AI-enabled solutions, explainability and ethical concerns for predicting sepsis in ICUs: a systematic review

Christina-Athanasia Alexandropoulou, Ilias Panagiotopoulos, Styliani Kleanthous, George Dimitrakopoulos, Ioannis Constantinou, Eleni Politi, Lakis Palazis, Constantinos S. Pattichis, Andreas S. Panayides

Abstract—**Artificial Intelligence (AI) advances are pushing the boundaries across research domains with AI-driven solutions in healthcare claiming a significant share. A key objective of these studies concerns the timely prediction of various pathological conditions. Sepsis is a life-threatening syndrome and one of the main causes of death in intensive care unit (ICU) patients. As it becomes a major health problem worldwide, sepsis early prediction could assist healthcare professionals towards making informed clinical decisions, and thereby, significantly reducing the sepsis' morbidity and mortality. A notable body of literature involving the use of AI for sepsis prediction exists. However, to the best of our knowledge, only a handful of studies focus on performing a systematic review of the AI enabled solutions for sepsis prediction in ICUs. In this context, the present paper aims to identify knowledge gaps, stimulate interest and yield motivations for future research. Moreover, to discuss ethical and explainability aspects and associated challenges. The literature search was conducted between February 2023 and April 2023 and considered eligible articles published within the last five years.**

Key words: AI-enabled solutions, ethical AI, explainability, sepsis prediction, ICUs, systematic review

Christina-Arhanasia I. Alexandropoulou, Ilias E. Panagiotopoulos and George J. Dimitrakopoulos are with the Department of Informatics and Telematics, Harokopio University of Athens, Greece (e-mail: calexand@hua.gr; ipanagio@hua.gr; gdimitra@hua.gr)

Kleanthous Styliani, Constantinos S. Pattichis and Andreas S. Panayides are with CYENS CoE, Cyprus (email: s.kleanthous@cyens.org.cy; c.pattichis@cyens.org.cy; a.panayides@cyens.org.cy)

Ioannis Constantinou is with the 3aHealth, Cyprus (e-mail: c.ioannis@3ahealth.com)

Eleni Politi is with Ubitech, Athens, GREECE (e-mail: epoliti@ubitech.eu)

Lakis Palazis is with the State Health Services Organization, Cyprus (e-mail: l.palazis@shso.org.cy)

I. Introduction

Healthcare is inundated with enormous amounts of data, including electronic health/medical record (EHR), clinical trial results, imaging and laboratory test results, continuous physiological parameters monitoring data, -omics data, as well as non-clinical data, such as demographics [1]. These data are usually disorganized and fragmented, as they come from different sources. This is a challenge for healthcare professionals and systems alike, as they need to manage large amounts of diverse data, which they need to collect, process, and interpret to make an informed clinical decision. The timely availability and analysis of this heterogeneous big healthcare data are key prerequisites for the development of any healthcare system contributing to the early diagnosis, prognosis or treatment [2].

Technologies like Artificial Intelligence (AI) and Machine Learning (ML) are being increasingly used in the healthcare sector, trying to address the aforementioned challenges, while enhancing clinical decision-making [3]. These technologies focus on the development of AI algorithms and corresponding software, by leveraging the efficacy of the learning healthcare systems, patient monitoring systems and medical diagnostic systems. The goal is to facilitate the availability of responsive AI-enabled predictive analysis systems that underpin big and heterogeneous healthcare data management, towards timely and informed decisions about patients' care and treatment [4].

Complex conditions, such as sepsis are ideal for the application of AI in healthcare, as it is a life-threatening disease with high morbidity and mortality rates. Worldwide, an estimated 30 million people are diagnosed with sepsis in ICUs and 6 million people die from sepsis every year [5]. In addition, the hospital cost for treatment of sepsis is increasing every year. The study of Nemati et. al. (2018) supports that if the antibiotic treatment is delayed, the mortality is increased every hour [6]. In that context, early recognition of risk factors and immediate clinical intervention, before any sign of clinical symptoms, are crucial for reducing the mortality rates and financial burden of sepsis [7]. Many researchers are interested in developing diagnostic and prognostic tools through ML methods, for healthcare professionals to be able to identify patients with sepsis early and enhance prognosis by undertaking the most appropriate treatment strategy with high accuracy [5, 6, 7, 8].

Despite the enthusiastic deployment of ML solutions for predictive prognosis, potential ethical implications need to be taken into account. ML approaches are relying heavily on datasets used for training the algorithms in identifying patterns in a new set of data [9]. There have been a number of cases where ML algorithms produced biased results or acted in an unfair way towards certain groups of the population [10]. Thus, it is important to understand what might be the potential ethical issues that might arise from the use of ML also for predicting sepsis in the ICU. Furthermore, the opaque nature of these approaches makes the interpretation of potential results very difficult, and any possible error almost impossible to be identified [11]. Explainability techniques in AI (XAI) have been introduced as a medium through which black-box ML algorithms can be explained and interpreted by a lay user [12]. There is a need to understand how these approaches can be utilized for aiding healthcare practitioners make better decisions in a predictive prognosis context.

In the light of the above, the present paper aims to conduct a systematic literature review on AI-enabled solutions for sepsis prediction, by identifying the knowledge gaps and highlighting the importance for further research on AI applications for sepsis. Ethical aspects and challenges of using AI in predicting sepsis, as well as the explainability of AI-enabled solutions, are also investigated.

The structure of this paper is as follows: Section II presents the methodology followed in the present paper, section III includes the literature review about applications of AI in the early prediction and diagnosis of sepsis, section IC ethical aspects and challenges of using AI in predicting sepsis are presented and section V includes the explainability of AI enabled solutions. Finally, discussion and perspectives for future research, as well conclusions are drawn in section VI.

