IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON EMERGING DEEP LEARNING THEORIES AND METHODS FOR BIOMEDICAL ENGINEERING

Received August 29, 2020, accepted November 22, 2020, date of publication December 10, 2020,
date of current version January 11, 2021.

Digital Object Identifier 10.1109/ACCESS.2020.3043732

A Deep Learning-Based Sepsis Estimation Scheme

BILAL YASEEN AL-MUALEMI AND LU LU, (Member, IEEE)
School of Computer Science and Engineering, South China University of Technology, Guangzhou 510000, China
Corresponding author: Lu Lu (lul@scut.edu.cn)

ABSTRACT The objective of this research is to design and implement a machine learning (ML) based technique that can predict cases of septic shock and extreme sepsis and assess its effects on medical practice and the patients. The study is a retrospective cohort type, which is used to algorithmic deduction and validation, along with pre- and post-impact assessment. For non-ICU cases, the algorithm was deduced and validated for specific periods. The classifiers used for the study have been deduced and validated by employing electronic health records (EHR), which were silent initially but alerted the clinical personnel concerning the sepsis prediction. For training the classification system, the chosen patients should have had ICD and the latest codes concerning extreme sepsis or septic shock. Moreover, the patients should have had positive blood culture during their interaction with the hospital, where there were indications of either systolic blood pressure (SBP) or lactate levels. The classification algorithms demonstrated a 93.84%, 93.22%, 95.25% accuracy, sensitivity and specificity respectively. The pattern used for clinical detection, in the context of the alerting system, led to a small but statistically significant increase in IV usage and lab tests. The values used for the alerting system were found to have no statistically significant difference in the context of different ICU wards since data from the laboratory tests serve as the primary early indicator of septic shock by confirming the presence of toxins.

INDEX TERMS Sepsis estimation, machine learning, deep learning, features optimization, clinical detection modeling.

I. INTRODUCTION

Considering medical syndromes, sepsis is among the most widely occurring syndromes in the country. Sepsis was previously categorised using a three-tier progression, namely sepsis, severe sepsis, and septic shock. Recently, the definition has been revised to a two-tier progression, namely sepsis (which includes severe sepsis), and septic shock [1]. Sepsis may have a life-threatening outcome, and sepsis-related mortality rate ranges between 25% to 40%, as understood from the latest literature [1], [2]. Sepsis is responsible for an economic burden of about $24 billion on the healthcare system in the U.S. [3]. Prompt diagnosis and rapid intervention may prevent sepsis from progressing to septic shock, thereby leading to higher patient survival and decrease stay at the hospital [4]. Sepsis comprises organ dysfunction arising from an infection-triggered systemic immune response, thereby causing inflammation. Nevertheless, the infection source and patient response can be very different across individuals, thereby creating challenges in the swift detection of sepsis. Hence, medical research has recently focused on automatic inpatient surveillance to facilitate the swift detection of sepsis. Health information is widely digitised using electronic health records (EHRs), thereby facilitating the development of automated systems that help with clinical decision making and prediction, which, in turn, lead to enhanced surveillance and intervention for complex cases [5]. Such information systems may offer suggestions and alert healthcare personnel by transforming electronic health records into a clinically accessible form, thereby increasing the quality of healthcare [6]. In the context of sepsis diagnosis, rules-based systems are presently the gold standard where alerts, suggestions, and minimal predictive information is provided [7]. Even for expert doctors, providing precise and timely sepsis diagnosis is challenging since there may be sepsis-related symptoms being caused by comorbidities [8], [9]. In a clinical scenario, the Systemic Inflammatory Response Syndrome (SIRS) may be used as the basis for establishing the rules for sepsis-specific scoring [10]. Additionally, the Modified Early Warning Score (MEWS) [11], [12] and the Sequential Organ Failure Assessment (SOFA) score [13] are other systems for rule-based criteria. Numerous studies have

The associate editor coordinating the review of this manuscript and approving it for publication was Yudong Zhang.

incorporated such scoring systems into EHR systems. Such tools may have appropriate sensitivity; however, they usually lack specificity, which indicates that they are not developed for sepsis prediction and progression. Moreover, rule-based scoring systems may not account for the differences in infection source and patient diversity. Table 1 specifies the SIRS criteria used for sepsis assessment. In the context of this paper, Section 2 discusses the significant works pertaining to Sepsis Estimation.

TABLE 1. Units for magnetic properties.

Parameter	Value
Heart rate (HR)	>92 beats/min
Body core temperature (BCT)	>38°C or 36°C
Breathing rate (BR)	22 breaths/min
alveolar carbon dioxide tension (PaCO2)	<32 mm Hg
Leukocyte cell count	12 × 109 cells/L

II. BACKGROUND (LITERATURE REVIEW)

There is literature supporting the use of machine learning (ML) models for early sepsis detection. These systems could be trained using individual Electronic Health Records (EHRs) [14]–[17]. It has been found the ML-based sepsis prediction models have significantly higher predictive capabilities compared to the score-based early warning systems like the National Early Warning System (NEWS) [18]–[20]. Shimabukuro *et al.*, illustrated numerous advantages of using an ML-based classification system for sepsis detection. The authors used a randomised clinical trial (RCT), and the results indicated that in-hospital mortality reduced to 13.5% (p=0.018), while the average stay at the hospital reduced from 13 days to 10 days (p=0.042). These factors were employed for sepsis prediction and complication analyses [21]. Nevertheless, present studies using ML have their drawbacks. Most of the studies have the ML models designed for a limited number of clinical indicators like vitals, which must necessarily be collected before ML-based predictions could be employed in clinical settings. While Intensive Care Units (ICU) or emergency departments often record vitals, it may not necessarily be the case for other medical departments [22]. Recently, convolutional neural networks (CNN), which facilitate deep learning (DL), have received a lot of attention in the context of pattern recognition and computer vision and as a methodology for artificial intelligence (AI). Machine learning is widely implemented using the neural network framework. In the context of deep learning, several layers form the neural network. Convolutional layers are those where the data passes through numerous filters, which facilitates effective pattern recognition. Traditional machine learning algorithms first extract the features and then use them for learning. In contrast, a convolutional neural network permits the use of the image during learning [23].

The classification process gets restricted because of the commonalities between the differences in perspective in the context of other classes and their own. There is a vast body of research that assessed the performance of typical landmark CNN classifiers. When used for classifying massive or fine-grained medical datasets, these classifiers produced excellent scores as compared to the present best fine-grained classifiers [24].

Behavioural activities like these may also be classified using Local Field Potential (LFP) output from the subthalamic nucleus (STN) [34]. Nevertheless, in the context of intensive care units (ICUs), predicting interventions in real-time is a challenge. A research team compared convolutional neural networks (CNN) and long short-term memory (LSTM) networks in the context of predicting five intervention activities, namely non-invasive ventilation, invasive ventilation, colloid boluses, crystalloid boluses, and vasopressors, which are the most commonly used medical interventions inside an intensive care ward to handle septic shock [25], [26].

