34] and Lin et al. [35] predicted in-hospital mortality on AKI patients. Lin et al. [35] predicted mortality based on five important variables (urine output, systolic blood pressure, age, serum bicarbonate level, and heart rate). In addition, the study by Lin et al. [35] also revealed that the effect of kidney injury markers, such as cystatin C and neutrophil gelatinase-associated lipocalin on subclinical injury, had not yet been analyzed, which can provide AKI prognostic information. This is due to lack of data availability in MIMIC. Garcia-Gallo et al. [36] and Kong et al. [37] predicted mortality on sepsis patients, and specifically, Garcia-Gallo et al. [36] identified patients that are on 1-year mortality trajectory. Anand et al. [38] claimed that the risk of mortality in diabetic patients could be better predicted using a combination of limited variables: HbA1c, mean glucose during stay, diagnoses upon admission, age, and type of admission. To compute the “diagnosis upon admission” variable, the study utilized Charlson Comorbidity Index, Elixhauser Comorbidity Index, and Diabetic Severity Index. The authors further claimed that combining diabetic-specific metrics and using the fewest possible variables would result in better mortality risk prediction in diabetic patients.

Studies have used both traditional ML (10 studies) and DL methods (11 studies) to predict mortality. In traditional ML techniques, Random Forest, Decision Tree, and Logistic Regression were the most commonly used algorithms. However, recent studies by Caicedo-Torres et al. [31], Du et al. [28], and Zahid et al. [29] have used DL methods for mortality prediction with a promising accuracy ranging from 0.86–0.87. Traditional ML models can be easily interpretable when compared to DL models that have many levels of features and hidden layers to predict outcomes. Understanding the

features that contribute towards the prediction plays an important role for clinical decision-making [39][40]. For example, one of the most cited studies by Pirracchio et al. [23] developed a mortality prediction algorithm (Super Learner) using a combination of traditional ML models; the results of which were easily interpretable by clinical researchers. In general, DL techniques are employed to improve prediction accuracy by training on large volumes of data [12]. Zahid et al. [29] developed a DL model (Self-Normalizing Neural Network (SNN)) that performed marginally better than the Pirracchio et al. [23] mortality prediction rate (Area Under the Receiver Operating Characteristic curve (AUROC) of SNN: 0.86 and Super Learner: 0.85). However, interpreting the results of DL models is challenging because of multiple hidden layers and they are often treated as black-box models. To address this limitation, Caicedo-Torres et al. [31] and Sha et al. [32] demonstrated the interpretability of the model in visualizations that will allow clinicians to make informed decisions.

2.2. Acute Kidney Injury (AKI) Prediction

AKI is one of the common complications among adult patients in the intensive care unit (ICU). AKI patients are at risk for adverse clinical outcomes such as prolonged ICU and hospitalization stays, high morbidity, and mortality. Application of ML in AKI care has been mainly focused on early prediction of an AKI event and risk stratification. In our review, studies employed traditional ML techniques to predict AKI events and XGBoost was the most commonly used algorithm.

Using the MIMIC dataset, Zimmerman et al. [41], Sun et al. [42], and Li et al. [43] predicted AKI after 24 h of ICU admission. Sun et al. [42] and Li et al. [43] used clinical unstructured notes generated during the first 24 h of ICU stay, whereas Zimmerman et al. [41] used structured clinical variables for prediction. The AUROC of predicting AKI within the first 24 h in Sun et al. [42], Zimmerman et al. [41], and Li et al. [43] was reported as 0.83, 0.783, and 0.779, respectively.

To define and classify AKI, three standard guidelines have been published and used in clinical settings: (1) Risk, Injury, Failure, Loss, End-Stage (RIFLE), (2) Acute Kidney Injury Network (AKIN), and (3) Kidney Disease: Improving Global Outcomes (KDIGO). In our results, most studies used KDIGO guidelines to create ground truth labels and is based on serum creatinine (SCr) and urine output. The SCr is one of the important predictor variables in AKI; however, it is a late marker of AKI, which delays diagnosis and care [44]. In clinical settings, it is highly desirable to early predict the AKI event for better intervention strategies. To address the aforementioned clinical need, Zimmerman et al. [41] predicted SCr values for 48 and 72 h based on 24 h SCr values and other clinical variables. Li et al. [43] extracted key features from clinical notes, such as diuretic and insulin medications using NLP instead of completely depending on SCr. Even though urine output is one of the defined metrics of AKI, Zimmerman et al. [41] reported it as not a significant predictor [45]. Further investigation should focus on the effect of urine output on predicting AKI and its impact.