II. Literature Selection and Methodology

The aim of the present paper was to investigate the AI enabled solutions for septic shock prediction through a systematic review of the literature. This process is considered the most appropriate approach for medical searches in order to identify, select and critically appraise relevant studies [13]. To find information for this paper, we searched narrative reviews, systematic reviews and research papers in PubMed during February-April 2023, published during the last five years, which were then evaluated for eligibility. Abstracts without full text were excluded. The search terms used to find relevant literature included: (“artificial intelligence” OR “machine learning” OR “deep learning” OR “Internet Of Things”) AND (“sepsis” OR “septic shock”).

TABLE I. Summary of the Results From Related Works on the Prediction of Sepsis Onset.

Authorship	Year	Features	Best Model / Algorithm	AUROC	References
Shashikumar et. al.	2021	40	NN	0.8	[14]
Nemati et. al.	2018	65	Artificial Intelligence Sepsis Expert Algorithm	0.85	[6]
Guo et. al.	2022	46	CNN	0.92	[16]
Hou et. al.	2020	11	XGboost	0.857	[17]
Goh et al.	2021	100	NLP	0.94	[15]
Kong et. al.	2020	86	GBM	0.829	[18]
Persson et. al.	2021	20	CNN	0.9	[19]
Wang et. al.	2021	55	RF	0.91	[21]
Kuo-Ching et. al.	2020	106	XGboost	0.89	[20]
Alireza et. al.	2021	14	NN	0.86	[22]
Bai et. al.	2022	27	AdaBoost	0.895	[23]
Misra et. al.	2021	65	RF	0.943	[24]
Kok et. al.	2020	40	TCN	0.98	[25]
Xin Zhao et. al.	2021	25	LightGBM	0.979	[26]
Saqib et. al.	2018	47	RF	0.696	[27]
Ghias et. al.	2022	6	XGboost	0.98	[28]
Bedoya et al	2022	86	MGP–RNN	0.882	[29]
Lauritsen et. al.	2020	30	LSTM / CNN	0.856	[30]
Scherpf et al.	2019	101	RNN	0.81	[31]

NN: Neural Networks, CNN: Convolutional Neural Networks, XGboost: eXtreme Gradient Boosting, NLP: Neuro-Linguistic Programming, GBM: Gradient Boosting Machine, RF: Random Forest, AdaBoost: Adaptive Boosting, TCN: Temporal Convolutional Network, Light GBM: Light Gradient Boosting Machine, MGP-RNN: Multitask Gaussian Process-Recurrent Neural Network, LSTM: Long Short-term memory, RNN: Recurrent Neural Network

A total of 359 articles were initially identified with these search terms, of which 326 abstracts without full text were excluded, leading to a final count of 33 papers. In the present study, we highlighted the papers related to the development of models / algorithms that can predict the sepsis onset in ICUs, resulting in 19 papers as presented in Table 1.

III. AI Systems for the Early Prediction and Diagnosis of Sepsis

From the literature review, it is evident that there is a steady upward trend in the number of papers being published since 2018, which is likely attributed to the increasing availability and adoption of AI, ML and DL technologies over the same period. Researchers are showing an increasing interest, which is expected to reach even higher levels in the coming years, as new techniques for innovative health technologies are constantly being developed.

Driven by the observation that key to sepsis treatment is the early identification and immediate intervention, as every hour of delay in antimicrobial treatment increases mortality. A number of research studies have developed automated diagnostic tools for the early and accurate prediction, management and treatment of sepsis.

Shashikumar et. al. (2021) developed the deep learning model COMPOSER to predict onset of sepsis four to forty eight hours 4 prior to time of clinical suspicion. This model consists of three modules: the first module includes timing information and clinical variables, the second module incorporates a conformal prediction network, which yields a statistical framework for identification of the out-of-distribution samples and the third module uses a feed-forward neural network for sepsis prediction. This model involves 40 clinical variables and can be applied only in ICUs [14]. Another study introduced the AI-sepsis algorithm for the early prediction of sepsis in ICU patients. This algorithm includes 65 variables of interest on hourly basis and can precisely predict the sepsis almost four to twelve hours prior to clinical recognition [6]. Moreover, Goh et al. (2021) designed and developed an AI-enabled solution, named SERA, for the early detection of sepsis. This algorithm combines the EHR data with the clinical notes of health professionals, increasing the early detection of sepsis by up to 32% and reducing false positives by up to 17% [15]. Another model based on ML and DL technology was presented in determining the sepsis severity and characterizing the patients' phenotype. This model was based on DCQMFF and CNN for predicting the 28-day survival rate, and K-means to classify the sepsis phenotype. MIMIC-III and MIMIC-IV databases were used and achieved good performance [16]. MIMIC-III was also used by Hou et. al. (2020) for predicting the 30-days mortality patients with sepsis-3. In this study, an ML approach was developed, using a conventional logistic regression model, a SAPS-II score prediction model and an XGBoost algorithm model. XGBoost was proved to be clinically useful and assisted health professionals in the management and treatment of patients with sepsis-3 [17]. Moreover, Kong et. al. (2020) developed an ML-enabled tool for the prediction of the risk of patients' death with sepsis in ICUs. They used the MIMIC III database for the development and validation of the proposed tool and 86 predictor variables including comorbidities, laboratory tests and demographics. This model was based on the gradient boosting machine (GBM), least absolute shrinkage and selection operator (LASSO), traditional logistic regression (LR) method and random forest (RF). Amongst them, the GBM model showed the best performance for the prediction of the risk of sepsis death [18]. Furthermore, Persson et. al. (2021) developed a machine learning algorithm, named NAVOU sepsis, by using convolutional neural networks, based on MIMIC-III clinical data from ICU patients for predictions up to three hours before sepsis onset [19].

