Multitask Deep Learning for Cost-Effective Prediction of Patient's Length of Stay and Readmission State Using Multimodal Physical Activity Sensory Data

Sajid Ali, Shaker El-Sappagh, Farman Ali, Muhammad Imran, *Member, IEEE*, and Tamer Abuhmed

***Abstract*—In a hospital, accurate and rapid mortality prediction of Length of Stay (LOS) is essential since it is one of the essential measures in treating patients with severe diseases. When predictions of patient mortality and readmission are combined, these models gain a new level of significance. Therefore, the most expensive components of patient care are LOS and readmission rates. Several studies have assessed readmission to the hospital as a single-task issue. The performance, robustness, and stability of the model increase when many correlated tasks are optimized. This study develops multimodal multitasking Long Short-Term Memory (LSTM) Deep Learning (DL) model that can predict both LOS and readmission for patients using multi-sensory data from 47 patients. Continuous sensory data is divided into eight sections, each of which is recorded for an hour. The time steps are constructed using a dual 10-second window-based technique, resulting in six steps per hour. The 30 statistical features are computed by transforming the sensory input into the resulting vector. The proposed multitasking model predicts 30-day readmission as a binary classification problem and LOS as a regression task by constructing discrete time-step data based on the length of physical activity during a hospital stay. The proposed model is compared to a random forest for a single-task problem (classification or regression) because typical machine learning algorithms are unable to handle the multitasking challenge. In addition, sensory data combined with other cost-effective modalities such as demographics, laboratory tests, and comorbidities to construct reliable models for personalized, cost-effective, and medically acceptable prediction. With a high accuracy of 94.84%, the proposed multitask multimodal DL model classifies the patient's readmission status and determines the patient's LOS in hospital with a minimal Mean Square Error (MSE) of 0.025 and Root Mean Square Error (RMSE) of 0.077, which is promising, effective, and trustworthy.**

***Index Terms*—Deep learning, healthcare predictive model, medical sensory data fusion, multimodal and multitasking, physical activity, and time-series data analysis.**

Manuscript received 30 March 2022; revised 5 July 2022 and 16 August 2022; accepted 23 August 2022. Date of publication 29 August 2022; date of current version 6 December 2022. This work was supported in part by the Ministry of Science and ICT, Korea through ICT Creative Consilience Program under Grant IITP-2021-2020-0-01821 supervised by the Institute for Information and Communications Technology Planning and Evaluation, and in part by the National Research Foundation of Korea funded by Korea Government under Grant 2021R1A2C1011198. *(Corresponding author: Tamer Abuhmed.)*

Sajid Ali is with the Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon-si 16419, South Korea (e-mail: sajidali@skku.edu).

Shaker El-Sappagh is with the College of Computing and Informatics, Sungkyunkwan University, Suwon-si 16419, South Korea, and also with the Faculty of Computer Science and Engineering, Galala University, Suez 435611, Egypt (e-mail: sh.elsappagh@inha.ac.kr).

Farman Ali is with the Department of Software, Sejong University, Seoul 05006, South Korea (e-mail: farmankanju@sejong.ac.kr).

Muhammad Imran is with the Institute of Innovation, Science and Sustainability, Federation University Australia, Ballarat, VIC 3350, Australia (e-mail: dr.m.imran@ieee.org).

Tamer Abuhmed is with the College of Computing and Informatics, Sungkyunkwan University, Suwon-si 16419, South Korea (e-mail: tamer@skku.edu).

Digital Object Identifier 10.1109/JBHI.2022.3202178

I. Introduction

HEALTH expenditure in many countries' accounts for a significant portion of their gross domestic product [1]. In many nations, increasing numbers of patients are causing government funding to deviate significantly from the cost of patient care. As a result, cost reduction is currently the most pressing issue in healthcare, particularly in light of the present scenario of reduced bed availability and rising costs [2]. Hospitalization and readmission of patients are the most expensive aspects of patient care; hence, healthcare administration is a priority. For example, inpatient hospitalization accounts for nearly a third (29%) of total healthcare expenditure in the United States. Length of Stay (LOS) is defined as the number of days a patient is hospitalized in a healthcare institution during one admission. It is calculated as the period from the patient's admission to discharge [3]. Patients, health personnel, and insurance organizations use the LOS as a critical indicator. It helps with hospital planning and management, as well as determining: (1) hospital resource requirements, (2) hospital mortality, (3) treatment costs, and (4) disease severity. Therefore, predicting LOS in healthcare facilities is crucial. However, it is a complex, poorly-structured, and dynamic problem because it is influenced by many factors such as the patient's demography, the complexity of treatment, comorbidities, laboratory tests, and other criteria unrelated to the disease [4]. The increasing use of social or community nursing care is an example of non-disease-related variables. Furthermore, the following factors may influence LOS: (1) organization of hospital management style; and (2) the patient readmission within the next 30 days is significantly associated with LOS. Predicting both measures greatly benefits patients, physicians, and hospital administration.

To determine both the cost of healthcare and the patient's experience, an accurate and optimal LOS prediction is a significant priority. Correctly estimating LOS has considerable advantages in critical care settings where expensive and complex procedures, such as acute disease management are performed [5]. A substantial amount of research has been conducted to determine the optimal attributes to improve the performance of the LOS prediction model. For example, LOS was strongly influenced by age, type of admission, and hospital type [6]. The best factors associated with LOS were found to be the hospital episode class, the physical services in which the patient is admitted, and the accompanying medical expertise [7]. On the other hand, these criteria are insufficient to predict LOS reliably [8]. Other parameters, such as nutritional state before admission and serum creatinine concentration, should be considered to establish a personalized and tailored LOS time prediction. According to [9], in patients treated with ventilators, suffering from hypomagnesemia, delirium, and malnutrition, the stay was longer, while age and gender are not significant determinants.

However, the continuous measurement of an objective assessment of physical activity over a longer period is possible thanks to advanced sensor technology. Wearable body accelerometers are an effective and cost-effective tool for assessing the health of a hospitalized patient [10]. Based on sensor data, Abeles et al., [11] conducted a survey to determine the relationship between LOS and the patient's level of physical activity. This is one of the first studies to explore whether there is a link between physical activity and LOS. Furthermore, there is a significant relationship between a patient's LOS and their readmission to the hospital [12]. Readmission history has been considered a single-task problem in several studies [13], [14]. However, joint optimization of many correlated tasks improves the model's performance, stability, and robustness [15]. Successful prediction of patient LOS and hospital readmission requires a multimodal multitasking Deep Learning (DL) model that can account for the variability of multivariate time-series data rather than existing static constrained algorithms in such a dynamic environment.

Schricker et al., [16] developed a novel wrist-worn Bosch sensor that can measure numerous features of a patient's physical activity and ambient conditions. They state that new algorithms are needed to improve prediction performance and understand the importance of physical activity in hospitalized patients. To the best of our knowledge, no research in the literature has explored Bosch sensory data to date. In this work, we proposed a multimodal Long Short-Term Memory (LSTM) model that, utilizing Bosch sensory data, can predict both LOS and readmission as a multitask. By formulating time-step data based on the length of physical activity, the proposed model predicts LOS as a regression task and 30-day readmission as a binary classification task. The use of sensory data based on other cost-effective modalities such as demographics, laboratory tests, and comorbidities can help develop reliable models for predicting tailored, cost-effective, and medically acceptable LOS measurements. The main contributions of the study are as follows:

- We proposed a multitasking LSTM-based DL model that can predict the patient's LOS as a regression task and readmission within 30 days as a binary classification task based on multimodal time series data. In this proposed architecture, we explore the role of the LSTM DL model in predicting two tasks based on heterogeneous time series data.
- The proposed model was optimized using the Bayesian optimizer against the real dataset of the patient's physical activity in the hospital with the new Bosch accelerometer sensor collected by the Robert-Bosch hospital in Stuttgart, Germany [16]. To the best of our knowledge, this novel time series dataset has not been used in the literature, and this study is the first to build a multitask DL model to shed light on the importance of a patient's physical activity in medical diagnosis.
- To assess the performance of the proposed model, we conducted several experiments including: (1) a single regression task for LOS prediction, (2) a single patient readmission prediction task, (3) a single LOS regression task after early fusion of sensory data with laboratory tests, comorbidities and demographics data, (4) a single readmission classification task after combining four modalities, (5) multitask modeling to jointly predict LOS and patient readmission. This assignment was based on an early fusion of the four modalities. Further, we explored the effect of applying the recursive feature elimination technique to feature selection.
- To further evaluate the efficiency of our model, we performed temporal spacial complexity analyses of the model using a combined dataset of sensory, demographic, and laboratory test time series data.
- The proposed model was compared with classical Machine Learning (ML) models, especially with a well-known set of random forests. Random forest was have been trained using statistical features computed from sensory time series data. These calculated statistical features were fused with demographics and laboratory tests.