There are extensive data registration protocols applicable to the units where prediction models are deployed; hence, the knowledge of the deployment and application potential of such models remains limited. Additionally, model evaluation typically compromises the reporting of only the receiver operating characteristic (ROC) plot and the area under the ROC curve (AUROC). There have been claims suggesting that AUROC only measures the predictive ability of the system, where prevalence is not considered and, therefore, it does not measure expected clinical effectiveness [26]. If the AUROC is used with data sets having a significant positive-negative sample imbalance, which is commonly observed in health sciences, misleading results could be generated. Moreover, most of the studies are assessed using ROC curves at a specified time from the onset of sepsis. In practical application, patient arrival at the hospital should be the ideal trigger point for beginning evaluation, while the algorithmic inference should be used several times after that [27].

This paper presents a sepsis estimation framework based on deep learning. The data set employed in the framework comprises several variables, which are applicable to both inside- and outside-ICU scenarios. Clinical utility is emphasised and is assessed in the context of straightforward concepts concerning accuracy. The suggested scheme provides a way to handle various data types as well as recording periods for the evaluation and estimation process. The validation processes are achieved with additional machine learning interpretations. Figure 1 depicts the proposed scheme. The core deep learning network used is CNN architecture [28], [29] of Long short-term memory (LSTM) composed of 34 input layers, 100 hidden layers with maximum number of batches of 18 and with two types of classes. The rectified linear unit (ReLU) applied for folding process of data sequence.

III. MATERIALS AND METHODS

A. GENERAL DESCRIPTION

The general scheme block diagram is described in figure 2. The gathered data are fed first to pre-processing stage, where it is an important issue to prepare data for using in next

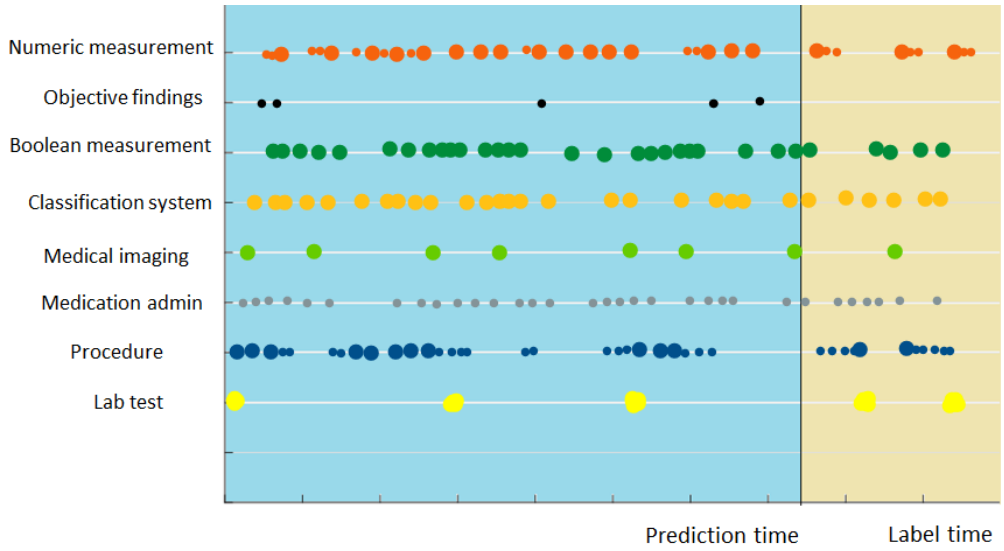

FIGURE 1. The proposed formulation parameters of clinical procedures, vital signals and objective findings for sepsis detection.

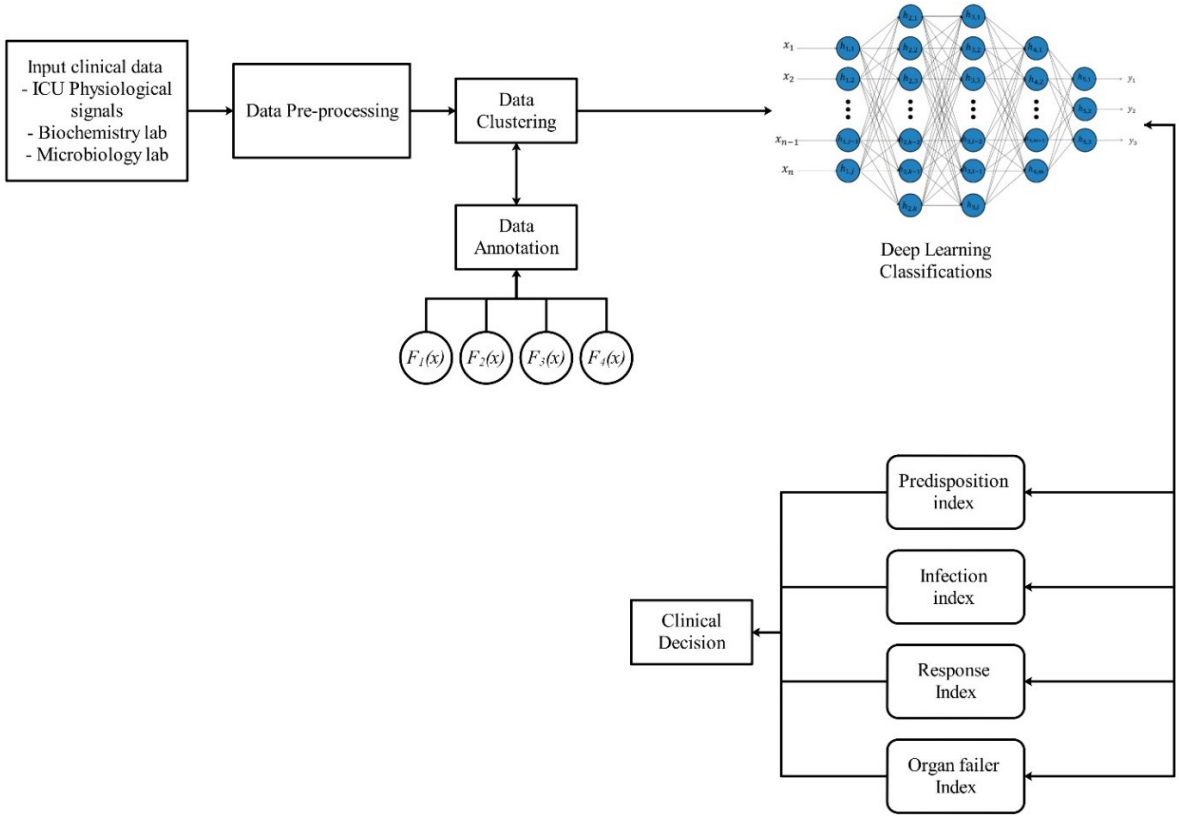

FIGURE 2. The proposed Sepsis estimation scheme.

stages. The second stage is the data clustering. In this stage, the required data are clustered and annotated according to their original orientation. This action will help to improve the diagnosis quality and reduce the unnecessary further process as well. Then the intelligent classification section is the next step. The outcome from intelligent section will guide the physician to estimate the patient's condition through groups of certain criteria named her ''Index''. By identifying each index, we can recognise the early sepsis cases then reduce the mortality.