Kuo-Ching et. al. (2020) also developed an AI-enabled algorithm for early diagnosis of sepsis in ICU, by implementing ML methods like XGBoost, 5-fold cross-validation and decision-tree. They used 106 variables and real-time data, collected by EHR. Among ML methods, XGBoost was the most appropriate for the timely diagnosis of sepsis with an accuracy greater than 80% [20]. On the other hand, Wang et. al. (2021) developed an AI-enabled algorithm for the early prediction of sepsis in ICU patients, using the random forest ML method and the 5-fold cross-validation. The results of this analysis showed the random forest was the most suitable in predicting sepsis patients with high accuracy [21]. The study of Alireza et. al. (2021) proposed a deep neural network architecture, named SSP (Smart Sepsis Predictor), for sepsis prediction in ICU patients. They used the 2019 PhysioNet/CinC Challenge dataset and achieved to predict sepsis up to 12 h before onset [22]. In addition, the study of Bai et. al. (2022) developed a ML diagnostic model for the prediction of sepsis associated Acute Respiratory Distress Syndrome (ARDS) in ICU patients, by using early clinical indicators that were easily accessible. They used five different ML methods (Naive Bayes, Logistic Regression, Gradient Boosted Trees, AdaBoost model, Random Forest) and collected data from the e-ICU and the MIMIC-IV database [23]. Misra et. al. (2021) developed a clinical decision support system for septic shock prediction where clinical data from EHR and eight different ML algorithms (Decision Trees, Random Forest, Bayes Generalized Linear Model, XGBoost, C5.0, Logistic Regression, Boosted Logistic Regression, Regularized Logistic and Support Vector Machine) were used. The best model was based on Random Forest, with a specificity of 88.1% and sensitivity of 83.9% at one, three, and six hours from the time of admission [24]. Kok et al. (2020) employed a deep temporal convolution network, which can predict sepsis rapidly with high accuracy. They used the open-source dataset released for the PhysioNet Computing in Cardiology 2019 Challenge. The data obtained were based on sepsis-3 criteria and each patient's record comprised 40 features: 26 laboratory measurements, 6 demographic variables and 8 vital signs recorded hourly. They also used the Gaussian Process Regression (GPR) to predict the variations of potential values for each feature, a temporal convolutional network (TCN) which convolutes over the time domain and 10-fold cross validation to estimate the model [25]. Xin Zhao et. al. (2021) used the original data from the physiological ICU database from three independent hospital systems in order to propose a processing method that can predict sepsis six hours prior to onset [26]. Saqib et. al. (2018) developed a model that can predict sepsis 24 and 36 hours prior to its onset using vital signs and lab results. They used the MIMIC III dataset to test ML techniques including traditional methods, such as RF, LR and DL techniques [27]. Ghias et. al. (2022) proposed a model using ML algorithms (Linear learner, Multilayer perceptron neural networks, Random Forest, Lightgbm and XGboost) in order to predict sepsis at the admission time of patients in ICU. They used the Physio Net data and six vital signs extracted from patient records over the age of 18 years. XGboost model achieved a highest accuracy of 0.98, precision of 0.97, and recall 0.98 under the precision-recall curve on the publicly available data [28]. Bedoya et al. (2022) developed and validated a novel DL model to detect sepsis four hours prior to its onset. They used MGP–RNN, random forest (RF), Cox regression (CR), and penalized logistic regression (PLR), and three clinical scores, SIRS, quick Sequential Organ Failure Assessment (qSOFA), and National Early Warning Score (NEWS). MGP–RNN proved to be the most suitable in predicting sepsis patients with high accuracy [29].Lauritsen et. al. (2020) presented a deep learning system for early detection of sepsis using a combination of a long short-term memory network and a convolutional neural network. The data were taken by multiple Danish hospitals over a seven-year period and the results showed performance ranging from AUROC 0.856 (3 hours before sepsis onset) to AUROC 0.756 (24 hours before sepsis onset) [30]. Scherpf et al. (2019) proposed a recurrent neural network based approach for the prediction of sepsis onset and compared it to InSight algorithm. According to the results, the RNN showed an overall higher performance than the InSight algorithm with a maximum AUROC 0.81 and 0.72 respectively. The performance decreases with increasing prediction time for both models [31].

IV. Ethical Concerns of Using AI in Sepsis

Ethical guidelines and discussions have been around in healthcare and medical diagnosis for decades. The principles of medical ethics (beneficence, non-maleficence, justice, and human autonomy) [32] are guiding healthcare professionals over their daily duties for protecting vulnerable patients in uncertain contexts. The development of predictive analytics in the medical domain focused on helping healthcare professionals better serve those patients at risk, especially in ICU conditions. With the rapid development of AI, the discussion on ethics shifted towards the ethical implications in using ML for prognosis. ML algorithms - the computational approaches that are employed in predictive analytics - lack transparency and are usually not replicable. This is true for several AI-enabled systems that are used also in other non-life critical applications [33].

In an attempt to regulate the development and use of these applications the European Commission has issued a number of guidelines, focusing particularly on situations involving vulnerable people and the potential imbalance of information or power. Additionally, AI-enabled applications must adhere to the fundamental rights, societal values, and the ethical principles of explicability, prevention of harm, fairness, and human autonomy [34]. For the safe use of optimized systems, the satisfaction of these criteria is of crucial importance. However, most of these criteria cannot be fully quantified, i.e., it is extremely difficult to design the controls and tests needed for a clinical decision support system (CDSS), as transparency is required in predictions and decisions. ML algorithms are trained on historic data collected from patients over the years.