2.3. Sepsis and Septic Shock

Sepsis is one of the leading causes of death among ICU patients and hospitalized patients overall [46]. As sepsis progresses, patients from pre-shock state are highly likely to develop septic shock. Early recognition of sepsis and initiation of treatment will reduce mortality and morbidity [47]. In our study, eight studies applied ML techniques to predict sepsis or septic shock events. Of these, four applied traditional ML algorithms and the other four used DL methods. XGBoost and LSTM were the most commonly used algorithms, of which the details of variables, sample size, and model performances are provided in the [Supplementary Table S1](#) (could download from the original paper link below).

The Scherpf et al. [48] model predicted sepsis 3 h prior to the onset with an AUROC of 0.81. The results of our review also reveal that most studies focused on early predicting of pre-shock state using hemodynamic measurements. The common variables used in ML models are arterial pressure, heart rate, labs, risk scores including Glasgow Coma Scores (GCS) and Sequential Organ Failure Assessment (SOFA) scores, and respiratory rate.

For predicting pre-shock state, Liu et al. [49] and Kam et al. [50] used a combination of these variables along with lab findings with the Area under the Curve (AUC) performance reported as 0.93 and 0.929, respectively. One of the interesting findings of the Liu et al. [49] study was that serum lactate was the primary predictor variable indicating a patient's risk level of entering septic shock, and is used as a biomarker for sepsis patient risk stratification. The study also reported, "A patient with serum lactate concentration one standard deviation above the population mean is approximately five times as likely to transition into shock than a patient with average serum lactate concentration" [49]. The hemodynamic measurements can be derived from waveform data or can be extracted as discrete data elements from EHR. Ghosh et al. [51] used three waveforms: mean arterial pressure, heart rate, and respiratory rate to derive hemodynamic predictor variables, whereas Liu et al. [49] and Kam et al. [50] used discrete measurements.

2.4. ICU Readmission

Intensive Care Units (ICU) provide care to critically ill patients, which is often costly and labor-intensive. Prolong ICU stays increases cost burden to both patients and hospitals. Early predicting unplanned readmissions may help in ICU resources allocation and improve patient health outcomes. Details of the studies qualified in this theme are listed in the [Supplementary Table S1](#) (could download from the original paper link below). Desautels et al. [52] identified patients who are likely to suffer unplanned ICU readmission: his model reported an AUROC of 0.71. Rojas et al. [53] and Lin et al. [54] focused on identifying patients that were re-admitted within 30 days of discharge. The best AUROC reported by Lin et al. [54] and Rojas et al. [53] is 0.791 and 0.78, respectively. The common predictor variables used in all three studies include: vital signs, demographics, comorbidities, and labs. Our findings revealed that there has been limited research done on predicting readmissions and the reported model AUROCs in literature are not promising (less than 80%) using MIMIC data.

2.5. ML Model Optimization

The performance of a given model heavily depended on data pre-processing, feature identification, and model validation. The missing data problem is arguably the most common issue encountered by machine learning practitioners when analyzing real-world healthcare data [55]. Researchers in general choose to address the missing data by either imputing or removing the observations [55]. The imputation can be done using simple-to-complex techniques: for example, in the study done by Lin et al. [35], missing observations were imputed using the mean value of the variable, whereas Davoodi et al. [26] and Zhang et al. [56] used sophisticated imputation techniques, Gaussian and Multivariate Imputation by Chained Equation (MICE), respectively. Substituting observed values with estimated observations introduces bias that may distort the data distribution or introduce spurious associations influencing model accuracy. To minimize this, imputation methods should be carefully selected, especially for prospective data. Imputation methodology depends on aim of the study, importance of data elements, percentage of missing data, and ML model used.

Feature importance technique is often employed to identify the highest ranked features. ML models with only important features improve the accuracy and computing time [57]. Cross-validation (CV) of a ML algorithm is vital to estimate a model's predictive power and generalized performance on the unseen data [58][59]. K-fold CV is often used to reduce the pessimistic bias by using more training data to teach the model. Our analysis found 52 studies used various validation techniques. Five-fold and 10-fold CV were the most common validation method used.