B. THE DATA SOURCE

Three hospitals were chosen for data collection pertaining to ICU patients. In this context, data obtained from two hospitals shall be made available to the public, while the remaining data set would be restricted and employed for scoring. Patient data will be stored in the single pipe-delimited text format. The rows indicate hourly data, while it is decided to allot the same header to every file. Moreover, as specified in tables 2, 3, and 4, patient covariates comprise laboratory values, vitals, and demographic data. Note that the data which listed in table 4 has been excluded where it provides a general and uncorrelated interpretation data that may mislead the artificial learners [30].

TABLE 2. The vital signs and signals used in sepsis prediction as part of electronic health record for patients inside ICU wards.

Sign	Description
HR (HR)	Heart rate (beats per minute)
O2Sat	Pulse oximetry (%)
Temp	Temperature (°C)
SBP	Systolic BP (mm Hg)
MAP	Mean arterial pressure (mm Hg)
DBP	Diastolic BP (mm Hg)
Resp	Respiration rate (breaths per minute)
EtCO2	End tidal carbon dioxide (mm Hg)

TABLE 3. The biochemistry laboratory Values.

Sign	Description
Base Excess	Measure of excess bicarbonate (mmol/L)
HCO3	Bicarbonate (mmol/L)
FiO2	Fraction of inspired oxygen (%)
pH	N/A
PaCO2	Partial pressure of carbon dioxide from arterial blood (mm Hg)
SaO2	Oxygen saturation from arterial blood (%)
AST	Aspartate transaminase (IU/L)
BUN	Blood urea nitrogen (mg/dL)
Alkalinephos	Alkaline phosphatase (IU/L)
Calcium	(mg/dL)
Chloride	(mmol/L)
Creatinine	(mg/dL)
Bilirubin_direct	Bilirubin direct (mg/dL)
Glucose	Serum glucose (mg/dL)
Lactate	Lactic acid (mg/dL)
Magnesium	(mmol/dL)
Phosphate	(mg/dL)
Potassium	(mmol/L)
Bilirubin_total	Total bilirubin (mg/dL)
TroponinI	Troponin I (ng/mL)
Hct	Hematocrit (%)
Hgb	Hemoglobin (g/dL)
PTT	partial thromboplastin time (seconds)
WBC	Leukocyte count (count*10$\hat{3}$/µL)
Fibrinogen	(mg/dL)
Platelets	(count*10$\hat{3}$/μL)

TABLE 4. The anthropomorphics and demographics data for embedded in the EHR (EMR).

Sign	Description
Age	Years (100 for patients 90 or above)
Gender	Female (0) or Male (1)
Unit1	Administrative identifier for ICU unit (MICU)
Unit2	Administrative identifier for ICU unit (SICU)
Hospital Admission Time	Hours between hospital admit and ICU admit
ICULOS	ICU length-of-stay (hours since ICU admit)

C. DEFINITION OF EARLY SEPSIS DETECTION IN THE STUDY TARGET

After including clinical data obtained from hospital admission records, every admission was subjected to binary classification, where a sepsis positive or negative class was assigned. An individual is considered to have sepsis based on a doctor's evaluation. If an internal medicine physician has reasonable cause to suspect systemic infection regardless of bacteraemia, and the patient's EHR is updated with a sepsis event using the SIRS criteria, the dataset concerning the patient is updated with a definite sepsis diagnosis. The classification provides information on whether the patients meets the gold standard in the context of sepsis, as defined, with consensus, in 2001. Which means suspected infection and the indication of two or more SIRS criteria. With formulated parameters involved in the sepsis detection using ML methods, this primarily formulation shaped in a form of two intervals (prediction time of sepsis events) and (labelling time of sepsis events). Labelling interval in this context lasting for about one third of the clinical time intervention, as this time the classification system finalising annotation procedure for the end-results of medical assessment and transfer final knowledge base to the user for display and human interpretation. These tabulated numbers will be used as the basis for the detectability of sepsis onset and the early detection paradigm. A widely known sepsis identification model is being reused from existing literature. This model has demonstrated promising results in a randomised study and, therefore, may be used as the basis for the ML classification technique used in clinical scenarios.

D. DATA PRE-PROCESSING

The sample used in this study comprised 40,336 individuals. For each subject, the data included demographics, vital signs, laboratory values, onset time of sepsis, and sepsis label. We evaluated several approaches to pre-process the data. In particular, the prevalence of missing values and a tailed sequence-length distribution were the primary issues that affected the design of our pre-processing pipeline. The philosophy behind the PREP processing pipeline is to operate the pre-processing steps needed to normalize data intake hooked on a form that is effective for wide application range,

FIGURE 3. illustrates the pre-processing steps used in this study.

while maintaining as much of the signal content as possible. Unfortunately, in the context of development, there might be some incompatibility between standardisation and specialisation. Moreover, there could be complicated effects of the precise pre-processing requirements like up/down sampling, filtering on the downstream components [31]. In contrast, collections having unprocessed and nonstandard clinical data may be challenging for the same reason. Development on a massive scale fulfills the crucial objectives of testing the classification algorithm's robustness and contrasting neurological observations across tests and individuals. Base data sets must be well documented and ready for analysis for such comparisons to begin. A necessary stage in the context of automated data processing at massive scales is the recognition and elimination of poor channels because algorithms may face several challenges processing poor signals.

Figure 3 illustrates the pre-processing steps used in this study.

Data normalisation, also referred to as standardisation, is the process of scaling and adjusting data without making any changes to the nature or fundamentals of the data points. In the context of machine learning, pre-processing of data is a commonly used step. The primary objective of data normalisation is to transform the values of the data set to fit a standard scale without disturbing the differences between data points. Commonly, new boundary definitions are used (most common being (0, 1), and (−1, 1)), and the data is transformed as required. Pre-processing using this technique is helpful for distance- (e.g., K-means, or KNN) or neural network-based algorithms. The primary attribute is normalised on the Z-scale using the values of the mean and standard deviation. The normalised value Ui for an initial value Vi for attribute A is specified as:

$$U_i = \frac{V_i - Avg(A)}{std(A)}. \tag{1}$$

where $Avg(A)$ and $Std(A)$ respectively denote the average and standard deviation for attribute A.

IV. MODELLING OF CLINICAL DETECTION

All learning performed in this study was done using RNN-LSTM, SVM and adaptive CNN platform as a classifier [32], [33]. These ML classifiers isolated the output from several ''vulnerable'' ML learners, which would themselves have been insufficient to address the required learning objectives and building a reliable learning system. The weak learners comprise decision trees that are formed by dividing the feature space iteratively. Thresholds are determined where the feature classification leads to the maximum decrease in entropy, thereby enhancing information classification within the created groups. The final classifier is employed to conduct the necessary branching checks, wherein the tree is traversed until a leaf node (and the associated risk score) is reached. Risk scores from the trees are combined to generate the overall risk score [34], [35]. The training mechanisms used to train the classifiers comprise every feature set, which includes the vital signs along with the physiological indicators (heart rate (HR), systolic and diastolic blood pressure (SBP/DBP), respiratory rate (RR), peripheral oxygen saturation (SpO2), and temperature) (refer to Table 2), the biochemistry lab values (refer to Table 3), the demographic and anthropomorphic aspects (refer to Table 4). As specified previously, measurements of all these parameters should be obtained at a point in time so that these are appropriate for consideration in statistical analyses [36]–[38]. Moreover, the values for white blood cell count (WBC), platelet count (plt), and the Glasgow Coma Score (GCS) was also considered, if available. The gold standard indicators and the associated feature indicators were fed into arrays that would serve as inputs for training and testing processes [39]. The constructed features having well-defined aggregate information and protected ordering information is retained. Five different features represent every set of seven vital sign indicators to constitute the average for the current hour, the two preceding hours, and the drift (transition detection phase) among two consecutive hours.