Organization: The remainder of the paper is organized as follows. Section II discusses the related work on LOS and readmission. Section III discusses the introductory elements of the cell's internal architecture, the experimental environment, and evaluation matrices. Section IV explains in detail the proposed multimodal and multitasking model. Section V discusses the experimental results and analysis. Section VI is the future direction, and conclusion of the study.

II. Related Work

The importance of prompt discharge of patients cannot be overstated. The longer the patient stays, the greater the costs for the hospital and the greater the reduction in the number of new patients that can be admitted. As a result, the patient's LOS has become a standard indicator of treatment quality [17]. LOS has been explored using a variety of methodologies based on the features mentioned above. LOS has been predicted in two ways: (1) without considering a particular disease or condition [18], and (2) considering a particular disease or condition [2], [5]. The LOS is measured using arithmetic approaches, such as average and median [19]. However, it demonstrated that the average LOS can be misleading and thus recommended using the statistical survival analysis methodology [20]. Linear regression and logistic regression are utilized as particular examples of survival models for LOS prediction. These simple deterministic techniques are usually insufficient to provide accurate and reasonable assessment in a clinical setting, that is a complex stochastic system [21]. Furthermore, based on simple assumptions, the generated models require manual adjustments related to medical expertise, which lowers their predicted accuracy.

Hospital management and healthcare professionals can benefit from ML-based models that can predict a patient's LOS in a single visit. ML models can help with: (1) planning preventive measures, (2) optimizing the use of hospital resources, and (3) evaluating the patient's health [8]. ML algorithms were applied to predict LOS in hospitals, focusing on the patient's primary health records on admission [5]. For example, Caetano et al., [7] employed 14 parameters to train a random forests regressor to predict the number of LOS days. It suggested a predictive random forest classifier if the patient remains in hospital for 3, 3–5, or >5 days [2]. The study had an accuracy rate of 80% and found five factors: heart rate, systolic and diastolic blood pressure, age, and insurance status. In addition, [22] evaluated the effectiveness of XGBoost, artificial neural networks, random forests, support vector machines, and gradient boosting decision trees in LOS prediction solely from baseline data. The baseline includes (1) patient-level records such as age, gender, admission, and discharge dates, (2) hospital-level records such as laboratory tests, and (3) discharge diagnostic records. Multi-disease networks and patient similarity networks are used to create features from the baseline data. The performance of all models has been enhanced by the inclusion of multimodal features; for example, the coefficient of determination (R^2) of XGBoost was increased by 18.7%.

Moreover, Triana et al., [23] developed a Deep Neural Network (DNN) regressor that achieved a Mean Absolute Error (MAE) of 1.685 for predicting post-surgery LOS in patients after coronary artery bypass surgery. The length of intubation, age, pretreatment creatinine concentration, and the number of intraoperative red blood cell transfusions were shown to be four out of 116 promising indicators. [5] reviewed existing ML algorithms for estimating LOS and reported that the performance of existing models varied greatly depending on the parameters employed. The study prompted them to conduct further research into the role of ML in predicting LOS for medical patients. Based on the selection of emergency departments and medical records, [24] proposed a DNN model based on natural language processing to predict the patient's LOS. The study used the binary classification problem to predict whether a patient would stay for more than two days, with an accuracy of 82%. However, in the medical area, it is not recommended to design a predictive model based only on clinical notes. [18] presented a DNN to learn other variables such as age and combine them with clinical notes to predict a patient's discharge over the next two days, with an Area Under the Curve (AUC) of 80%.

However, Kalgotra et al., [4] stated that previous LOS estimation approaches had low accuracy and real-world applicability because relying on a single (baseline) modality to predict LOS is not appropriate. Several data-driven methodologies for predicting LOS are available in the literature but the prevailing data-driven and statistical methodologies have failed to account for the inherent uncertainty, complexity, and variability of health processes. [25] and [26] proposed modeling the time dimension to solve the aforementioned challenge. DL algorithms such as convolutional neural network and LSTM can more accurately account for multivariate time-series data [27]. The abstractive interpretations of the raw heterogeneous time-series data can be learned sequentially by DL [28], and the learned features can be optimized together in a comprehensive way [15], [29]. For example, [4] explored using a comorbidity network to develop an LSTM model for LOS prediction. They linked the patient's medical history to prospective comorbidities (determined using a neural network) to investigate the relationship between the diseases. The fusion of comorbidities showed an average mean absolute percentage error of 29.8%. Markov models [26], phase-type distributions [30], and conditional phase-type distributions [31] are some of the other techniques presented.

According to current literature analysis, the levels of physical activity measured by the accelerometer are a helpful proxy for a patient's health status and can be used to assess mobility and functional capacity. Physical activity sensory records can be utilized to reliably predict the patient's LOS [32]. Two methods for determining physical activity are the standard metrics listed below [33]. (1) The self-report method is a simple questionnaire completed by the patient; however, it is based on an assessment of patient mobility rather and not actual performance. Because these quality-of-life assessments are frequently subjective, memory error and understatement can lead to erroneous results [11]. (2) Performance-based methods, such as the 6-minute walk test, objectively assess a patient's activity. Preparing a test session involves a significant investment of time and equipment inadequate for physicians and patients [34]. These assessments only evaluate activity at a single point in time.

In addition, hospital readmission rates are often used as measures of treatment quality. More importantly for hospitals, reducing readmission is a vital component of the Patient Protection and Affordable Care Act for large-scale cost savings [35]. Hospitals that have excessive 30-day readmissions due to certain diseases will face heavy penalties under the Medicare reimbursement system, according to Section 3025 of the Act. Previous research has provided a limited understanding of modifiable variables contributing to readmission probability. These include prevention of adverse events during the first stay [35], a more careful review of medications before discharge, and communicating with patients about self-care after discharge [36].

Studies examining risk factors employ survival models to predict readmissions as a binary classification. According to a recent analysis [37], the Cox regression model and associated

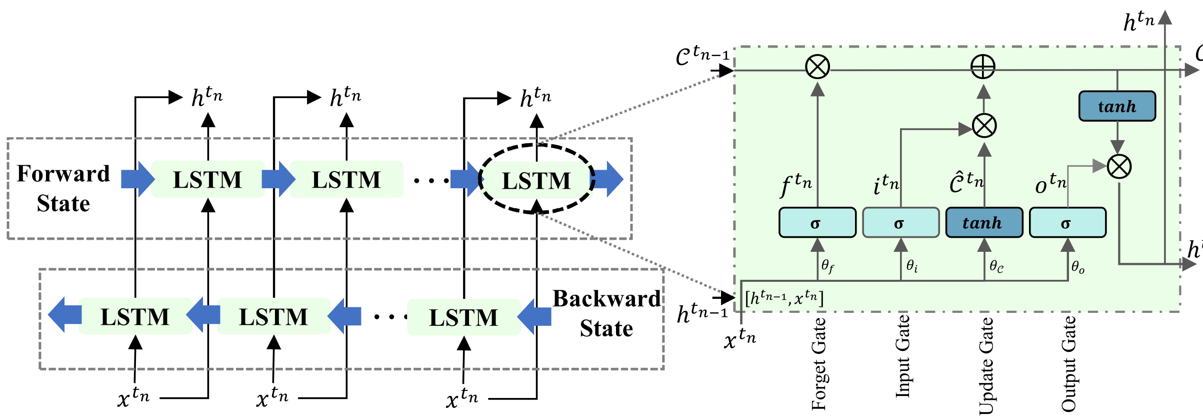

Fig. 1. The internal mechanism of the LSTM cell.

modifications [38] have been the most commonly used survival strategy in predicting readmission. Increasingly, ML techniques are considered to be better at capturing highly non-linear and complicated relationships than traditional methodologies. Artificial neural networks [39], support vector machines [40], random forests [41], and decision trees [42] commonly use ML approaches.