Recent work attempted to look into ethical implications of AI use in CDSS [35, 36, 37, 38, 39], highlighting a number of challenges that need to be taken into account.

Firstly, as in many medical situations there is a need for acquiring a consent from either the patient or their relatives for collecting and using their data for either training the algorithms or as input to the CDSS. Governments have provided the necessary guidelines under which these data can be collected, anonymized, processed and used for future statistical applications [40]. Consequently, data should be collected following relevant ethical guidelines that medical institutions are well aware of.

Secondly, non-representative data used in training the algorithms might lead to inequalities and biases in the prediction, which can have serious consequences. People coming from different ethnic backgrounds or gender, might have different medical needs and experiencing certain symptoms at a different pace. Hence, using gender and culturally relevant demographic data for ML training is vital in this context. Research, gained a lot of experience in this area from different domains (e.g., Recruitment, Justice, ImageTagging) [41] where AI-enabled systems treated people coming from different groups of the population in an unfair way. Similarly, using historical data of empirical diseases might lead to prognostically misleading outcomes. We need to understand that the above wrongly performed cases are the result of the use of historic data that are either baring human biases or are outdated. By using this data to train ML algorithms it will result in perpetuating biases or mistreating people. For example, Seymour et. al. followed an unsupervised learning method to identify four novel phenotypes of sepsis which differ with respect to biomarkers and mortality. Through this new approach, patients with sepsis could receive treatment that is more timely and appropriate [42]. Thus, in the context of predicting sepsis, we need to make sure that the datasets on which the developed AI-enabled systems are trained, will be reflecting the respective population.

Thirdly, the opaqueness of ML approaches makes it difficult for people to trust their outputs and foster accountability of actions. Physicians are usually discussing and explaining their thought process for identifying potential mistakes or preventing some. In a black-box system it is very difficult and in most occasions impossible to understand how the algorithm provided a certain output. Hence, knowing under what circumstances the system's decision should be trusted can be challenging. In ICU predictions, including sepsis prognosis, healthcare providers should be able to understand how the system ended up making a prediction, and whether this should be trusted. Human - AI interaction community is investigating different approaches for overcoming this issue (e.g., through explanations), however there is a lot of work that still needs to be done in this area, particularly in the context of ICU predictions [43]. Another very important factor is that of moral and legal accountability. There is a lot of discussion in the field on this concept particularly in relation to the use of AI in the medical domain [44]. In the case that a physician makes a mistake with any implications, there is usually a process that is followed for identifying the source of the error and who is to be held accountable. However, in the case of a wrongful prediction by the system, there is the issue of "who is to be held accountable?". Due to the opaqueness of ML models, it is very difficult to backtrack and understand which individual part of the chain is to be held accountable. In ICU conditions, where mortality rates are high, issues of accountability need to be addressed legally and also morally.

There are many complex ethical issues involved in the adoption of AI in ICU that can only be addressed if different disciplines work together (computer science, medicine, social science, psychology). Human-AI interaction is advocating for human - AI complementarity [45] where AI will support humans in exceeding performance of humans or AI alone.

V. Explainable AI and Sepsis

The above ethical and complexity issues that are entailed in AI CDSS, are highlighting the necessity for communicating critical information to the user. Explainability or the area of XAI (Explainable Artificial Intelligence) [46, 47], aims to achieve confidence, trustworthiness, accessibility, causality and transferability in predictions, so that health professionals can understand and correlate the results with the clinical practice [48, 49]. Over the last years, XAI has focused on improving the interaction between health professionals and AI systems.

Explanations that accompany the system's decision, proven to enhance trust between Human-AI [50, 51], however there is evidence that different types and levels of explanations are needed for different contexts [52, 53]. Local explanations focus on explaining a particular output; global explanations explain how a set of outputs emerges from a particular input; and counterfactual explanations attempt to help the user understand how their input could change the output of the system by resembling everyday human conversation [54]. Explanations can take different forms, for example, the use of natural language [55], visualizations [56], uncertainty of the model decision (e.g., confidence score) [57], example-based [58], etc. Whatever the type of explanation, the user should be able to understand the information provided by the system thus, explanations should be informative and easy to be interpreted by a specific audience. Research in the area of XAI, stresses the fact that explanations are context dependent. That emphasizes the need for specific studies to be performed for understanding what might be suitable explanation types and forms specifically for ICU applications in the context of sepsis prognosis. Furthermore, how algorithm predictions are presented to clinicians, and the extent to which they are accompanied by additional information or even recommendations, are key determinants of clinician acceptance of CDSS.

Explanations for the prediction of sepsis in ICU, in the form of graph visualizations were developed in [59, 60]. These studies have used data collected from real cases in ICU to evaluate their algorithmic model, however they do not provide any information on the impact that the explanations had on the clinical staff perception of the results. More recently, a comparative performance analysis of different feature selection for mortality prediction in ICU was performed in a set of data collected from COVID-19 patients. LIME was then used for providing explanation on the results [61]. Similarly, in [62] the authors developed two local level explanation techniques for assisting the understanding of the model prediction on the effect of Fibronectin on the survival of sepsis. According to the literature, explainable AI solutions for sepsis predictions are currently at early stages. Most of the researchers are exploring the predictive performance of different AI models, by using retrospective data and incorporating simplistic explainability methods. It is well known that the accurate sepsis prediction will help health professionals to choose appropriate treatment methods and better prognosis for sepsis patients [63]. However, all the above research studies have not evaluated their developed explanations for understanding - how the respective user groups perceive the explanations provided? whether the chosen XAI approach was efficient, if it enhances the clinical staffs' efficiency, and most importantly if it improves their trust towards the system.