V. CLINICAL PROCEDURE FOR DETECTION INDEX

The running scores provided by the systemic inflammatory response syndrome (SIRS) and the logistic organ dysfunction syndrome (LODS) were used to benchmark the sepsis prediction algorithms and their effectiveness obtained by employing the high-profile intelligent ML classifier. In the context of these experiments, the patients were in the age bracket of 2 to 17 years, and these patients were treated as one population set. Four ''folds'' of similar size were created by splitting this set to facilitate four-fold cross-validation (CV) [40], [41]. The cross-validation technique can handle performance generalisation and associated variability, along with the ability to handle hourly SIRS and LODS scores. In the context of the original clinical data set, the lab had specified the results as normal or abnormal. LODS sub-scores used an encoding of 0 points for a normal result and 1 point for an abnormal result. Several metrics measuring classifier performance (LODS and SIRS) were computed for the test folds [42]. Single tail paired T-tests were performed to assesses statistical significance, where every pair comprised the performance of

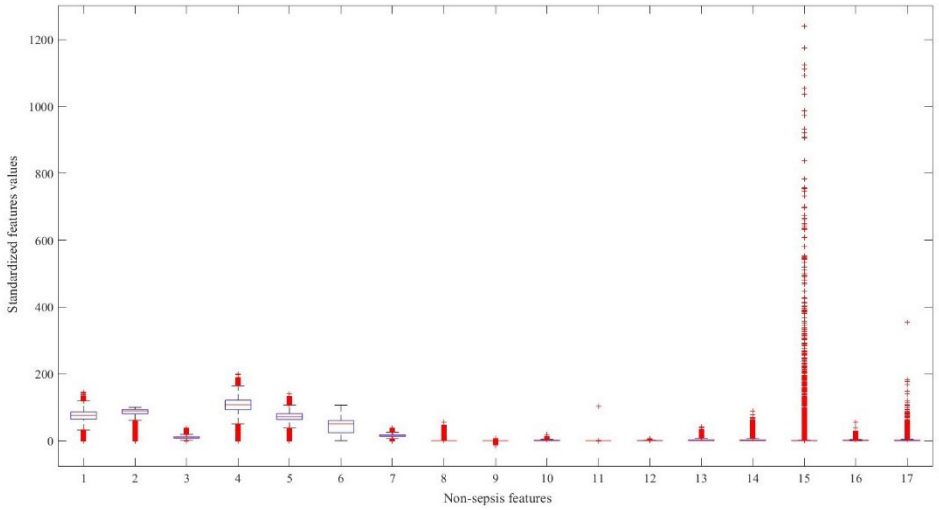

FIGURE 4. The 1-17 features of non-sepsis cases.

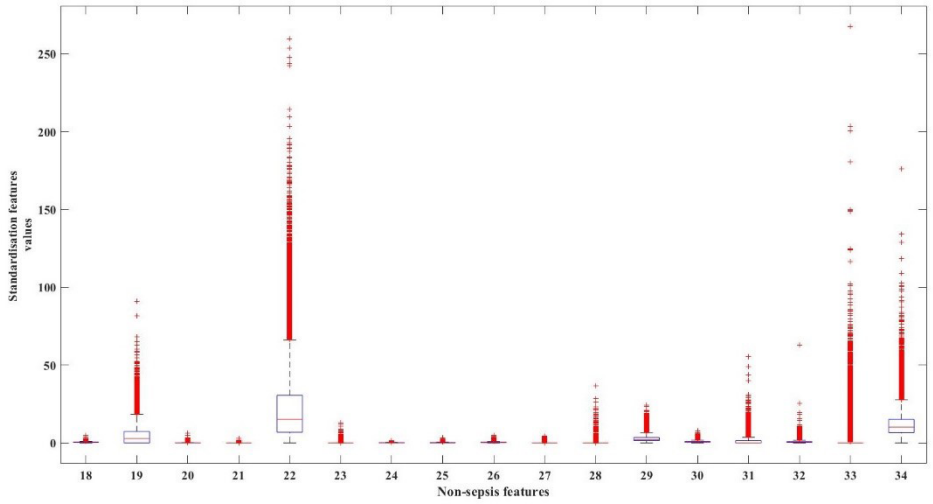

FIGURE 5. The 18-34 non-sepsis features.

two different classifiers used on a specific test fold. There were six samples for the paired T-test set since there is a comparison between six AUC/ROC scores. The test set on which the tests were performed had the significance threshold set at p=0.05 for all assessments [43], [44].

VI. RESULTS AND DISCUSSION

The requirement of data set standardisation commonly restricts machine learning estimator (MLE) implementation. The estimators may have poor performance if the features are not approximately normally distributed, which is Gaussian distribution with the mean and standard deviations being zero and one, respectively. For example, several elements of the objective function in the context of a learning technique are assumed to have the features having their centre at zero and similar variance. If variance for one feature is more substantial than its peers by several orders of magnitude, such a feature may have undue influence on the objective function and create problems for the estimator to learn appropriately using the other features. Figures 4 and 5 depict the non-sepsis Z-score normalisation (Standardised) gained features, while figures 6 and 7 depict the sepsis normalised features. Because of using many features, it divided into two figures for more clear vision, where first division holds features 1 to 17, while the second division holds the remaining features 18 to 34

The ML implementation comprises 34 features of which, seven values were selected from the six data quantification channels. The physiochemical prediction model is designed using decision-tree-based SVM classifiers. As specified in similar research [10], the gradient boosting model has every tree split for a maximum of six times, while tree aggregation

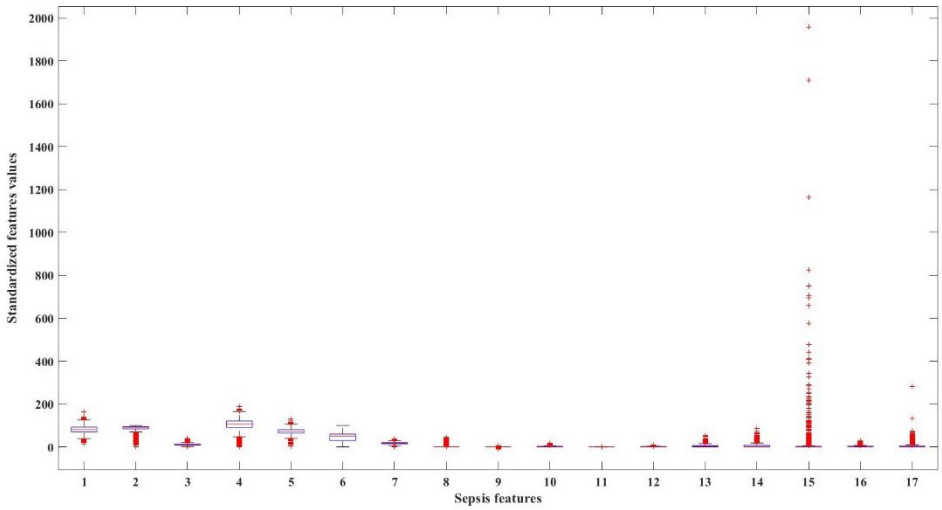

FIGURE 6. The 1-17 Sepsis features.