III. Formulation of Problem

In multivariate time-series data analysis, the value of one feature at a specific time often depends on the historical values of that feature and other features. For k time series of features $X_1, X_2, \ldots, X_k$, the predicted value of X_1 at time $t+1$ depends on the value of X_1 in all previous time steps $X_1^t, X_1^{t-1}, \ldots, X_1^{t-n}$. It depends on n historical values of other features, i.e., $\hat{X}_1^{t+1} = f(X_1^t, \ldots, X_k^t; \ldots; X_1^{t-1}, \ldots, X_k^{t-2}; X_1^{t-n}, \ldots, X_k^{t-n})$. The LSTM DL model can utilize the temporal correlation between the historical values of each time-series feature, and the correlation information between multivariate time-series in the prediction process [15], [27]. The LSTM utilizes three primary gating units in its basic structure: (1) an input unit (i^{t_n}) for updating as; $i^{t^n} = \sigma(\theta_i \cdot [h^{t_{n-1}}, x^{t_n}] + b_i)$, (2) a forget unit (f^{t_n}) for maintenance as; $f^{t^n} = \sigma(\theta_f \cdot [h^{t_{n-1}}, x^{t_n}] + b_f)$, and (3) an output unit (o^{t_n}) for deleting as; $o^{t^n} = \sigma(\theta_o \cdot [h^{t_{n-1}}, x^{t_n}] + b_o)$, shown in Fig. 1.

Each gating unit has a control parameter. For time instance t_n, C^{t_n} denotes the current status of the cell, $C^{t_{n-1}}$ represents a step before the cell's status, $\hat{C}^{t_n}$ takes the updated value of the cell's status. Furthermore, h^{t_n} describes the value of the hidden layer. The terms are defined as below:

$$\hat{C}^{t_n} = tanh(\theta_C \cdot [h^{t_{n-1}}, x^{t_n}] + b_C)$$

$$C^{t_n} = f^{t_n} \otimes C^{t_{n-1}} \oplus i^{t_n} \otimes \hat{C}^{t_n}$$

$$\hat{h}^{t_n} = o^{t_n} \otimes tanh(C^{t_n})$$

where the output of the hidden layer's memory cell is $h^{t_{n-1}}$ and the set of weight matrices and biases for the gating units are depicted by θ_* and b_*, respectively. In addition, the standard logistic sigmoid function is σ. The Hadamard product is $\otimes$. The concatenation operation is performed by $\oplus$. The output activation function is ϕ, such as $SoftMax$ or $Tanh$.

The proposed LSTM model jointly learns two correlated tasks: (1) y^1: hospital patient's LOS as a regression task, and (2) y^2: hospital readmission within 30 days as a binary classification task. The study depends on an accelerometer, gyroscope, ambient measurements sensors, patient-level, and laboratory test modalities represented as $X = X^1, \ldots, X^5$. Each modality, $X^m = \{x_1^m, \ldots, x_i^m, \ldots, x_N^m\}$ has N patients where each patient x_i^m is a multivariate time-series sequence, $x_i^m = \{x_{i1_t}^m, x_{i2_t}^m, x_{if_t}^m\}$ for $t = 1, \ldots, s$ time-steps for the one hour dataset. As a result, each i^{th} patient is defined as $x_i = \{x_i^1, \ldots, x_i^m, \ldots, x_i^M, y_i^1, y_i^2\}$, $i = 1, \ldots, N$, where y_i^1 is the continuous value of the regression task, and y_i^2 is the class label of the classification task for the i^{th} instance. The joint optimization is performed on shared parameters (θ^{sh}) based on the specified task (θ^t). The per-task hypothesis is parametrized as $f^t(x, \theta^{sh}, \theta^t) : X \rightarrow y$, and the specific loss functions of classification and regression are $L^t(./.) : y^t \times y^t \rightarrow R^+$. The multi-objective optimization problem can be defined to minimize the model loss, as shown below:

$$\min_{\substack{\theta^{sh} \\ \theta^1, \theta^2}} L(\theta^{sh}, \theta^1, \theta^2) = \min_{\substack{\theta^{sh} \\ \theta^1, \theta^2}} \left(\hat{L}^1(\theta^{sh}, \theta^1), \hat{L}^2(\theta^{sh}, \theta^2) \right)$$

where $\hat{L}^t(./.)$ is the specific loss function represented as $\hat{L}^t(\theta_{st}, \theta^t) = \frac{1}{N} \sum_i L(f^t(x_i; \theta^{sh}, \theta^t), y_i^t)$. We addressed both tasks to the common objective function defined as follows since m is the number of θ_{st} and θ_t parameters.

$$\hat{L}^t(\theta^{sh}, \theta^t) = -\left[\frac{1}{N} \sum_{i=1}^{N} y_i \log(p_i) + (1 - y_i) \log(1 - p_i) \right] + \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i^2) + \frac{\lambda}{m} \sum_{j=1}^{m} \theta_j^2$$

where the cross-entropy loss is used to solve the binary classification task with $p_i = \frac{1}{1 + exp(-W^T x_i)}$, and the mean squared loss for the regression task, the term at the end is the regularization. The actual and predicted values of the regression task for patient i are y_i and $\hat{y}_i$, respectively.

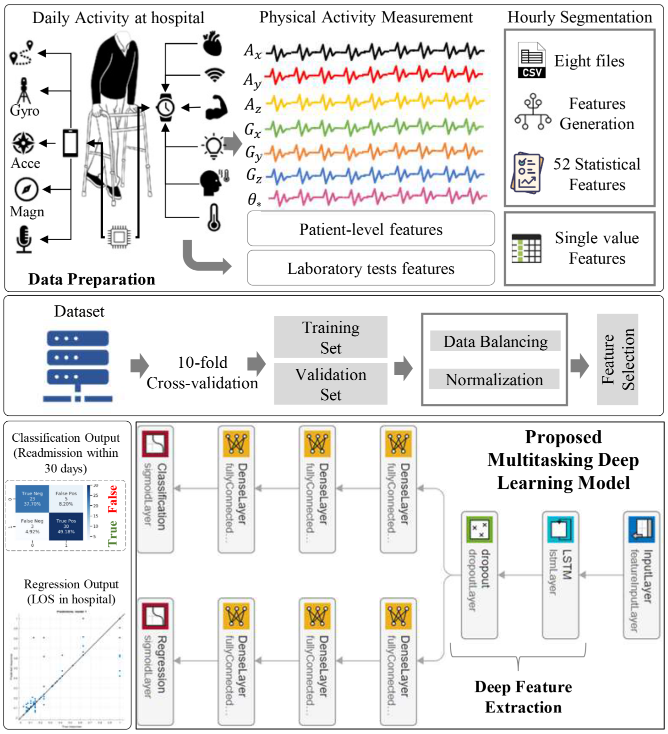

Fig. 2. Diagram of the proposed multitasking deep learning model.

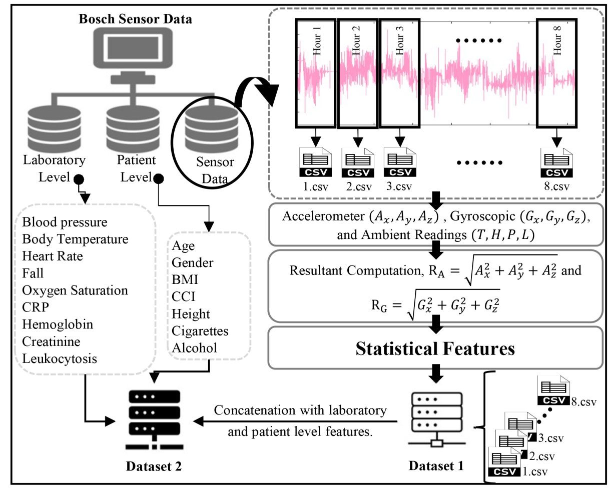

Fig. 3. Dataset structure and preprocessing pipeline of the sensory data.

IV. Design of Proposed Framework

This section discusses the proposed DL model and describes each module in the proposed multitasking model. The proposed multitasking prediction model is shown in Fig. 2. The main goal of the DL model is to accurately perform multitask prediction. The model has two layers: (1) a classification layer to predict readmission to a hospital within 30 days and (2) regression layers to predict a patient's LOS in a hospital. LSTM is a DL architecture that can extract temporal features from long time-series data. The LSTM is used to predict deep common representation from multimodal time-series data in the proposed model. Then, each task is optimized using a separate DL model. In the following subsections, we discuss each phase of the proposed model.