This review has some inherent limitations. First, there is the possibility of studies missed due to the search methodology. Second, we removed sixteen publications where full text was not available, and this may have introduced bias. Finally, a comparison of ML model performance was not possible in the quantitative analysis even though the studies used the MIMIC dataset for training and validating ML models. This is due to the fact that ML performance is dependent on the data elements selected for prediction, model parameters used, and size of the dataset.

2.6. Key Points and Recommendations

The aim of the study was to perform a comprehensive literature review on ML application in ICU settings using MIMIC dataset. The key points of our study and recommendations for future research provided therein are enlisted below.

1. The recent proliferation of publicly available MIMIC datasets allowed researchers to provide effective ML-based solutions in an attempt to solve complex healthcare problems. However, reproducibility of ML models is lacking due to inconsistent reporting of clinical variables selected, data pre-processing, and model specifications during the development. Future studies should follow standard reporting guidelines to accurately disclose model specifications.
2. Significant work has been done in predicting mortality within 6 to 72 h of hospital admission on retrospective data. However, prospective implementation is lacking. To adapt to dynamics of clinical events, we recommend exposing these models to prospective trials before moving it to routine clinical practice.

ML model performance heavily depends on clinical variables utilized. We identified and summarized the variables used by different model across the themes. Future studies should focus on performing a detailed analysis of these variables for improved performance.

3. Unstructured clinical notes have valuable and time-sensitive information critical for decision-making. Eight studies in our review tapped into clinical notes to mine important information. However, recent advancements in NLP techniques like Bidirectional Encoder Representations from Transformers (BERT) and Embeddings from Language Models (ELMo) have not been explored.

4. Interpretable ML models allow clinicians to understand and improve model performance. However, only two studies have resorted to visualization-based interpretations in the study.