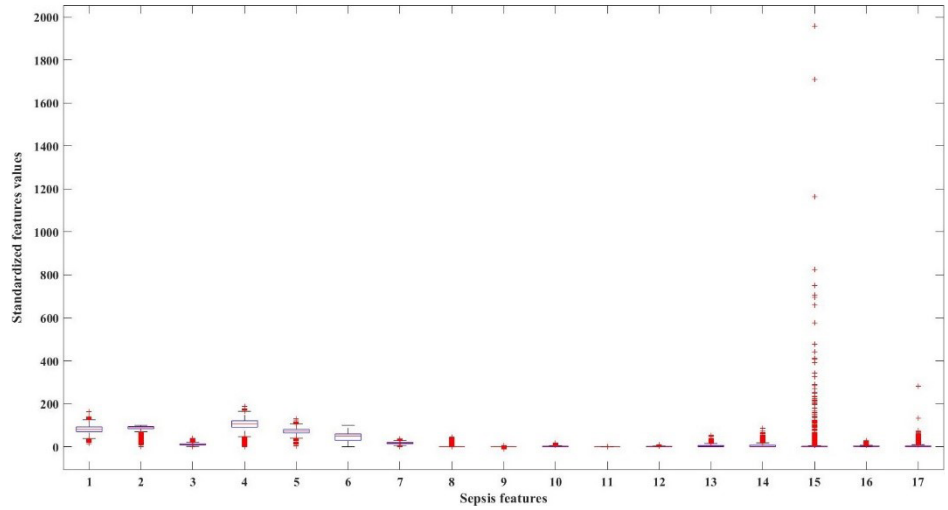

FIGURE 7. The 18-34 Sepsis features.

is limited to a thousand trees for predicting the risk values. As clear insight to the results obtained via several running (17-34) level of training we found that the sepsis features can be detected easily from the clinical datasets provided in the context of standardised feature values which can be tuned to be in the range of (0-300 iteration) for 17 feature in the range of (0-2000 iteration) for 34 feature regime as shown in figures 8 and 9 respectively. These results come with different projection on the main feature set classification, as part of recursive training paradigm used in the pre-processing stage. These classified features incorporated in sequence unfolding scheme to validate the result of classification. Using such type of validation will contain different classification uncertainties during learning process.

For specific classification of sepsis feature, we discriminate four different classes used in the prediction process of sepsis onset and related complication, these feature can be formulated as ($\theta 1$, for highly classified feature, and $\theta 2$ for low classified one) which had been attained from 2000 iteration training epoch) as shown in figures 4 and 5 respectively. By contrast, we achieve to formulate two other feature set ($\theta 3$, low classified feature and $\theta 4$, for high classified one) based on 300 training iteration. These augmented features can be used in successive steps for retraining the ML estimator and associated ACNN networks.

Features are the main summary of the four indexes used in this study (predisposition, infection, response, and organ failure index). Four of them will be intermingled to form the final ''Clinical detection'' which signals for the sepsis alerting system implemented in central monitoring platform inside ICU wards.

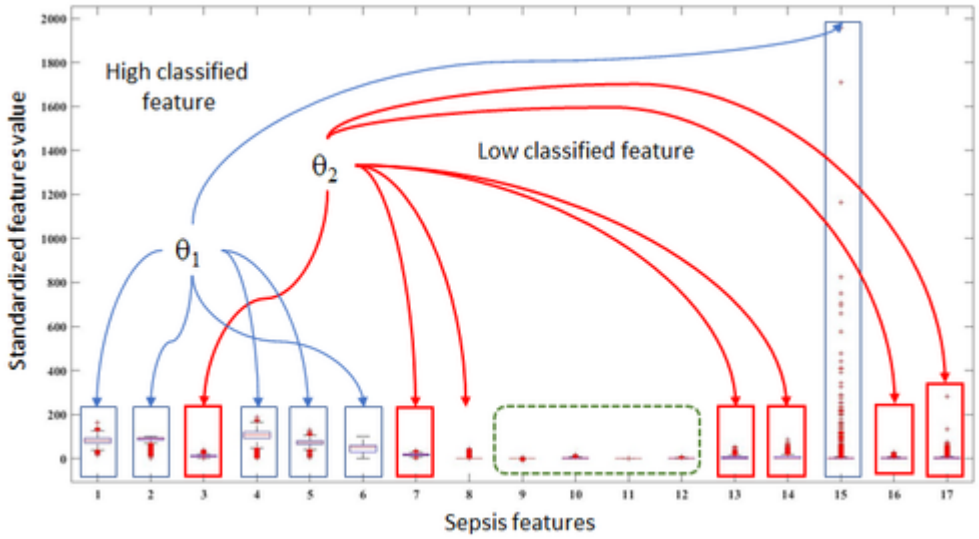

FIGURE 8. Classification of sepsis features, by using 2000 iterations detected form the HER clinical data annotated and processed using ACNN classifier algorithm proposed in this study, the theta class for features with (high features classified (θ1)) and with (low features classified (θ2).

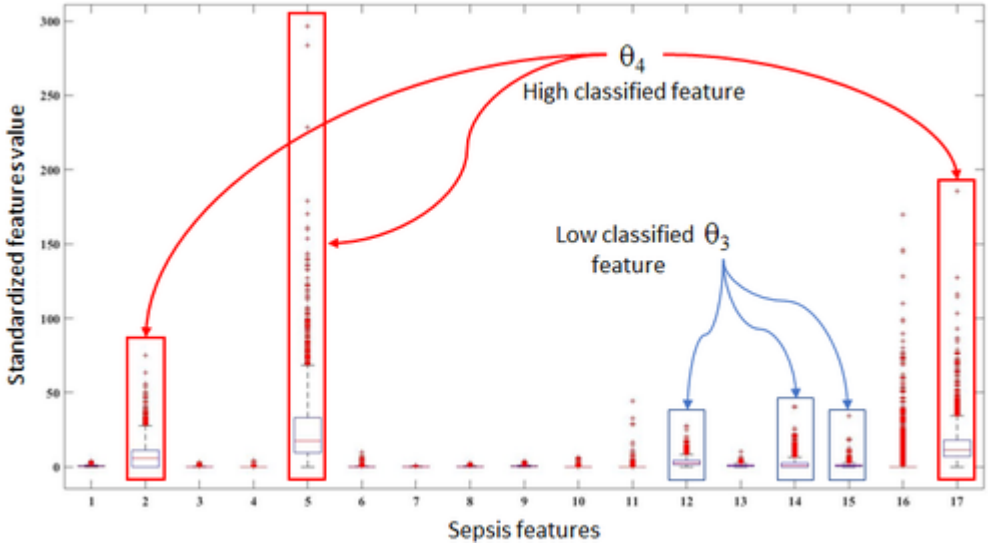

FIGURE 9. Classification of sepsis features, by using 300 iterations detected form the HER clinical data annotated and processed using ACNN classifier algorithm proposed in this study, the theta class for features with (low features classified (θ3)) and with (high features classified (θ4)).