A. Description of Dataset

The proposed framework is trained and validated using a physical activity dataset collected in the hospital using a wrist-worn Bosch sensor system. It is an open-source clinical database of 47 patients from Robert-Bosch hospital in Stuttgart, Germany, recorded in 2016 [16]. For 30 days, the sensor records numerous aspects of the physical activity of patients of varying ages treated in internal medicine departments - the main goal of creating algorithms for LOS prediction and 30-day readmission. Acute infections, heart failure, exacerbations of Chronic Obstructive Pulmonary Disease (COPD), and fluid overload are among the most common diagnoses when admitting patients. In addition, the dataset includes clinical assessment variables such as age, gender, height, weight, alcohol and cigarette consumption, and Charlson Comorbidity Index (CCI) scores. The patient's laboratory data include white blood cell counts, C-reactive Protein (CRP), creatinine, heart rate, blood oxygen saturation, body temperature, and hemoglobin concentration. Baseline statistics, as mentioned above, can be found in Table I.

The sensor consists of an accelerometer, gyroscope, and additional readings. The accelerometer and gyroscope measure the patient's physical activity such as walking and movement with a sampling rate of 100 Hz. Additional readings include ambient conditions such as Temperature (T), Humidity (H), air Pressure (P), and Light (L) had a sampling rate of 1 Hz.

B. Data Preparation

There are 58 subjects in the Bosch sensory dataset. Every day, each patient monitors their physical activities on an ongoing basis during their stay in the hospital. The following conditions must be met to guarantee a minimum amount of data per record: (1) at least 8 hours of data have been collected, and (2) the sensor is worn for at least 80% of the total recorded time. In addition, the minimum length of hospital stay for patients is expected to be 24 hours. We included 47 participants who had at least one day (24 hours) of the sensor reading.

The sensor data was divided into eight parts, each covering one hour, as shown in Fig. 3. We retrieved the sensor data for each hour using a dual ten-second sliding window technique with two successive windows that do not overlap. The window size is based on the assumption that no activity will take more than ten-second to complete [43]. Each hour segment comprises 360,000 samples with a window size of 60,000 samples (ten-second). Each patient has six steps in one-hour, twelve steps in two-hour, and 48 steps in the eight-hour sensory dataset. Each inertia sensor (accelerometer and gyroscope) has three dimensions. Each sensor reading may have a vector structure $v = [x, y, z]$ at each point in time, where x,y, and z represent the three dimensions of the sensor record. As described below, we derive the statistical features of the sensory data:

TABLE I
STATISTICS ON THE BASELINE CLINICAL VARIABLE AND LABORATORY DATA

Feature Name	Total	Male/Female M/F	Maximum (M/F)	Minimum (M/F)	Standard Deviation (M/F)	Unit of Measurement
Number of Patients	47	24/23	-	-	-	-
Number of readmissions after 30 days	16	8/8	-	-	-	-
Number of not readmission after 30 days	31	16/15	-	-	-	-
Average BMI-norm for readmission patients	24.694	28.53/20.85	42.2908/27.0552	21.7994/0	6.7855/8.6999	kg/m^2
Average BMI-norm for not readmission patients	22.953	21.781/24.205	37.874/37.0242	0/16.7062	11.6891/5.9647	kg/m^2
Average CCI for readmission patients	6.562	6.87/6.25	12/11	3/4	2.6959/2.6049	Scale 1-16
Average CCI for not readmission patients	5.677	5.750/5.600	12/14	1/1	3.2558/4.0497	Scale 1-16
Average Height for readmission patients	158.68	173.87/143.50	187/179	163/0	7.5107/58.7513	cm
Average Height for not readmission patients	153.19	142.187/164.933	190/173	0/156	70.8747/6.1233	cm
Average Cigarettes for readmission patients	0.312	0.625/0	5/1	0/0	1.7678/0	per day
Average Cigarettes for not readmission patients	1.322	1.625/1	20/15	0/0	5.1235/3.873	per day
Average Alcohol for readmission patients	0.25	0.50/0	2/1	0/0	0.75593/0	per day
Average Alcohol for not readmission patients	0.322	0.5625/0.0667	3/1	0/0	0.96393/0.2582	per day
Average Age for readmission patients	71.5	64.75/78.25	85/93	31/66	18.8358/8.2419	years
Average Age for not readmission patients	69.03	70.875/67.066	88/94	48/23	13.4456/21.7074	years
Average Body Temperature for readmission patients	13.856	4.625/23.087	37/37	0/0	13.0815/19.1184	°C
Average Body Temperature for not readmission patients	20.271	15.956/24.873	36.9/39	0/0	18.6874/18.2144	°C
Average Heart Rate for readmission patients	32.625	15/50.25	120/92	0/0	42.4264/42.3109	BPM
Average Heart Rate for not readmission patients	45.161	34.875/56.133	100/110	0/0	41.3697/42.5372	BPM
Average Oxygen Saturation for readmission patients	35.937	11.25/60.62	90/100	0/0	31.8198/50.2932	%
Average Oxygen Saturation for not readmission patients	36.835	36.118/37.600	98/98	0/0	48.1689/47.7027	%
Average CRP for readmission patients	0.218	0.100/0.337	0.8/1.2	0/0	0.28284/0.52082	mg/L
Average CRP for not readmission patients	5.851	2.562/9.360	20.5/31.2	0/0	5.3052/10.1019	mg/L
Average Hemoglobin for readmission patients	31.687	18.87/44.50	151/128	0/0	53.3866/61.8477	g/dL
Average Hemoglobin for not readmission patients	71.184	57.043/86.266	153/142	0/0	66.6642/46.6238	g/dL
Average Creatinine for readmission patients	0.387	0.2125/0.5625	1.7/2.6	0/0	0.60104/0.93799	mol/L
Average Creatinine for not readmission patients	1.223	1.618/0.800	8.9/2.5	0/0	2.9276/0.70204	mol/L
Average Leukocytosis for readmission patients	1.993	1.575/2.412	12.6/6.7	0/0	4.4548/3.3323	count/L
Average Leukocytosis for not readmission patients	5.697	5.543/5.860	20.3/14.2	0/0	6.864/4.3711	count/L

CCI - Charlson Comorbidity Index, CRP - C-reactive protein, BMI - Body Mass Index, BPM - Bears per Minute, mg/L - milligram per liter, g/dl - grams per deciliter, μmol/L - micromole per liter.

TABLE II
TIME SERIES CONSTRUCTED FEATURES OF THE SENSOR DATA. ALL THE FEATURES ARE COMPUTED FOR ACCELEROMETER AND GYROSCOPE READINGS. FOR THE AMBIENT CONDITION (T, H, P, AND L), MEAN AND STANDARD DEVIATION ARE COMPUTED

Statistical Features	Description	Equation
Mean	Mean or average of the elements of the sensor data sequence.	$\mu = \frac{1}{N}\sum_{i=1}^{N}\mathcal{A}_i$
Variance	It is used to measure the dispersion of data points from the mean value.	$v = \frac{1}{N-1}\sum_{i=1}^{N}\lvert\mathcal{A}_i-\mu\rvert^2$
Standard Deviation	Find the amount of spread of the sensory data. This is the square root of the variance.	$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\lvert\mathcal{A}_i-\mu\rvert^2}$
Coefficient of Variation	It measures the relative dispersion of data points around the mean value.	$\mathcal{CV} = \frac{\sigma}{\mu}$
Skewness	It measures the degree of symmetry or distortion from a normal distribution in the sensory data distribution.	$\mathcal{S} = \frac{\in(\mathcal{A}_i-\mu)^3}{\sigma^3}$; $\in$ is the expected value
Kurtosis	It measures the degree of tailedness relative to a normal distribution in the sensory data distribution.	$\mathcal{K} = \frac{\in(\mathcal{A}_i-\mu)^4}{\sigma^4}$
Maximum	The maximum value of the sensory data.	$\max(\mathcal{A}_i, axis=1)$
Minimum	The minimum value of the sensory data.	$\min(\mathcal{A}_i, axis=1)$
Range	It measures the range of a particular data set by subtracting the highest and lowest values.	$\mathcal{R} = \max(\mathcal{A}_i, axis=1) - \min(\mathcal{A}_i, axis=1)$
Coefficient of Range	It measures the relative dispersion of data points.	$\mathcal{CR} = \frac{\mathcal{R}}{\max(\mathcal{A}_i, axis=1)+\min(\mathcal{A}_i, axis=1)}$
Quartiles	It divides the data points into quarters. We computed the quartiles for 25%, 50%, 75%, and 95%.	$\mathcal{Q} = (N+1)\times\frac{i}{4}$; i is the percentile.
Mean Absolute Deviation	It measures the average distance between each data point with respect to the mean value.	$\mathcal{MAD} = \frac{1}{N}\sum_{i=1}^{N}\lvert\mathcal{A}_i-\mu\rvert$

T - Temperature, H - Humidity, P - Pressure, and L - Light

- The resultant (R) of raw sensor data is computed for both the accelerometer and gyroscope.
- The statistical features in each sliding window are determined based on the resultant. As shown in Table II, we constructed 15-time domain statistical features in each time window.
- We constructed 38 features ($2 \times sensors \times 15 + 4 \times ambient\ \ measurements \times 2$) for dataset 1.
- For the dataset 2, we merged laboratory tests (9 variables) and patient-level (7 variables) with the dataset 1 to make 52 features.