References

1. Jiang, F.; Jiang, Y.; Zhi, H.; Dong, Y.; Li, H.; Ma, S.; Wang, Y.; Dong, Q.; Shen, H.; Wang, Y. Artificial intelligence in healthcare: Past, present and future. *Stroke Vasc. Neurol.* 2017, 2, 230–243.
2. Martins, J.; Magalhães, C.; Rocha, M.; Osório, N.S. Machine Learning-Enhanced T Cell Neoepitope Discovery for Immunotherapy Design. *Cancer Inform.* 2019, 18, 31205413.
3. Ahmed, Z.; Mohamed, K.; Zeeshan, S.; Dong, X. Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. *Database J. Biol. Databases Curation* 2020, 2020.
4. Figueroa, R.L.; Zeng-Treitler, Q.; Kandula, S.; Ngo, L.H. Predicting sample size required for classification performance. *BMC Med. Inform. Decis. Mak.* 2012, 12, 8.
5. Senders, B.J.T.; Arnaout, O.; Karhade, B.A.V.; Dasenbrock, H.H.; Gormley, W.B.; Broekman, M.L.; Smith, T.R. Natural and Artificial Intelligence in Neurosurgery: A Systematic Review. *Neurosurgery* 2017, 83, 181–192.
6. Uddin, S.; Khan, A.; Hossain, E.; Moni, M.A. Comparing different supervised machine learning algorithms for disease prediction. *BMC Med. Inform. Decis. Mak.* 2019, 19, 281.
7. De Langavant, L.C.; Bayen, E.; Yaffe, K. Unsupervised Machine Learning to Identify High Likelihood of Dementia in Population-Based Surveys: Development and Validation Study. *J. Med. Internet Res.* 2018, 20, e10493.
8. Belkin, M.; Hsu, D.; Ma, S.; Mandal, S. Reconciling modern machine-learning practice and the classical bias–variance trade-off. *Proc. Natl. Acad. Sci. USA* 2019, 116, 15849–15854.
9. Borstelmann, S.M. Machine Learning Principles for Radiology Investigators. *Acad. Radiol.* 2020, 27, 13–25.
10. Chartrand, G.; Cheng, P.M.; Vorontsov, E.; Drozdal, M.; Turcotte, S.; Pal, C.J.; Kadoury, S.; Tang, A. Deep Learning: A Primer for Radiologists. *Radiographics* 2017, 37, 2113–2131.
11. Torres, J.J.; Varona, P. Modeling Biological Neural Networks. In *Handbook of Natural Computing*; Rozenberg, G., Bäck, T., Kok, J.N., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 533–564.
12. Chen, D.; Liu, S.; Kingsbury, P.; Sohn, S.; Storlie, C.B.; Habermann, E.B.; Naessens, J.M.; Larson, D.W.; Liu, H. Deep learning and alternative learning strategies for retrospective real-world clinical data. *NPJ Digit. Med.* 2019, 2, 43.
13. Jang, J.; Ankit, A.; Kim, J.; Jang, Y.J.; Kim, H.Y.; Kim, J.H.; Xiong, S. A Unified Deep-Learning Model for Classifying the Cross-Country Skiing Techniques Using Wearable Gyroscope Sensors. *Sensors* 2018, 18, 3819.
14. Giger, M.L. Machine Learning in Medical Imaging. *J. Am. Coll. Radiol.* 2018, 15, 512–520.
15. Downing, N.L.; Rolnick, J.; Poole, S.F.; Hall, E.; Wessels, A.J.; Heidenreich, P.; Shieh, L. Electronic health record-based clinical decision support alert for severe sepsis: A randomised evaluation. *BMJ Qual. Saf.* 2019, 28, 762–768.
16. Kizzier-Carnahan, V.; Artis, K.A.; Mohan, V.; Gold, J.A. Frequency of Passive EHR Alerts in the ICU: Another Form of Alert Fatigue? *J. Patient Saf.* 2019, 15, 246–250.
17. Huddar, V.; Desiraju, B.K.; Rajan, V.; Bhattacharya, S.; Roy, S.; Reddy, C.K. Predicting Complications in Critical Care Using Heterogeneous Clinical Data. *IEEE Access* 2016, 4, 7988–8001.
18. Giannini, H.M.; Ginestra, J.C.; Chivers, C.; Draugelis, M.; Hanish, A.; Schweickert, W.D.; Fuchs, B.D.; Meadows, L.; Lynch, M.; Donnelly, P.J.; et al. A Machine Learning Algorithm to Predict Severe Sepsis and Septic Shock: Development, Implementation, and Impact on Clinical Practice. *Crit. Care Med.* 2019, 47, 1485–1492.
19. Johnson, A.E.W.; Pollard, T.J.; Shen, L.; Lehman, L.-W.H.; Feng, M.; Ghassemi, M.; Moody, B.; Szolovits, P.; Anthony Celi, L.; Mark, R.G. MIMIC-III, a freely accessible critical care database. *Sci. Data* 2016, 3, 160035.
20. Sanchez-Pinto, L.N.; Luo, Y.; Churpek, M.M. Big Data and Data Science in Critical Care. *Chest* 2018, 154, 1239–1248.
21. Panch, T.; Szolovits, P.; Atun, R. Artificial intelligence, machine learning and health systems. *J. Glob. Health* 2018, 8, 020303.
22. Marafino, B.J.; Boscardin, W.J.; Dudley, R.A. Efficient and sparse feature selection for biomedical text classification via the elastic net: Application to ICU risk stratification from nursing notes. *J. Biomed. Inform.* 2015, 54, 114–120.
23. Pirracchio, R.; Petersen, M.L.; Carone, M.; Rigon, M.R.; Chevret, S.; van der Laan, M.J. Mortality prediction in intensive care units with the Super ICU Learner Algorithm (SICULA): A population-based study. *Lancet Respir. Med.* 2015, 3, 42–52.