Sepsis is easy to treat but hard to diagnose at the early stages of the disease. The opposite applies to later stages of the sepsis disease. In most cases, the diagnosis of sepsis is based on the organic dysfunction findings, the laboratory data and the general clinical situation of the patient. An explainable AI-enabled sepsis diagnosis tool could analyze a large number of features and variables related to sepsis, by giving a precise outcome for each ICU patient. As a result, health professionals may have an estimation of the possibility of sepsis appearance, which can help them change this clinical situation [64].

However, there are many difficulties for the development of explainable AI-enabled solutions of sepsis prediction. Not all hospitals are using EHR these days. These data are essential to be used on the one hand for training the ML algorithms used in CDSS and as on the other hand as input in an explainable AI-based system. Furthermore, according to the study of Beam and Kohane (2018), AI "is not a magic device that can spin data into gold" directly [65]. The above highlights the necessity for important scientific and human efforts in order for explainable AI-enabled tools to be used in each specific scenario.

As discussed in the previous section, large datasets should be extracted properly and processed in a systematic way for CDSS to provide accurate predictions and precise outcomes [66]. Currently, common sources of biomedical and demographic data are frequently used for the training of sepsis AI-related models, instead of sepsis related, ICU datasets [67]. This limits the efficiency and accuracy of the CDSS and of the explanations that accompany it. Hence, sepsis prognosis specific datasets are required to be extracted and curated for providing more accurate outcomes. Explainable tools that will accompany these outcomes will then be more informative, efficient, aid in building trust between the health workers and the CDSS and contribute towards accountability.

VI. CONCLUSIONS

In the present paper, we studied the applicability of AI enabled solutions for sepsis prediction, using a systematic literature review of various papers, which were evaluated for eligibility. The analysis reveals exponential growth in the number of research works published the last few years, while authors are interested in the development of AI algorithms. Also, the present analysis points out that those technologies can improve the field of health in a variety of ways, but further developments are essential for system security, accuracy, data collection and management, and privacy protection.

In light of the above, it can be stated that AI-enabled solutions are a very promising tool to improve sepsis prediction and detection and management. On the other hand, predicting the patients' mortality in ICUs from sepsis is more challenging, as is the wide adoption of such systems in standard clinical practice. This initial work points at several ethical implications of the use of AI - enabled systems particularly in the ICU. The opaqueness of the algorithmic approaches that are employed in these systems make it difficult for defining clear ethics guidelines and monitoring mechanisms. Explainability techniques appear to be a promising direction towards providing more transparency in CDSS and help the medical staff develop trust towards the outcomes of these systems. However, further work is needed in understanding what are suitable explainability methods and formats and how the different user groups perceive those. This paper was written by computer and data science experts, as well as health professionals. Therefore, this combination allows us to address essential aspects from both health and computer science fields.

On the other hand, there are some limitations. The present study is a literature review paper, and thus, not all available research studies on this topic of interest were considered. The aim of the present study was to provide an overview of the existing AI-enabled solutions for sepsis prediction, by following the above study design and selection.

Future research should focus on discussing clustering sepsis into different phenotypes. Also, future studies should focus on predicting the patients' mortality in ICUs from sepsis.

ACKNOWLEDGEMENT

This research has being performed in the context of the project Hospital Transformation through Artificial Intelligence – HOSPAITAL, CODEVELOP-ICT-HEALTH/0322/0071, which is co-financed by the Republic of Cyprus through the Research and Innovation Foundation and the European Regional Development Fund.

REFERENCES

[1] A. S. Panayides, M. Pattichis, S. Leandrou, C. Pitris, A. Constantinidou and C. S. Pattichis, "Radiogenomics for Precision Medicine With A Big Data Analytics Perspective," in IEEE Journal of Biomedical and Health Informatics, vol. 23, no. 5, pp. 2063-2079, Sept. 2019. doi: 10.1109/JBHI.2018.2879381.

[2] S. Purkayastha, J.W. Gichoya, Siva, A. Addepally, Implementation of a Single Sign- On System between Practice, Research and Learning Systems, 2017, https://doi. org/10.4338/ACI-2016-10-CR-0171.

[3] A. S. Panayides, A. Amini, N.D. Filipovic, A. Sharma, S. A. Tsaftaris, A. Young, D. Foran, N. Do, S. Golemati, T. Kurc, K. Huang, K. S. Nikita, B.P. Veasey, M. Zervakis, J.H. Saltz, C.S. Pattichis, "AI in Medical Imaging Informatics: Current Challenges and Future Directions," in IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 7, pp. 1837-1857, July 2020, doi: 10.1109/JBHI.2020.2991043.

[4] Lysaght, T., Lim, H.Y., Xafis, V. et al. AI-Assisted Decision-making in Healthcare. ABR 11, 299–314 (2019). https://doi.org/10.1007/s41649-019-00096-0.

[5] Zabihi, M., Kiranyaz, S. and Gabbouj, M. (2019). Sepsis prediction in intensive care unitusing ensemble of xgboost models. In 2019 Computing in Cardiology (CinC), page 1, IEEE.

[6] Nemati S, Holder A, Razmi F, Stanley MD, Clifford GD, Buchman TG. An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU. Crit Care Med. 2018 Apr;46(4):547-553. doi: 10.1097/CCM.0000000000002936. PMID: 29286945; PMCID: PMC5851825.

[7] R.P. Dellinger, M.M. Levy, A. Rhodes, et al., Surviving sepsis campaign, Crit. Care Med. 41 (2) (2013) 580–637, https://doi.org/10.1097/CCM.0b013e31827e83af.