Data Split. The dataset is divided into 70% and 30% ratios for training and validation, respectively. A 10-fold cross-validation technique with the 70% partition was used to train and optimize the proposed DL model using 30% of the validation set. This process is repeated ten times, and the average results are reported with standard deviations.

Data Normalization. We employed the $\min - \max$ normalization [44] approach of the created statistical features before introducing them to the DL model in order to make the model converge quickly and reduce training time. We use the following method to convert each feature, which can be designated as p_i; where $i = 1, 2, \ldots, n$:

$$p_i^N = \frac{p_i - \min_{i=1}^{n}(p_i)}{\max_{i=1}^{n}(p_i) - \min_{i=1}^{n}(p_i)}$$

where the normalized feature is $p_i^N \in [0, 1]$.

Data Balancing. Our data is imbalanced. For the first class (readmitted), we have 16 records, and for the second class, we have 31 records (i.e., not readmitted). There are many balancing techniques that generate synthesis data samples [45]. it is not recommended in the medical industry to generate artificial data for data balancing as physicians will be suspicious of the results. To meet the challenges, scientists use the oversampling technique [46]. As a result, we independently oversampled the training and validation datasets. With additional examples from the rare class, this approach balances the dataset by increasing the number of rare samples (e.g., repetition, bootstrapping, or a synthetic minority).

Model Optimization. A Bayesian optimizer [47] with a stratified 10-fold CV was used to determine the optimal hyperparameters. Formally, let $\mathcal{F}$ be a well-behaved function defined on the subset $\mathcal{X} \subset R^d$. The goal is then to solve the following global optimization problem:

$$x^* = \arg\max_{\mathrm{x} \in \mathcal{X}} \mathcal{F}(\mathrm{x})$$

Bayesian optimization aims to find the global optimum of the black-box function $\mathcal{F}(.)$ by constructing a probabilistic model and then using this model to make decisions about the training set $\mathcal{X}$ to evaluate the function, while integrating out uncertainty.

Model Configuration. The proposed regression, classification, and multitask models use the ADAM optimizer with the optimization of the learning rate equal to 0.001. The number of epochs and the training batch size are 24 and 100, respectively. For feature extraction, we use a single LSTM layer with 128 hidden units, followed by a l_2-norm and dropout of 0.01 each. Regression and classification layers are given the obtained learned features. Each task layer comprises three dense layers that work with the ReLU function. For classification layers, there are 64, 32, and 32 hidden units, whereas regression layers include 128, 64, and 32 hidden units.

Evaluation Metrics. We use accuracy, precision, recall, and F1-score metrics to evaluate the performance of the classification task [15]. The performance of the regression task is evaluated using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) metrics [27].

Feature Selection. Recursive XGBoost and SULOV (Searching for Uncorrelated List Of Variables) methods is used to reduce features in order to select the best features for the model [48]. We assume that a data set is $\mathcal{D} = \{(x_i, y_i) : i = 1, \ldots, n, x_i \in R^m, y_i \in R^m\}$, having n samples with m features. Let $\hat{y}_i$ be defined as the value predicted by the model:

$$\hat{y}_i = \sum_{i=1}^{n} \mathcal{F}_i(x_i)$$

where $\mathcal{F}_i$ represents an independent regression tree and $\mathcal{F}_i(x_i)$ denotes the prediction score given by the i^{th} tree to the n^{th} sample. The set of functions $\mathcal{F}_i$ in the regression tree model can be learned by minimizing the objective function as follows:

$$\mathcal{O} = \sum_{i=1}^{n} \mathcal{L}(y_i, \hat{y}_i) + \sum_{i=1}^{n} \Omega(\mathcal{F}_i)$$

TABLE III
THE PERFORMANCE COMPARISON BETWEEN RANDOM FOREST REGRESSOR AND THE PROPOSED DL WITH ONLY A REGRESSION LAYER IS REPORTED FOR EACH HOUR OF SENSORY DATA WITH A DIFFERENT TIME-STEP. DATASET 1 IS EMPLOYED BY BOTH MODELS

Sensory Data	Time Step	Random Forest		Proposed Model	
		MAE	RMSE	MAE	RMSE
Hour1	6	1.256±1.09	1.418±1.52	1.027±0.504	1.173±0.931
Hour2	12	0.971±1.28	1.335±1.44	0.911±0.479	1.143±0.785
Hour3	18	0.849±1.35	1.253±1.37	0.843±0.728	1.109±0.819
Hour4	24	0.834±1.27	1.209±1.14	0.785±0.826	1.045±0.537
Hour5	30	0.763±1.05	1.174±0.98	0.693±0.638	1.012±0.463
Hour6	36	0.733±0.94	1.154±1.05	0.686±0.714	0.984±0.604
Hour7	42	0.710±1.10	1.132±0.85	0.654±0.864	0.953±0.714
Hour8	48	0.638±0.89	1.094±0.71	0.586±0.340	0.914±0.525

Multi-Sensors Features Fusion. The features fusion approach is essential for combining different sensory inputs in order to increase the performance of DL models. We combine the features of three sensors (accelerometer, gyroscope, and ambient measurement), patient-level, and laboratory test features. We prepared two different datasets as discussed in Section IV-B. The features fusion helps the DL model to better comprehend the training data and maintains the feature independence in the result matrix for each sensor. At each data point in the figures, the standard error is represented as an error bar.

V. RESULTS AND DISCUSSION

The performance of the proposed multitasking DL model is evaluated in this section from various aspects, with the following key objectives: (1) comparison of the regression and classification as a single task with traditional ML, (2) performance with time steps variation, (3) comparison between single-task and multitask, and (4) effectiveness of the feature fusion method. The evaluation metrics are presented as (mean $\pm$ standard deviation) of 10-fold cross-validation.

EXPERIMENT 1. The traditional ML regressor is compared to our proposed multitasking DL approach. In this experiment, we only used the regression task settings to compare the proposed model with ML. Random forest regressor is trained with 200 estimators, the number of trees equal to 60, and a maximum depth of 8 using the grid search optimization. The number of grid divisions is 10×10. Dataset 1 is fed into the ML model (refer to Section IV-B for more information on dataset preparation). Table III shows the regression evaluation matrices, i.e., MAE and RMSE, for the random forest model and the proposed multitasking model with only a regression layer.

The proposed DL regression model outperformed the traditional random forest regressor. Overall, the DL model is more accurate and stable. For example, a random forest regressor gave the MAE of 1.256 and the RMSE of 1.418 at Hour1. The suggested DL regressor has a lower MAE of 18.23% and a lower RMSE of 17.27%. As the number of hours increases, more time steps are added to the dataset. As a result, the error measure eventually decreases in both ML and DL.

TABLE IV
THE PERFORMANCE COMPARISON BETWEEN THE RANDOM FOREST CLASSIFIER AND THE PROPOSED DL WITH ONLY THE CLASSIFICATION LAYER MODEL IS REPORTED FOR EACH HOUR OF SENSORY DATA WITH VARYING TIME-STEP. DATASET 1 IS EMPLOYED BY BOTH MODELS

Sensory Data	Time Step	Random Forest				Proposed Model			
		Accuracy	Precision	Recall	F1-Score	Accuracy	Precision	Recall	F1-Score
Hour1	6	60.47±4.44	60.58±3.81	60.97±3.74	59.59±4.01	63.23±3.59	57.74±3.78	59.20±2.97	57.88±3.30
Hour2	12	65.82±4.03	50.47±3.73	51.36±3.15	50.80±3.87	72.57±3.07	76.98±2.31	77.18±2.23	73.34±2.03
Hour3	18	69.35±3.83	52.81±3.50	52.97±3.06	52.90±3.57	73.79±3.03	74.62±2.23	71.52±2.13	72.45±2.04
Hour4	24	72.36±3.13	74.75±3.00	72.57±2.63	72.15±3.03	76.04±2.83	76.86±1.94	70.28±1.93	75.74±1.94
Hour5	30	73.35±2.84	74.87±2.73	73.18±2.38	71.83±2.49	76.85±2.78	79.12±1.86	80.50±1.75	77.78±1.87
Hour6	36	75.49±2.62	64.87±2.12	55.35±2.59	56.87±2.69	79.12±2.66	80.45±1.85	77.87±1.67	71.45±1.73
Hour7	42	76.16±2.53	77.21±2.03	76.95±2.53	57.38±2.63	80.37±2.47	82.92±1.61	84.34±1.45	81.26±1.47
Hour8	48	77.82±2.48	78.24±2.02	77.68±2.23	77.28±2.61	81.61±2.17	83.38±1.57	82.52±1.16	82.91±1.27

As shown in Table III, as the number of steps in both ML and the proposed DL network increases, the error rate decreases significantly. However, it can be seen that the MAE difference is small compared to the RMSE difference between the random forest and the proposed DL model. In ML, the MAE is significantly reduced by about 49%, and the RMSE is reduced by about 23% from step 6 in the first-hour dataset to step 48 in the eight-hour dataset. On the other hand, MAE goes down by 42.94%, and RMSE drops by 22.08% in DL suggested.