The risk-score derived from the clinical detection paradigm or as we introduced before as (algorithmic sepsis predictor) can be further processed to be compatible with SIRS and LODS criteria as part of the septic shock classification and segmentation.

The predictive capability of the RNN, ACNN, and SVM models was evaluated using hourly training and testing slots that began at the onset of sepsis and continued through the six hours before sepsis symptoms were apparent.

In addition, the ACNN classifier, two more intelligent classifiers are used to validate the proposed scheme for validation process LSTM-RNN and SVM classifiers.

TABLE 5. The scheme obtained results for training phase.

Method	Results (accuracy %)
RNN-LSTM	92.72%
SVM-quadratic kernel	78.00%
Adaptive CNN	**93.84%**

The Receiver Operating Characteristics (ROC) values for the SVM are depicted in Figures 10(a) and (b), which highlight the sensitivity-specificity trade-off. Specificity is the fraction of sepsis-negative individuals classified as having severe sepsis, while sensitivity denotes the fraction of individuals having severe sepsis and were classified the same.

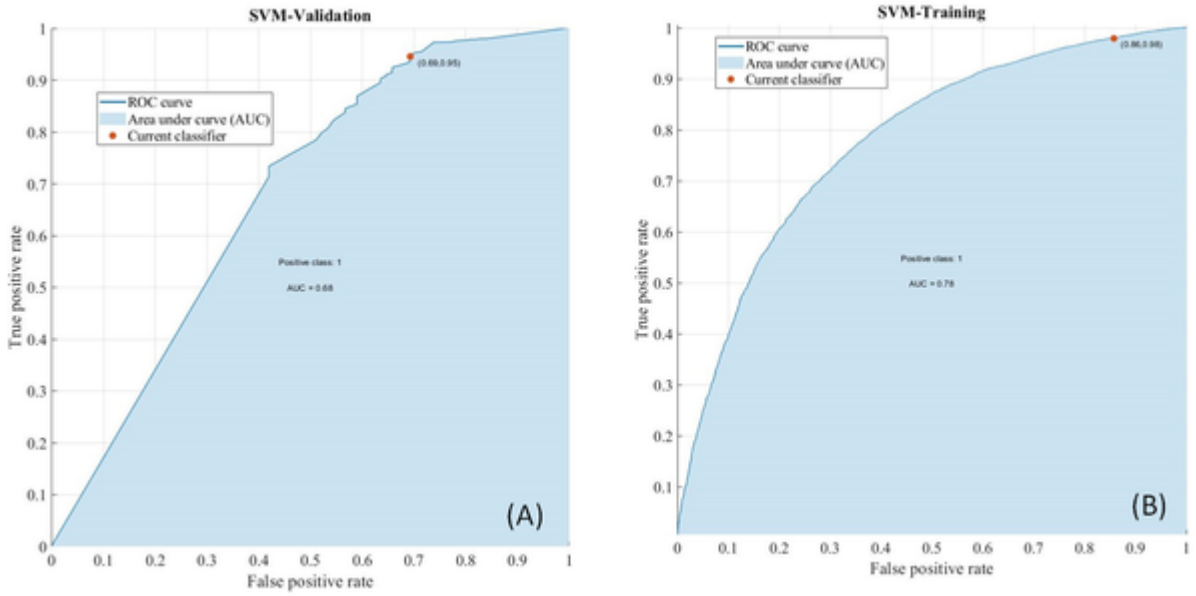

FIGURE 10. (A) ROC curve depicting the average across the four folds for LODS and SIRS algorithm at sepsis onset, having an AUC of 0.68. (B) ROC curve depicting the average across the four folds for LODS and SIRS with 4 hrs sepsis symptoms (pre-onset) having an AUC of 0.78.

TABLE 6. The scheme obtained results for testing phase.

Method	Results (accuracy %)
RNN-LSTM	91.10%
SVM-quadratic kernel	68.00%
Adaptive CNN	**93.18%**

TABLE 7. Accuracy comparison between several state-of-the-art classification techniques in sepsis detection and computer aided diagnosis.

Author	accuracy %	Sensitivity%	Specificity%
Faisal [37] CNN	78.00 %	77.94%	68.00%
Kam [38] (Insight model)	92.90%	91.40%	94.40%
Lauritsen [39] (CNN-LSTM)	85.60%	67.00%	63.00%
Sarafrazi1 [40] (CNN/tree model)	82.00%	-	-
Bartona [41] (SVM)	88.00%	80.00%	78.00%
Xin [[42] (CNN/RANN)	92.70%	-	-
Jau-Woei Perng [43] CNN+SoftMax	87.01%	-	-
Wang [44] RF-CFOA-KELM	81.60%	89.75%	65.77%
Authors proposed method	**93.84%**	**93.22%**	**95.25%**

Compared to the characteristics of the LODS and SIRS curves, the ROC curve for the ML-based predictor is superior with a higher area under the curve, thereby highlighting enhanced accuracy in (B) as compared to (A). It is clear from Table 5 and Table 6 that, the better classifier that works under the condition of this scheme is ACNN. For the ACNN classifier, the gained sensitivity and specificity are 93.22%, and 95.25% respectively. The sensitivity and specificity results refer to the low false negative and false positive rate which reflect the quality of the intelligent classifier.

Table 7 depicts the benchmarking with other related studies that achieved regarding sepsis detection. It is noted that the proposed scheme shows an outstanding accuracy, sensitivity, and specificity rate that leads to consider it and use it in hospitals and medical clinics as well.

The study is limited in the aspect that the proposed technique cannot be used on the MIMIC-III database because the two cohorts comprise completely varying patient data. The MIMIC-III database comprises critical care ward data at tertiary health care centres, as compared to other databases, which have a mix of data specific to urban and rural individuals from several centres.

VII. CONCLUSION AND FUTURE WORK

In the context of this study, the ML framework that was formulated and tested is capable of prediction and estimation. It is highly specific but lacks sensitivity, which indicates the model's ability to predict severe sepsis correctly, the onset of shock, and complications. The clinical measures were not impacted to a great extent by the predictive warnings issued by the algorithms. Further augmentation may be done by optimising the design, refining the algorithm, and expressing a clinical perspective about the implementation.

Whereas the other studies comprise the development and implementation of small-scale predictive alerting systems using ML-based estimation algorithms, this study is the first that reports clinicians' perspective of these tools for early detection of sepsis. The results of the study highlight the probable roadblocks to the positive reception of ACNN, which include 1) patient clinical stability when alerted, 2) lack of

algorithmic transparency in the context of clinical aspects, 3) confidence in doctors' decision, and 4) ambiguity in generating alerts specific for high-risk individuals for whom health deterioration has not begun. These aspects could be used generally for ML-based alerting systems.

The clinical relevance and accuracy of the system could be viewed as inferior, given the clinical stability of the patients when the alerts were issued. ACNN is an intelligent predictor that had a median trigger alert time of 4 hours, while, for some cases, the alerts were sent many days in advance of the onset of severe sepsis or associated shock. It is possible that clinical personnel understood the ACNN as a traditional tool and may have disregarded the alerts as being erroneous or not helpful when there were no symptoms of clinical deterioration. Potentially, the need for instant bedside assessment could have created this incorrect perspective that the alerts were issued for rapid response in cases of decompensation. While implementation-based promotions could address such incorrect perceptions to some extent, an appropriate lead time in the context of predictive alerting is not yet clear.