[8] V.M. Quinten, M. van Meurs, J.J. Ligtenberg, J.C. Ter Maaten, Prehospital antibiotics for sepsis: beyond mortality? Lancet Respir. Med. 6 (3) (2018) e8, https://doi.org/10.1016/S2213-2600(18)30061-4.

[9] Ricardo Baeza-Yates. 2018. Bias on the web. Commun. ACM 61, 6 (June 2018), 54–61. https://doi.org/10.1145/3209581

[10] Pinar Barlas, Kyriakos Kyriakou, Olivia Guest, Styliani Kleanthous, and Jahna Otterbacher. 2021. To "See" is to Stereotype: Image Tagging Algorithms, Gender Recognition, and the Accuracy-Fairness Trade-off. Proc. ACM Hum.-Comput. Interact. 4, CSCW3, Article 232 (December 2020), 31 pages. https://doi.org/10.1145/3432931

[11] Mittelstadt BD, Allo P, Taddeo M, Wachter S, Floridi L (2016) The ethics of algorithms: mapping the debate. Big Data & Society 2:1–21

[12] David Gunning et al. ,XAI—Explainable artificial intelligence.Sci. Robot.4,eaay7120(2019).DOI:10.1126/scirobotics.aay7120

[13] Hannah Snyder, Literature review as a research methodology: An overview and guidelines, Journal of Business Research, Volume 104, 2019, Pages 333-339, ISSN 0148-2963,

[14] Shashikumar SP, Wardi G, Malhotra A, Nemati S. Artificial intelligence sepsis prediction algorithm learns to say "I don't know". NPJ Digit Med. 2021 Sep 9;4(1):134. doi: 10.1038/s41746-021-00504-6. PMID: 34504260; PMCID: PMC8429719.

[15] Goh, K.H., Wang, L., Yeow, A.Y.K. et al. Artificial intelligence in sepsis early prediction and diagnosis using unstructured data in healthcare. Nat Commun 12, 711 (2021). https://doi.org/10.1038/s41467-021-20910-4.

[16] Guo F, Zhu X, Wu Z, Zhu L, Wu J, Zhang F. Clinical applications of machine learning in the survival prediction and classification of sepsis: coagulation and heparin usage matter. J Transl Med. 2022 Jun 11;20(1):265. doi: 10.1186/s12967-022-03469-6. PMID: 35690822; PMCID: PMC9187899.

[17] Hou N, Li M, He L, Xie B, Wang L, Zhang R, Yu Y, Sun X, Pan Z, Wang K. Predicting 30-days mortality for MIMIC-III patients with sepsis-3: a machine learning approach using XGboost. J Transl Med. 2020 Dec 7;18(1):462. doi: 10.1186/s12967-020-02620-5. PMID: 33287854; PMCID: PMC7720497.

[18] 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 20, 251 (2020). https://doi.org/10.1186/s12911-020-01271-2.

[19] Persson I., Östling A., Arlbrandt M., Söderberg J., Becedas D., A Machine Learning Sepsis Prediction Algorithm for Intended Intensive Care Unit Use (NAVOY Sepsis): Proof-of-Concept Study JMIR Form Res 2021;5(9):e28000.

[20] Kuo-Ching Yuan, Lung-Wen Tsai, Ko-Han Lee, Yi-Wei Cheng, Shou-Chieh Hsu, Yu-Sheng Lo, Ray-Jade Chen, The development an artificial intelligence algorithm for early sepsis diagnosis in the intensive care unit, International Journal of Medical Informatics, Volume 141, 2020, 104176, ISSN 1386-5056, https://doi.org/10.1016/j.ijmedinf.2020.104176.

[21] Wang D, Li J, Sun Y, Ding X, Zhang X, Liu S, Han B, Wang H, Duan X and Sun T (2021) A Machine Learning Model for Accurate Prediction of Sepsis in ICU Patients. Front. Public Health 9:754348. doi: 10.3389/fpubh.2021.754348.

[22] Alireza Rafiei, Alireza Rezaee, Farshid Hajati, Soheila Gheisari, Mojtaba Golzan, SSP: Early prediction of sepsis using fully connected LSTM-CNN model, Computers in Biology and Medicine, Volume 128, 2021, 104110, ISSN 0010-4825, https://doi.org/10.1016/j.compbiomed.2020.104110.

[23] Bai Y, Xia J, Huang X, Chen S and Zhan Q (2022), Using machine learning for the early prediction of sepsis-associated ARDS in the ICU and identification of clinical phenotypes with differential responses to treatment. Front. Physiol. 13:1050849. doi: 10.3389/fphys.2022.1050849.

[24] Misra D, Avula V, Wolk DM, Farag HA, Li J, Mehta YB, Sandhu R, Karunakaran B, Kethireddy S, Zand R, Abedi V. Early Detection of Septic Shock Onset Using Interpretable Machine Learners. Journal of Clinical Medicine. 2021; 10(2):301. https://doi.org/10.3390/jcm10020301.