EXPERIMENT 2. This experiment compares the classic random forest with our suggested multitasking DL model in the classification task. We solely compared the proposed model to ML using the classification task settings. The grid search optimization is used to train a random forest classifier with maximum splits of 73, which has a maximum deviation reduction criterion. The grid has a total of 10×10 divisions. The random forest classifier is fed with Dataset 1 (refer to Section IV-B for more detailed dataset preparation). Table IV illustrates the classification evaluation matrices for the random forest classifier and the proposed model with only the classification layer activated in terms of accuracy, precision, recall, and F1-score.

The suggested DL classification model outperforms the standard random forest classification model. Overall, the DL model works better than ML with varying time steps. For example, a random forest classifier yielded an accuracy of 60.47% with a high degree of variance (SD = 4.44) at Hour1. Similarly, the other matrices show about 60% performance. The suggested DL classification achieved 3% greater accuracy with lower variance (SD = 3.59). However, the precision, recall, and F1-score were weaker. As the number of hours increases, more time steps are added to the dataset. In ML and DL, model performance and stability improve as additional time steps are added.

The classification accuracy of both the random forest classifier and the proposed DL model continues to improve as the number of steps increases, as shown in Table IV. However, in other matrices, the performance in the 4^{th} hour shows a strong transition. The random forest classifier outperformed the proposed model in terms of precision, recall, and F1-score in the 4^{th} hour. The performance of the proposed predictive model improved by almost 24% after integrating 8-hour sensory data, while the random forest classifier's performance improved by 17.35%.

EXPERIMENT 3. Traditional random forest regressor is compared to the proposed multitasking DL multimodal. We only used the regression task settings in this experiment. A random forest regressor with the same parameters as in experiment 1 is employed. The only difference is that Dataset 2 (refer to Section IV-B for more details on the dataset preparation) is loaded into a random forest and the proposed multimodal. Table V show the regression evaluation matrices, i.e., MAE and RMSE, for the random forest model and the proposed multimodal with only a regression layer.

TABLE V
THE PERFORMANCE COMPARISON BETWEEN RANDOM FOREST REGRESSOR AND PROPOSED MULTIMODAL DL WITH REGRESSION ONLY LAYER IS REPORTED FOR EACH HOUR OF SENSORY DATA WITH VARYING TIME-STEP. DATASET 2 IS EMPLOYED BY BOTH MODELS

Sensory Data	Time Step	Random Forest		Proposed Model	
		MAE	RMSE	MAE	RMSE
Hour1	6	1.171±1.01	1.301±1.41	0.1292±0.49	0.1339±0.31
Hour2	12	0.92±1.06	1.110±1.37	0.1274±0.28	0.1309±0.28
Hour3	18	0.818±0.96	1.130±1.28	0.1215±0.25	0.1215±0.73
Hour4	24	0.794±0.97	1.150±1.10	0.1176±0.98	0.1184±0.57
Hour5	30	0.685±0.95	1.113±0.90	0.1081±0.53	0.1136±0.26
Hour6	36	0.700±0.81	1.147±1.01	0.1068±0.32	0.1157±0.43
Hour7	42	0.681±0.84	1.121±0.81	0.1074±0.72	0.1118±0.46
Hour8	48	0.605±0.72	1.081±0.68	0.1055±0.41	0.1024±0.25

The proposed DL regression model outperformed the random forest regressor. Overall, the proposed multimodal is more accurate and stable. For example, a random forest regressor gave the MAE of 1.171 and the RMSE of 1.301 in Hour1. Whereas the LSTM-based multimodal DL regressor has the MAE of 0.1292 (88.96% lower) and the RMSE of 0.1339 (89.70% lower). As shown in Table V, more time steps are added to the dataset as the number of hours increases. As a result, the error measure eventually decreases in both ML and DL. However, it can be seen that the difference between MAE is small compared to the difference in RMSE. In ML, the MAE is significantly reduced by about 48.33%, and the RMSE decreases by about 16.91% from step 6 in the first-hour dataset to step 48 in the eight-hour dataset. On the other hand, the MAE lowers by 18.34%, and the RMSE drops by 23.52% in the proposed multimodal DL. A comparison between the proposed multimodal and random forest regressor based on the MAE and RMSE evaluation matrix is shown in Fig. 4. After incorporating demographic and laboratory test features into the sensory data, both the MAE and RMSE evaluation matrix for the proposed multimodal have decreased drastically. However, with additional time steps, there is no apparent improvement. As time steps are added, the

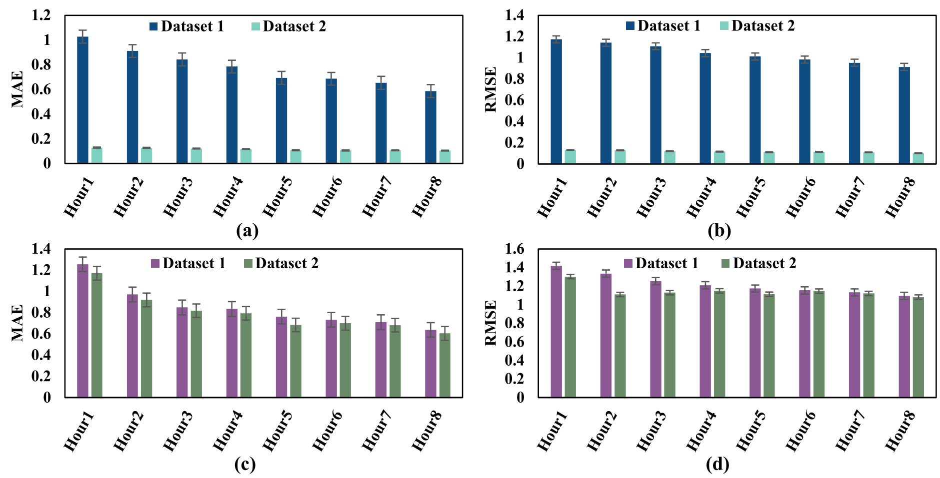

Fig. 4. Illustrates the proposed multimodal comparison based on (a) the MAE matrix and (b) the RMSE matrix between dataset 1 and dataset 2. The bottom row represents the random forest regressor comparison based on (c) the MAE matrix and (d) the RMSE matrix between dataset 1 and dataset 2. For dataset 1 and dataset 2, refer to Section IV-B.

TABLE VI
THE PERFORMANCE COMPARISON BETWEEN THE RANDOM FOREST CLASSIFIER AND THE PROPOSED DL WITH ONLY THE CLASSIFICATION LAYER MODEL IS REPORTED FOR EACH HOUR OF SENSORY DATA WITH VARYING TIME-STEP. DATASET 2 IS EMPLOYED BY BOTH MODELS

Sensory Data	Time Step	Random Forest				Proposed Model			
		Accuracy	Precision	Recall	F1-Score	Accuracy	Precision	Recall	F1-Score
Hour1	6	61.55±3.84	61.60±3.72	61.97±3.50	60.57±3.33	71.37±3.13	72.71±2.87	71.59±2.50	70.27±2.88
Hour2	12	66.80±3.32	51.35±3.47	52.75±2.69	51.77±3.47	73.79±3.13	74.62±3.53	73.79±2.73	72.64±3.04
Hour3	18	70.19±3.61	54.34±2.70	54.27±2.10	53.91±3.20	76.48±3.17	72.65±3.06	72.76±2.51	62.85±2.92
Hour4	24	73.51±3.01	76.47±2.97	73.50±2.08	73.19±2.04	79.11±2.91	67.47±2.81	68.90±2.09	66.10±2.71
Hour5	30	73.55±2.54	76.48±1.80	74.52±1.86	72.78±2.45	85.51±2.77	79.34±2.60	80.51±2.01	83.69±2.63
Hour6	36	76.89±2.30	65.35±1.39	56.79±2.36	57.30±1.81	85.98±2.48	86.49±2.51	87.29±1.85	85.80±2.47
Hour7	42	76.31±2.11	78.57±1.54	77.65±2.04	58.40±1.72	86.39±2.12	86.18±2.01	86.37±1.73	85.28±2.22
Hour8	48	78.55±1.97	78.96±1.44	78.67±1.61	77.87±1.81	87.16±2.05	87.19±1.71	87.71±1.50	82.46±1.70

random forest regressor gradually decreases. Similarly, when using dataset 2 instead of dataset 1, it has a lower MAE and RMSE.