24. Hoogendoorn, M.; El Hassouni, A.; Mok, K.; Ghassemi, M.; Szolovits, P. Prediction Using Patient Comparison vs. Modeling: A Case Study for Mortality Prediction. In Proceedings of the Annual International Conference of the IEEE Engineering in Medi-Cine and Biology Society IEEE Engineering in Medicine and Biology Society Annual Conference, Orlando, FL, USA, 16–20 August 2016; pp. 2464–2467.
25. Awad, A.; Bader-El-Den, M.; McNicholas, J.; Briggs, J. Early hospital mortality prediction of intensive care unit patients using an ensemble learning approach. *Int. J. Med. Inform.* 2017, 108, 185–195.
26. Davoodi, R.; Moradi, M.H. Mortality prediction in intensive care units (ICUs) using a deep rule-based fuzzy classifier. *J. Biomed. Inform.* 2018, 79, 48–59.
27. Weissman, G.E.; Hubbard, R.A.; Ungar, L.H.; Harhay, M.O.; Greene, C.S.; Himes, B.E. Inclusion of Unstructured Clinical Text Improves Early Prediction of Death or Prolonged ICU Stay. *Crit. Care Med.* 2018, 46, 1125–1132.
28. Hao, D.; Ghassemi, M.M.; Mengling, F. The Effects of Deep Network Topology on Mortality Prediction. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Annual Conference, Orlando, FL, USA, 16–20 August 2016; pp. 2602–2605.
29. Zahid, M.A.H.; Lee, J. Mortality prediction with self normalizing neural networks in intensive care unit patients. In Proceedings of the 2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), Las Vegas, NV, USA, 4–7 March 2018; pp. 226–229.
30. Tang, F.; Xiao, C.; Wang, F.; Zhou, J. Predictive modeling in urgent care: A comparative study of machine learning approaches. *JAMIA Open* 2018, 1, 87–98.
31. Caicedo-Torres, W.; Gutierrez, J. ISeeU: Visually interpretable deep learning for mortality prediction inside the ICU. *J. Biomed. Inform.* 2019, 98, 103269.
32. Sha, Y.; Wang, M.D. Interpretable Predictions of Clinical Outcomes with An Attention-based Recurrent Neural Network. *ACM BCB* 2017, 2017, 233–240.
33. Zhang, D.; Yin, C.; Zeng, J.; Yuan, X.; Zhang, P. Combining structured and unstructured data for predictive models: A deep learning approach. *BMC Med. Inform. Decis. Mak.* 2020, 20, 280.
34. Celi, L.A.; Galvin, S.; Davidzon, G.; Lee, J.; Scott, D.; Mark, R. A Database-driven Decision Support System: Customized Mortality Prediction. *J. Pers. Med.* 2012, 2, 138–148.
35. Lin, K.; Hu, Y.; Kong, G. Predicting in-hospital mortality of patients with acute kidney injury in the ICU using random forest model. *Int. J. Med. Inform.* 2019, 125, 55–61.
36. García-Gallo, J.; Fonseca-Ruiz, N.; Celi, L.; Duitama-Muñoz, J. A machine learning-based model for 1-year mortality prediction in patients admitted to an Intensive Care Unit with a diagnosis of sepsis. *Med. Intensiv.* 2020, 44, 160–170.
37. Kong, G.; Lin, K.; Hu, Y. Using machine learning methods to predict in-hospital mortality of sepsis patients in the ICU. *BMC Med. Inform. Decis. Mak.* 2020, 20, 251.
38. Anand, R.S.; Stey, P.; Jain, S.; Biron, D.R.; Bhatt, H.; Monteiro, K.; Feller, E.; Ranney, M.L.; Sarkar, I.N.; Chen, E.S. Predicting Mortality in Diabetic ICU Patients Using Machine Learning and Severity Indices. *AMIA Jt. Summits Transl. Sci. Proc.* 2018, 2017, 310–319.
39. Davenport, T.; Kalakota, R. The potential for artificial intelligence in healthcare. *Futur. Health J.* 2019, 6, 94–98.
40. Elshaw, R.; Al-Mallah, M.H.; Sakr, S. On the interpretability of machine learning-based model for predicting hypertension. *BMC Med. Inform. Decis. Mak.* 2019, 19, 146.
41. Zimmerman, L.P.; Reyfman, P.A.; Smith, A.D.R.; Zeng, Z.; Kho, A.; Sanchez-Pinto, L.N.; Luo, Y. Early prediction of acute kidney injury following ICU admission using a multivariate panel of physiological measurements. *BMC Med. Inform. Decis. Mak.* 2019, 19, 16.
42. Sun, M.; Baron, J.; Dighe, A.; Szolovits, P.; Wunderink, R.G.; Isakova, T.; Luo, Y. Early Prediction of Acute Kidney Injury in Critical Care Setting Using Clinical Notes and Structured Multivariate Physiological Measurements. *MedInfo* 2019, 264, 368–372.
43. Li, Y.; Yao, L.; Mao, C.; Srivastava, A.; Jiang, X.; Luo, Y. Early Prediction of Acute Kidney Injury in Critical Care Setting Using Clinical Notes. In Proceedings of the 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Madrid, Spain, 3–6 December 2018.
44. De Geus, H.R.H.; Betjes, M.G.; Bakker, J. Biomarkers for the prediction of acute kidney injury: A narrative review on current status and future challenges. *Clin. Kidney J.* 2012, 5, 102–108.
45. Tsai, T.-Y.; Chien, H.; Tsai, F.-C.; Pan, H.-C.; Yang, H.-Y.; Lee, S.-Y.; Hsu, H.-H.; Fang, J.-T.; Yang, C.-W.; Chen, Y.-C. Comparison of RIFLE, AKIN, and KDIGO classifications for assessing prognosis of patients on extracorporeal membrane oxygenation. *J. Formos. Med. Assoc.* 2017, 116, 844–851.