[25] Kok, C., Jahmunah, V., Oh, S. L., Zhou, X., Gururajan, R., Tao, X., … Acharya, U. R. (2020). Automated prediction of sepsis using temporal convolutional network. Computers in Biology and Medicine, 103957. doi:10.1016/j.compbiomed.2020.103

[26] Xin Zhao, Wenqian Shen, Guanjun Wang, "Early Prediction of Sepsis Based on Machine Learning Algorithm", Computational Intelligence and Neuroscience, vol. 2021, Article ID 6522633, 13 pages, 2021. https://doi.org/10.1155/2021/6522633

[27] Saqib, M., Sha, Y., & Wang, M. D. (2018). Early Prediction of Sepsis in EMR Records Using Traditional ML Techniques and Deep Learning LSTM Networks. 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). doi:10.1109/embc.2018.8513254

[28] Ghias N., Ul Haq S., Arshad H., Sultan H., Bashir F., Ghaznavi S.A., Shabbir M., Badshah Y., Rafiq M., Using Machine Learning Algorithms to predict sepsis and its stages in ICU patients, medRxiv, March 2022, doi: https://doi.org/10.1101/2022.03.15.22271655

[29] Armando D Bedoya and others, Machine learning for early detection of sepsis: an internal and temporal validation study, JAMIA Open, Volume 3, Issue 2, July 2020, Pages 252–260, https://doi.org/10.1093/jamiaopen/ooaa006

[30] Lauritsen SM, Kalør ME, Kongsgaard EL, Lauritsen KM, Jørgensen MJ, Lange J, Thiesson B. Early detection of sepsis utilizing deep learning on electronic health record event sequences. Artif Intell Med. 2020 Apr;104:101820. doi: 10.1016/j.artmed.2020.101820. Epub 2020 Feb 19. PMID: 32498999.

[31] Matthieu Scherpf, Felix Gräßer, Hagen Malberg, Sebastian Zaunseder, Predicting sepsis with a recurrent neural network using the MIMIC III database, Computers in Biology and Medicine, Volume 113, 2019, 103395, ISSN 0010-4825, https://doi.org/10.1016/j.compbiomed.2019.103395.

[32] Gillon R (2015) Defending the four principles approach as a good basis for good medical practice and therefore for good medical ethics. J Med Ethics 41:111–116

[33] Cohen, I. Glenn, Ruben Amarasingham, Anand Shah, Bin Xie, and Bernard Lo. "The legal and ethical concerns that arise from using complex predictive analytics in health care." Health affairs 33, no. 7 (2014): 1139-1147.

[34] High-Level Expert Group on Artificial Intelligence (2019) Ethics guidelines for trustworthy AI. European Commission, Brussels

[35] Morley, Jessica, Caio CV Machado, Christopher Burr, Josh Cowls, Indra Joshi, Mariarosaria Taddeo, and Luciano Floridi. "The ethics of AI in health care: a mapping review." Social Science & Medicine 260 (2020): 113172.

[36] Keskinbora, Kadircan H. "Medical ethics considerations on artificial intelligence." Journal of clinical neuroscience 64 (2019): 277-282.

[37] Beil M, Proft I, van Heerden D, Sviri S, van Heerden PV. Ethical considerations about artificial intelligence for prognostication in intensive care. Intensive Care Med Exp. 2019 Dec 10;7(1):70. doi: 10.1186/s40635-019-0286-6. PMID: 31823128; PMCID: PMC6904702.

[38] Bjerring, J.C., Busch, J. Artificial Intelligence and Patient-Centered Decision-Making. Philos. Technol. 34, 349–371 (2021). https://doi.org/10.1007/s13347-019-00391-6

[39] Cohen, I. Glenn, Ruben Amarasingham, Anand Shah, Bin Xie, and Bernard Lo. "The legal and ethical concerns that arise from using complex predictive analytics in health care." Health affairs 33, no. 7 (2014): 1139-1147.

[40] 45 C.F.R. Sect. 46.102(f) (2012)

[41] Pinar Barlas, Styliani Kleanthous, Kyriakos Kyriakou, and Jahna Otterbacher. 2019. What Makes an Image Tagger Fair? In Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization (UMAP '19). Association for Computing Machinery, New York, NY, USA, 95–103. https://doi.org/10.1145/3320435.3320442

[42] Seymour CW, Kennedy JN, Wang S, et al. Derivation, Validation, and Potential Treatment Implications of Novel Clinical Phenotypes for Sepsis. JAMA. 2019;321(20):2003–2017. doi:10.1001/jama.2019.5791

[43] Ngiam, Kee Yuan, and Wei Khor. "Big data and machine learning algorithms for health-care delivery." The Lancet Oncology 20, no. 5 (2019): e262-e273.

[44] Beil, Michael, Ingo Proft, Daniel van Heerden, Sigal Sviri, and Peter Vernon van Heerden. "Ethical considerations about artificial intelligence for prognostication in intensive care." Intensive Care Medicine Experimental 7, no. 1 (2019): 1-13. Morley, Jessica, Caio CV

[45] Inkpen, Kori, Shreya Chappidi, Keri Mallari, Besmira Nushi, Divya Ramesh, Pietro Michelucci, Vani Mandava, Libuše Hannah Vepřek, and Gabrielle Quinn. "Advancing Human-AI Complementarity: The Impact of User Expertise and Algorithmic Tuning on Joint Decision Making." arXiv preprint arXiv:2208.07960 (2022).

[46] Doshi-Velez, F. and Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608.

[47] AKM Bahalul Haque, A.K.M. Najmul Islam, Patrick Mikalef, Explainable Artificial Intelligence (XAI) from a user perspective: A synthesis of prior literature and problematizing avenues for future research, Technological Forecasting and Social Change, Volume 186, Part A, 2023, 122120, ISSN 0040-1625, https://doi.org/10.1016/j.techfore.2022.122120.

[48] Arrieta, A.B.; Díaz-Rodríguez, N.; Del Ser, J.; Bennetot, A.; Tabik, S.; Barbado, A.; García, S.; Gil-López, S.; Molina, D.; Benjamins, R.; et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Inf. Fusion 2020, 58, 82–115.

[49] Holzinger, A., Biemann, C., Pattichis, C. S., & Kell, D. B. (2017). What do we need to build explainable AI systems for the medical domain?. arXiv preprint arXiv:1712.09923.