EXPERIMENT 4. This experiment compares the random forest classifier with the proposed multimodal in the classification task settings. The random forest classifier with the same parameters as in experiment 2 was employed. The only difference is that Dataset 2 is entered into the random forest and the proposed multimodal. Table VI illustrates the classification evaluation matrices for the random forest classifier and the proposed model with only the classification layer activated in terms of accuracy, precision, recall, and F1-score.

The comparison between the proposed multimodal with classification layer and the random forest classifier based on the accuracy evaluation matrix is shown in Fig. 6. Other matrices (e.g., precision, recall, and F1-score) can be found in Table VI and Table IV. After incorporating demographic and laboratory test features into sensory data, the accuracy of the proposed multimodal continues to increase. However, after Hour5, there is no noticeable improvement. With the addition of time steps, the random forest classifier gradually improves when using dataset 2 instead of dataset 1.

EXPERIMENT 5. In this experiment, we conducted the multitask, i.e., regression and classification, simultaneously. We compared our proposed multitasking model in two different scenarios. To the best of our knowledge, we could not find any multitask (classification and regression) study based on traditional ML. The experimental results reveal two important observations: (1) the proposed multitasking model improves task performance when compared to a single-task individually, and (2) furthermore, the feature selection method enhances performance.

The performance of the proposed multitasking DL model is shown in Table VII. There has been a considerable improvement in the classification and regression assessment matrices. The multitasking classification approach outperforms the single-task classification model by 7.21% in 48-time steps. Other performance matrices, such as precision, recall, and F1-score, improved significantly by 6.53%, 4.82%, and 4.49%, respectively. The multitask regression approach enhances its performance in comparison with the single-task regression approach. As can be seen from Table VII, the MAE and RMSE matrices decrease by 0.078 and 0.004, respectively.

Experiment 6. After incorporating other modality features, the multitask model had a maximum accuracy of 94.37% with a minimum variance of 0.91. On the other hand, in the proposed multitask and multimodal model, as the number of time steps increases, the MSE and RMSE gradually decrease compared to the single-task regression model (with or without modality fusion).

TABLE VII
THE PROPOSED MULTITASKING DL MODEL'S PERFORMANCE IS REPORTED FOR EACH HOUR OF SENSORY DATA WITH VARYING TIME-STEP TAKING INTO ACCOUNT PATIENT LEVEL AND LABORATORY TESTS VARIABLES, I.E., DATASET 2

Sensory Data of	Time Step	Accuracy	Precision	Recall	F1Score	MAE	RMSE
Hour1	6	86.25±2.20	73.39±2.38	78.38±2.12	72.12±2.51	0.0415±0.012	0.1215±0.26
Hour2	12	89.20±2.08	78.72±2.14	78.26±2.06	75.16±2.47	0.0399±0.017	0.1184±0.31
Hour3	18	89.93±1.93	77.62±2.02	77.80±1.95	75.60±2.30	0.0368±0.016	0.1152±0.33
Hour4	24	90.26±1.89	89.71±1.90	89.25±1.79	89.40±2.11	0.0363±0.018	0.1119±0.24
Hour5	30	91.39±1.61	92.19±1.77	92.90±1.48	90.83±1.83	0.03290.0013	0.1029±0.21
Hour6	36	92.73±1.12	83.47±1.65	93.80±1.22	87.82±1.57	0.0325±0.018	0.1013±0.25
Hour7	42	94.75±1.01	88.17±1.43	87.95±1.15	84.68±1.12	0.0311±0.011	0.1007±0.21
Hour8	48	94.37±0.91	93.72±1.14	92.53±1.04	86.95±1.01	0.0276±0.015	0.0989±0.22

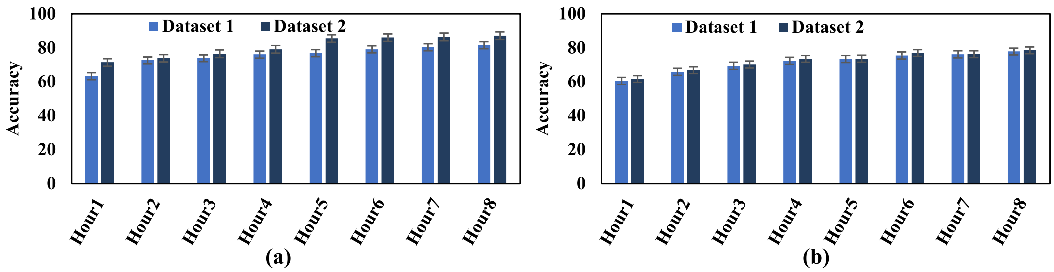

Fig. 5. It shows the comparison of the single classification task of dataset 1 and dataset 2 based on the accuracy matrix. (a) proposed multimodal for classification, and (b) random forest classifier. For dataset 1 and dataset 2, refer to Section IV-B.

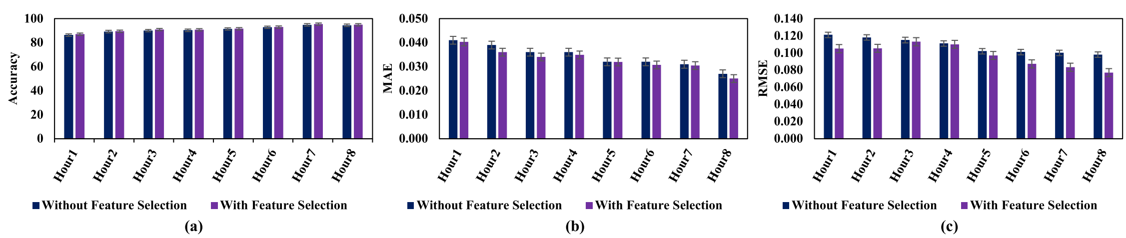

Fig. 6. A comparison is shown for the proposed multitask and multimodal with and without feature selection, employing dataset 2. (a) It shows the performance of the proposed multitask classification in terms of the model accuracy. (b) It shows the performance of the proposed multitasks regressor in terms of the MAE model matrix. (c) It shows the performance of the proposed multitasks regression in terms of the RMSE model matrix.

In conclusion, the study showed that increasing the number of time steps improves the performance of both the random forest and the DL model for single-task and multitasking. Furthermore, the best performance can be obtained by combining a wide variety of data with multi-sensory inputs.

In this experiment, we investigate the role of feature selection in training the proposed model. We employed a feature selection method (e.g., recursive XGBoost and SULOV) to evaluate the proposed multitask and multimodal. We have eight separate data files, each with an hourly metric. The data from each hour was subjected to a feature selection procedure, and the common features were included in the final experiment. Thirteen common features were determined, including ambient features (mean of humidity, light, pressure, and temperature), laboratory tests (BMI, CCI, heart rate, creatinine, age, and hours in hospital), and sensory measurement (minimum, quartile 25, quartile 50 of accelerometer). Fig. 5 shows a comparison of the proposed multitask DL model with and without the feature selection method.

The deep classification model, as shown in Fig. 5(a), initially performed the worst of all the models (at hour 1). In hour 2 and hour 3, it performs adequately with deep multitasking architecture with no other features. The single deep classification model outperformed the multitasking model by the fourth hour without modality fusion. After applying the feature selection method, there was a slight improvement in classification and regression as multitasking. The multitasking with feature selection approach outperforms the model without feature selection by 0.47% at 48-time steps. Other performance matrices, such as precision, recall, and F1-score, showed a similar trend of slight improvement of 0.75%, 0.13%, and 0.53%, respectively. In addition, the multitask regression approach enhances its performance. As can be seen in Figs. 5(b) and (c), the MAE and RMSE matrices are reduced by 0.002 and 0.021, respectively.