46. Ogundipe, F.; Kodadhala, V.; Ogundipe, T.; Mehari, A.; Gillum, R. Disparities in Sepsis Mortality by Region, Urbanization, and Race in the USA: A Multiple Cause of Death Analysis. *J. Racial Ethn. Health Disparities* 2019, 6, 546–551.
47. Liu, V.; Escobar, G.J.; Greene, J.D.; Soule, J.; Whippy, A.; Angus, D.C.; Iwashyna, T.J. Hospital Deaths in Patients With Sepsis From 2 Independent Cohorts. *JAMA* 2014, 312, 90–92.
48. Scherpf, M.; Gräßer, F.; Malberg, H.; Zaunseder, S. Predicting sepsis with a recurrent neural network using the MIMIC III database. *Comput. Biol. Med.* 2019, 113, 103395.
49. Liu, R.; Greenstein, J.L.; Granite, S.J.; Fackler, J.C.; Bembea, M.M.; Sarma, S.V.; Winslow, R.L. Data-driven discovery of a novel sepsis pre-shock state predicts impending septic shock in the ICU. *Sci. Rep.* 2019, 9, 6145.
50. Kam, H.J.; Kim, H.Y. Learning representations for the early detection of sepsis with deep neural networks. *Comput. Biol. Med.* 2017, 89, 248–255.
51. Ghosh, S.; Li, J.; Cao, L.; Ramamohanarao, K. Septic shock prediction for ICU patients via coupled HMM walking on sequential contrast patterns. *J. Biomed. Inform.* 2017, 66, 19–31.
52. Desautels, T.; Das, R.; Calvert, J.; Trivedi, M.; Summers, C.; Wales, D.J.; Ercole, A. Prediction of early unplanned intensive care unit readmission in a UK tertiary care hospital: A cross-sectional machine learning approach. *BMJ Open* 2017, 7, e017199.
53. Rojas, J.C.; Carey, K.A.; Edelson, D.P.; Venable, L.R.; Howell, M.D.; Churpek, M.M. Predicting Intensive Care Unit Readmission with Machine Learning Using Electronic Health Record Data. *Ann. Am. Thorac. Soc.* 2018, 15, 846–853.
54. Lin, Y.-W.; Zhou, Y.; Faghri, F.; Shaw, M.J.; Campbell, R.H. Analysis and prediction of unplanned intensive care unit readmission using recurrent neural networks with long short-term memory. *PLoS ONE* 2019, 14, e0218942.
55. Tang, F.; Ishwaran, H. Random forest missing data algorithms. *Stat. Anal. Data Mining ASA Data Sci. J.* 2017, 10, 363–377.
56. Zhang, Z.; Ho, K.M.; Hong, Y. Machine learning for the prediction of volume responsiveness in patients with oliguric acute kidney injury in critical care. *Crit. Care* 2019, 23, 112.
57. Khan, N.M.; Madhav, C.N.; Negi, A.; Thaseen, I.S. Analysis on Improving the Performance of Machine Learning Models Using Feature Selection Technique. *Adv. Intell. Systems Comput.* 2020, 69–77.
58. Tabe-Bordbar, S.; Emad, A.; Zhao, S.D.; Sinha, S. A closer look at cross-validation for assessing the accuracy of gene regulatory networks and models. *Sci. Rep.* 2018, 8, 6620.
59. Raschka, S. Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning. *ArXiv* 2018, arXiv:1811.12808.