Doctors and healthcare personnel could find it challenging to rely on alerts and predictions produced by algorithmic systems. ML algorithms are also referred to as ''black-box models'' since the variables affecting prediction may not necessarily be explicitly at the user's disposal. Given that ML algorithms can comfortably accommodate several hundred variables, the aspects contributing to a forecast could be unwieldy to the point where they may not be significant for clinicians. Moreover, since the ML techniques consider essential parameters that may not have previously associated outcomes, predictions provided using such parameters could be clinically less inherent. Clarity in the design of ML-based algorithms is challenging, but, if achieved, any alerts issued by such systems could prompt risk assessment by a clinician.

REFERENCES

[1] A. Rhodes, L. Evans, and R. P. Dellinger, ''Surviving sepsis campaign: International guidelines for management of sepsis and septic shock: 2016,'' *Intensive Care Med.*, vol. 43, no. 3, pp. 304–377, 2017.

[2] The ARISE Investigators and the ANZICS Clinical Trials Group, ''Goal-directed resuscitation for patients with early septic shock,'' *New England J. Med.*, vol. 371, no. 16, pp. 1496–1506, Oct. 2014.

[3] E. K. Stevenson, A. R. Rubenstein, G. T. Radin, R. S. Wiener, and A. J. Walkey, ''Two decades of mortality trends among patients with severe sepsis: A comparative meta-analysis,'' *Crit. care Med.*, vol. 42, no. 3, p. 625, 2014.

[4] A. F. Shorr, S. T. Micek, W. L. Jackson, Jr., and M. H. Kollef, ''Economic implications of an evidence-based sepsis protocol: Can we improve outcomes and lower costs?'' *Crit. Care Med.*, vol. 35, no. 5, pp. 1257–1262, 2007.

[5] O. A. Uthman, C. I. Okwundu, C. S. Wiysonge, T. Young, and A. Clarke, ''Citation classics in systematic reviews and meta-analyses: Who wrote the top 100 most cited articles?'' *PLoS ONE*, vol. 8, no. 10, Oct. 2013, Art. no. e78517.

[6] S. C. Smith, ''AHA/ACCF secondary prevention and risk reduction therapy for patients with coronary and other atherosclerotic vascular disease: 2011 update: A guideline from the American heart association and American college of cardiology foundation endorsed by the world heart federation and the preventive cardiovascular nurses association,'' *J. Amer. College Cardiol.*, vol. 58, no. 23, pp. 2432–2446, 2011.

[7] L. A. van Vught, P. Klouwenberg, and C. Spitoni, ''Incidence, risk factors, and attributable mortality of secondary infections in the intensive care unit after admission for sepsis,'' *Jama*, vol. 315, no. 14, pp. 1469–1479, 2016.

[8] R. P. Dellinger, ''Incidence surviving sepsis campaign: International guidelines for management of severe sepsis and septic shock, 2012,'' *Intensive Care Med.*, vol. 39, no. 2, pp. 165–228, 2013.

[9] H. M. Giannini, J. C. Ginestra, C. Chivers, M. Draugelis, A. Hanish, W. D. Schweickert, B. D. Fuchs, L. Meadows, M. Lynch, P. J. Donnelly, K. Pavan, N. O. Fishman, C. W. Hanson, and C. A. Umscheid, ''A machine learning algorithm to predict severe sepsis and septic shock: Development, implementation, and impact on clinical practice,'' *Crit. Care Med.*, vol. 47, no. 11, pp. 1485–1492, Nov. 2019.

[10] O. A. Usman, A. A. Usman, and M. A. Ward, ''Comparison of SIRS, SOFA, and NEWS for the early identification of sepsis in the emergency department,'' *Amer. J. Emergency Med.*, vol. 37, no. 8, pp. 1490–1497, Aug. 2019.

[11] D. R. Prytherch, G. B. Smith, P. E. Schmidt, and P. I. Featherstone, ''ViEWS—Towards a national early warning score for detecting adult inpatient deterioration,'' *Resuscitation*, vol. 81, no. 8, pp. 932–937, 2010.

[12] A. F. Hussein, S. J. Hashim, A. F. A. Aziz, F. Z. Rokhani, and W. A. W. Adnan, ''Performance evaluation of time-frequency distributions for ECG signal analysis,'' *J. Med. Syst.*, vol. 42, no. 1, p. 15, Jan. 2018.

[13] E. P. Raith, A. A. Udy, M. Bailey, and S. McGloughlin, ''Prognostic accuracy of the SOFA score, SIRS criteria, and SOFA score for in-hospital mortality among adults with suspected infection admitted to the intensive care unit,'' *Jama*, vol. 317, no. 3, pp. 290–300, 2017.

[14] D. W. Shimabukuro, C. W. Barton, M. D. Feldman, S. J. Mataraso, and R. Das, ''Effect of a machine learning-based severe sepsis prediction algorithm on patient survival and hospital length of stay: A randomised clinical trial,'' *BMJ Open Respiratory Res.*, vol. 4, no. 1, Nov. 2017, Art. no. e000234.

[15] S. Nemati, A. Holder, F. Razmi, M. D. Stanley, G. D. Clifford, and T. G. Buchman, ''An interpretable machine learning model for accurate prediction of sepsis in the ICU,'' *Crit. Care Med.*, vol. 46, no. 4, pp. 547–553, Apr. 2018.

[16] Q. Mao, M. Jay, J. L. Hoffman, J. Calvert, C. Barton, D. Shimabukuro, L. Shieh, U. Chettipally, G. Fletcher, Y. Kerem, Y. Zhou, and R. Das, ''Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU,'' *BMJ Open*, vol. 8, no. 1, Jan. 2018, Art. no. e017833.

[17] J. Futoma, S. Hariharan, M. Sendak, N. Brajer, M. Clement, A. Bedoya, C. O'Brien, and K. Heller, ''An improved multi-output Gaussian process RNN with real-time validation for early sepsis detection,'' 2017, *arXiv:1708.05894*. [Online]. Available: http://arxiv.org/abs/1708.05894

[18] R. Challen, J. Denny, M. Pitt, L. Gompels, T. Edwards, and K. Tsaneva-Atanasova, ''Artificial intelligence, bias and clinical safety,'' *BMJ Qual. Saf.*, vol. 28, no. 3, pp. 231–237, Mar. 2019.

[19] E. Bloch, T. Rotem, J. Cohen, P. Singer, and Y. Aperstein, ''Machine learning models for analysis of vital signs dynamics: A case for sepsis onset prediction,'' *J. Healthcare Eng.*, vol. 2019, pp. 1–11, Nov. 2019.

[20] M. Schinkel, K. Paranjape, R. S. Nannan Panday, N. Skyttberg, and P. W. B. Nanayakkara, ''Clinical applications of artificial intelligence in sepsis: A narrative review,'' *Comput. Biol. Med.*, vol. 115, Dec. 2019, Art. no. 103488.

[21] B. S. dos Santos, M. T. A. Steiner, A. T. Fenerich, and R. H. P. Lima, ''Data mining and machine learning techniques applied to public health problems: A bibliometric analysis from 2009 to 2018,'' *Comput. Ind. Eng.*, vol. 138, Dec. 2019, Art. no. 106120.