[50] Vivian Lai, Samuel Carton, Rajat Bhatnagar, Q. Vera Liao, Yunfeng Zhang, and Chenhao Tan. 2022. Human-AI Collaboration via Conditional Delegation: A Case Study of Content Moderation. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (New Orleans, LA, USA) (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 54, 18 pages. https://doi.org/10.1145/3491102.3501999

[51] Helena Vasconcelos, Matthew Jörke, Madeleine Grunde-McLaughlin, Tobias Gerstenberg, Michael S. Bernstein, and Ranjay Krishna. 2023. Explanations Can Reduce Overreliance on AI Systems During Decision-Making. Proc. ACM Hum.-Comput. Interact. 7, CSCW1, Article 129 (April 2023), 38 pages. https://doi.org/10.1145/3579605

[52] Lilian Edwards and Michael Veale. 2017. Slave to the algorithm: Why a right to an explanation is probably not the remedy you are looking for. Duke L. & Tech. Rev. 16 (2017), 18

[53] Maria Kasinidou, Styliani Kleanthous, Pınar Barlas, and Jahna Otterbacher. 2021. I agree with the decision, but they didn't deserve this: Future Developers' Perception of Fairness in Algorithmic Decisions. In Proceedings of the 2021 acm conference on fairness, accountability, and transparency. 690–700

[54] Leif Hancox-Li. 2020. Robustness in machine learning explanations: does it matter? In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20). Association for Computing Machinery, New York, NY, USA, 640–647. https://doi.org/10.1145/3351095.3372836

[55] Gagan Bansal, Tongshuang Wu, Joyce Zhou, Raymond Fok, Besmira Nushi, Ece Kamar, Marco Tulio Ribeiro, and Daniel Weld. 2021. Does the Whole Exceed its Parts? The Effect of AI Explanations on Complementary Team Performance. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21). Association for Computing Machinery, New York, NY, USA, Article 81, 1–16. https://doi.org/10.1145/3411764.3445717

[56] Alicioglu, Gulsum, and Bo Sun. "A survey of visual analytics for Explainable Artificial Intelligence methods." Computers & Graphics 102 (2022): 502-520.

[57] Zana Buçinca, Maja Barbara Malaya, and Krzysztof Z. Gajos. 2021. To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-assisted Decision-making. Proc. ACM Hum.-Comput. Interact. 5, CSCW1, Article 188 (April 2021), 21 pages. https://doi.org/10.1145/3449287

[58] Shi Feng and Jordan Boyd-Graber. 2019. What can AI do for me? evaluating machine learning interpretations in cooperative play. In Proceedings of the 24th International Conference on Intelligent User Interfaces (IUI '19). Association for Computing Machinery, New York, NY, USA, 229–239. https://doi.org/10.1145/3301275.3302265

[59] Yang, Meicheng BS1; Liu, Chengyu PhD1; Wang, Xingyao MS1; Li, Yuwen PhD1; Gao, Hongxiang BS1; Liu, Xing MD, PhD2; Li, Jianqing PhD1,3. An Explainable Artificial Intelligence Predictor for Early Detection of Sepsis. Critical Care Medicine 48(11):p e1091-e1096, November 2020. | DOI: 10.1097/CCM.0000000000004550

[60] Nesaragi, Naimahmed, and Shivnarayan Patidar. 2021. 'An Explainable Machine Learning Model for Early Prediction of Sepsis Using ICU Data'. Infections and Sepsis Development. IntechOpen. doi:10.5772/intechopen.98957.

[61] N. Tasnim and S. A. Mamun, "Comparative Performance Analysis of Feature Selection for Mortality Prediction in ICU with Explainable Artificial Intelligence," 2023 International Conference on Electrical, Computer and Communication Engineering (ECCE), Chittagong, Bangladesh, 2023, pp. 1-6, doi: 10.1109/ECCE57851.2023.10101553.

[62] Lemańska-Perek, Anna, Dorota Krzyżanowska-Gołąb, Katarzyna Kobylińska, Przemysław Biecek, Tomasz Skalec, Maciej Tyszko, Waldemar Gozdzik, and Barbara Adamik. 2022. "Explainable Artificial Intelligence Helps in Understanding the Effect of Fibronectin on Survival of Sepsis" Cells 11, no. 15: 2433. https://doi.org/10.3390/cells11152433

[63] Yang, C.C. Explainable Artificial Intelligence for Predictive Modeling in Healthcare. J Healthc Inform Res 6, 228–239 (2022). https://doi.org/10.1007/s41666-022-00114-1.

[64] Vincent JL. The Clinical Challenge of Sepsis Identification and Monitoring. PLoS Med. 2016 May 17;13(5):e1002022. doi:

10.1371/journal.pmed.1002022. PMID: 27187803; PMCID: PMC4871479.

[65] A.L. Beam, I.S. Kohane. Big data and machine learning in health care. Vol. 319, JAMA – Journal of the American Medical Association. American Medical Association, (2018), pp. 1317-1318

[66] M.A. Gianfrancesco, S. Tamang, J. Yazdany, G. Schmajuk. Potential biases in machine learning algorithms using electronic health record data. JAMA Intern Med, 178 (2018), pp. 1544. http://dx.doi.org/10.1001/jamainternmed.2018.3763.

[67] J. Wiens, S. Saria, M. Sendak, M. Ghassemi, V.X. Liu, F. Doshi-Velez, et al. Do no harm: a roadmap for responsible machine learning for health care. Nat. Med., 25 (2019), pp. 1337-1340. http://dx.doi.org/10.1038/s41591-019-0548-6.