[22] S. Halligan, D. G. Altman, and S. Mallett, ''Disadvantages of using the area under the receiver operating characteristic curve to assess imaging tests: A discussion and proposal for an alternative approach,'' *Eur. Radiol.*, vol. 25, no. 4, pp. 932–939, Apr. 2015.

[23] K. J. L. Bell, E. Beller, J. Sundstrom, K. McGeechan, A. Hayen, L. Irwig, B. Neal, and P. Glasziou, ''Ambulatory blood pressure adds little to framingham risk score for the primary prevention of cardiovascular disease in older men: Secondary analysis of observational study data,'' *BMJ Open*, vol. 4, no. 9, Sep. 2014, Art. no. e006044.

[24] T. Sainath, A.-R. Mohamed, B. B. Kingsbury, and K. Ramabhadran, ''Deep convolutional neural networks for LVCSR,'' in *Proc. Acoust., Speech Signal Process.*, May 2013, pp. 8614–8618.

[25] J. Suykens and J. Vandewalle, ''Least squares support vector machine classifiers,'' in *Proc. Acoust., Speech Signal Process.*, 1999, pp. 293–300.

[26] J. Wang, K. Feng, and J. Wu, ''SVM-based deep stacking networks,'' in *Proc. AAAI Conf. Artif. Intell.*, 2019, pp. 5273–5280.

[27] S. Anwar, N. Barnes, and L. Petersson, "A systematic evaluation: Fine-grained CNN vs. Traditional CNN classifiers," 2020, *arXiv:2003.11154*. [Online]. Available: http://arxiv.org/abs/2003.11154
[28] T. R. Mullen, C. A. E. Kothe, Y. M. Chi, A. Ojeda, T. Kerth, S. Makeig, T.-P. Jung, and G. Cauwenberghs, "Real-time neuroimaging and cognitive monitoring using wearable dry EEG," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 11, pp. 2553–2567, Nov. 2015.
[29] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: An overview and application in radiology," *Insights into Imag.*, vol. 9, no. 4, pp. 611–629, Aug. 2018.
[30] M. Reyna, S. Prajwal Shashikumar, B. Moody, P. Gu, A. Sharma, S. Nemati, and G. Clifford, "Early prediction of sepsis from clinical data: The PhysioNet/computing in cardiology challenge 2019," in *Proc. Comput. Cardiol. Conf. (CinC)*, Dec. 2019.
[31] C. Farabet, C. Couprie, L. Najman, and Y. LeCun, "Learning hierarchical features for scene labeling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1915–1929, Aug. 2013.
[32] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," 2014, *arXiv:1409.4842*. [Online]. Available: http://arxiv.org/abs/1409.4842
[33] Z.-R. Zhou, W.-W. Wang, Y. Li, K.-R. Jin, X.-Y. Wang, Z.-W. Wang, Y.-S. Chen, S.-J. Wang, J. Hu, H.-N. Zhang, P. Huang, G.-Z. Zhao, X.-X. Chen, B. Li, and T.-S. Zhang, "In-depth mining of clinical data: The construction of clinical prediction model with R," *Ann. Transl. Med.*, vol. 7, no. 23, p. 796, Dec. 2019.
[34] D. D. Bhuva, J. Cursons, G. K. Smyth, and M. J. Davis, "Differential co-expression-based detection of conditional relationships in transcriptional data: Comparative analysis and application to breast cancer," *Genome Biol.*, vol. 20, no. 1, pp. 1–21, Dec. 2019.
[35] C. A. Gonzalez-Amarillo, J. C. Corrales-Munoz, M. A. Mendoza-Moreno, A. M. Gonzalez Amarillo, A. F. Hussein, N. Arunkumar, and G. Ramirez-Gonzalez, "An IoT-based traceability system for greenhouse seedling crops," *IEEE Access*, vol. 6, pp. 67528–67535, 2018.
[36] M. A. Reyna, C. Josef, S. Seyedi, and R. Jeter, "Early prediction of sepsis from clinical data: The physionet/computing in cardiology challenge 2019," in *Proc. Comput. Cardiol.*, Sep. 2019, p. 1.
[37] M. Faisal, A. Scally, D. Richardson, K. Beatson, R. Howes, K. Speed, and M. A. Mohammed, "Development and external validation of an automated computer-aided risk score for predicting sepsis in emergency medical admissions using the patient's first electronically recorded vital signs and blood test results," *Crit. Care Med.*, vol. 46, no. 4, pp. 612–618, Apr. 2018.
[38] H. J. Kam and H. Y. Kim, "Computers in biology and medicine," *Crit. Care Med.*, vol. 89, pp. 248–255, 2017.
[39] S. M. Lauritsen, M. E. Kalør, E. L. Kongsgaard, K. M. Lauritsen, M. J. Jørgensen, J. Lange, and B. Thiesson, "Early detection of sepsis utilizing deep learning on electronic health record event sequences," *Artif. Intell. Med.*, vol. 104, Apr. 2020, Art. no. 101820.
[40] S. Sarafrazi, R. Choudhari, H. Mehta, C. Mehta, O. Japalaghi, K. Mehta, H. Han, and P. Francis-Lyon, "Cracking the sepsis code: Assessing time series nature of EHR data, and using deep learning for early sepsis prediction," in *Proc. Comput. Cardiol. Conf. (CinC)*, Singapore, Dec. 2019, pp. 1–4.
[41] C. Barton, U. Chettipally, Y. Zhou, Z. Jiang, A. Lynn-Palevsky, S. Le, J. Calvert, and R. Das, "Evaluation of a machine learning algorithm for up to 48-hour advance prediction of sepsis using six vital signs," *Comput. Biol. Med.*, vol. 109, pp. 79–84, Jun. 2019.
[42] X. Li, G. André Ng, and F. Schlindwein, "Convolutional and recurrent neural networks for early detection of sepsis using hourly physiological data from patients in intensive care unit," in *Proc. Comput. Cardiol. Conf. (CinC)*, Singapore, Dec. 2019, pp. 1–4.
[43] J.-W. Perng, I.-H. Kao, C.-T. Kung, S.-C. Hung, Y.-H. Lai, and C.-M. Su, "Mortality prediction of septic patients in the emergency department based on machine learning," *J. Clin. Med.*, vol. 8, no. 11, p. 1906, Nov. 2019.
[44] X. Wang, Z. Wang, J. Weng, C. Wen, H. Chen, and X. Wang, "A new effective machine learning framework for sepsis diagnosis," *IEEE Access*, vol. 6, pp. 48300–48310, 2018.

BILAL YASEEN AL-MUALEMI received the B.Sc. degree in information system from the Department of Computer Science, Sana'a University, Yemen, and the M.S. degree in information system from the Department of Computer Science, Arabic Academ, Yemen. He is currently pursuing the Ph.D. degree in deep learning. His research interests include image classifications and data analysis.

LU LU (Member, IEEE) received the Ph.D. degree from Xi'an Jiaotong University, in 1999. He is currently a Professor with the School of Computer Science and Engineering, South China university of Technology, China. His main research interests include software engineering, software testing, and software architecture design.