Complexity Analysis and Comparison. The proposed model is shown in Fig. 2 as having a single LSTM layer followed by a layer for recurrent projection. The architecture minimizes the traditional LSTM network's (each time-step $\mathcal{O}(n)$) high learning computational complexity. Therefore, the advantage of the proposed architecture is the minimization of the complexity of updating each weight to $\mathcal{O}(1)$, comparable to backpropagation in time [59]. The output space has two

TABLE VIII
COMPARISON OF THE PROPOSED MODELS FOR THE LOS PREDICTION AND READMISSION WITH THE MOST RECENT LITERATURE

References	Year	Disease	Time-series	Modality	Subjects	Method	Performance
Our*	2022	Heart failure Acute infections COPD exacerbation	■	Multimodality	47	LSTM Multitasking	Accuracy = 94.37% MAE = 0.0276
[55]	2021	Pneumonia	-	Multimodality	210	Statistical analysis Single-task	-
[56]	2021	Heart failure	■	Multimodality	4532	Hierarchical Attention Network Single-task	AUC-ROC = 0.87
[57]	2021	Heart failure	■	Multimodality	4529	Ensemble Learning Single-task	AUC-ROC = 0.86
[58]	2021	Torso Trauma	■	Multimodality	840	ANN & SVM Multitask	Accuracy = 83% ROC = 0.87
[59]	2020	Heart failure	■	Multimodality	4559	XGBoost Single-task	AUC-ROC = 0.857
[60]	2020	Heart failure	■	Multimodality	10664	Stacking Ensemble Model Single-task	Accuracy = 94.4% ROC = 0.933
[61]	2019	Heart failure	■	Multimodality	4535	Multi-scale Convolutional Single-task	ROC-AUC = 0.8735
[62]	2019	Kidney injury	■	Multimodality	19044	Random Forest Single-task	AUROC = 0.866
[63]	2018	Heart failure	■	Multimodality	10972	Deep Rule-based Fuzzy Single-task	Accuracy = 73.90%
[64]	2017	Heart failure	-	Single-modality	61	Statistical analysis Single-task	-

COPD - Chronic Obstructive Pulmonary Disease

complexities, implying that the hidden cell states are vectors of two dimensions. The proposed multitasking model is compared with the recent literature as summarized in Table VIII.

VI. CONCLUSION AND DIRECTION OF THE FUTURE

In this study, we proposed a multitask multimodal DL model to predict various tasks such as predicting LOS as a regression problem and patient rehospitalization as a classification problem. The proposed architecture includes a wrist-worn Bosch sensor dataset that collects sensory data for time-series activity in the hospital. The ambient conditional readings, patient-level measurement, and laboratory test data are combined with recorded sensory data of 47 patients' physical activity. The continuous sensory data was subdivided into eight sections, each with an hourly recording time. The dual 10-second window-based approach creates time steps, resulting in six steps per hour. By translating the sensory input into the resultant vector, 30 statistical attributes are calculated. To assess the proposed DL model, five experiments are conducted: (1) a random forest regressor is trained as a regression task without combining laboratory and patient-level data, (2) the proposed DL model is trained with regression settings on the same dataset, (3) a random forest classifier is trained as a classification task without fusion of laboratory and patient-level data, (4) The proposed DL model is trained on the same dataset with classification settings, and (5) two sub-experiments are conducted with the multitask problem to check the combination of sensory data with other laboratory data. Overall, the proposed multitask multimodal DL model predicts the patient's readmission status with high accuracy of 94.84% and predicts the patient's LOS in the hospital with a minimum MSE of 0.025 and RMSE of 0.077, which is a promising, effective, and reliable.

In future research, we intend to investigate the following areas. A deep convolutional neural network or autoencoder is considered for automatic feature extraction. On the other hand, we would like to increase the dataset by including a more significant number of people and sensors in many areas of the body simultaneously.

REFERENCES

[1] W. H. Organization et al., "Global spending on health: A world in transition," World Health Organization, Geneva, Switzerland, Tech. Rep. WHO/HIS/HGF/HFWorkingPaper/19.4, 2019.

[2] T. A. Daghistani, R. Elshawi, S. Sakr, A. M. Ahmed, A. Al-Thwayee, and M. H. Al-Mallah, "Predictors of in-hospital length of stay among cardiac patients: A machine learning approach," *Int. J. Cardiol.*, vol. 288, pp. 140–147, 2019.

[3] A. Lim and P. Tongkumchum, "Methods for analyzing hospital length of stay with application to inpatients dying in southern Thailand," *Glob. J. Health Sci.*, vol. 1, no. 1, pp. 27–38, 2009.

[4] P. Kalgotra and R. Sharda, "When will I get out of the hospital? modeling length of stay using comorbidity networks," *J. Manage. Inf. Syst.*, vol. 38, no. 4, pp. 1150–1184, 2021.

[5] S. Bacchi, Y. Tan, L. Oakden-Rayner, J. Jannes, T. Kleinig, and S. Koblar, "Machine learning in the prediction of medical inpatient length of stay," *Intern. Med. J.*, vol. 52, no. 2, pp. 176–185, 2022.

[6] A. Freitas et al., "Factors influencing hospital high length of stay outliers," *BMC Health Serv. Res.*, vol. 12, no. 1, pp. 1–10, 2012.

[7] N. Caetano, R. M. Laureano, and P. Cortez, "A data-driven approach to predict hospital length of stay," in *Proc. 16th Int. Conf. Enterprise Inf. Syst.*, 2014, vol. 1, pp. 407–414.

[8] A. Awad, M. Bader-El-Den, and J. McNicholas, "Patient length of stay and mortality prediction: A survey," *Health Serv. Manage. Res.*, vol. 30, no. 2, pp. 105–120, 2017.

[9] I. T. Peres, S. Hamacher, F. L. C. Oliveira, A. M. T. Thomé, and F. A. Bozza, "What factors predict length of stay in the intensive care unit? systematic review and meta-analysis," *J. Crit. Care*, vol. 60, pp. 183–194, 2020.

[10] K. Culhane, G. Lyons, D. Hilton, P. Grace, and D. Lyons, "Long-term mobility monitoring of older adults using accelerometers in a clinical environment," *Clin. Rehabil.*, vol. 18, no. 3, pp. 335–343, 2004.

[11] A. Abeles, R. M. Kwasnicki, C. Pettengell, J. Murphy, and A. Darzi, "The relationship between physical activity and post-operative length of hospital stay: A systematic review," *Int. J. Surg.*, vol. 44, pp. 295–302, 2017.

[12] S. T. Rinne et al., "Association between length of stay and readmission for COPD," *Amer. J. Managed Care*, vol. 23, no. 8, 2017, Art. no. e253.

[13] J. Todd, A. Gepp, S. Stern, and B. J. Vanstone, "Improving decision making in the management of hospital readmissions using modern survival analysis techniques," *Decis. Support Syst.*, vol. 156, 2022, Art. no. 113747.
[14] J. Zhou, X. Li, X. Wang, Y. Chai, and Q. Zhang, "Locally weighted factorization machine with fuzzy partition for elderly readmission prediction," *Knowl.-Based Syst.*, vol. 242, 2022, Art. no. 108326.
[15] T. Abuhmed, S. El-Sappagh, and J. M. Alonso, "Robust hybrid deep learning models for Alzheimer's progression detection," *Knowl.-Based Syst.*, vol. 213, 2021, Art. no. 106688.
[16] S. Schricker, N. Schmid, M. Schanz, M. Kimmel, and M. D. Alscher, "In-hospital physical activity measured with a new Bosch accelerometer sensor system," 2020 Accessed: Mar. 28, 2022. [Online].
[17] R. El-Bouri, T. Taylor, A. Youssef, T. Zhu, and D. A. Clifton, "Machine learning in patient flow: A review," *Prog. Biomed. Eng.*, vol. 3, no. 2, 2021, Art. no. 022002.
[18] S. Bacchiet al., "Mixed-data deep learning in repeated predictions of general medicine length of stay: A derivation study," *Intern. Emerg. Med.*, vol. 16, no. 6, pp. 1613–1617, 2021.
[19] M. G. Castillo, "Modelling patient length of stay in public hospitals in Mexico," Ph.D. dissertation, Univ. Southampton, Southampton, U.K., 2012.
[20] A. Marshall, C. Vasilakis, and E. El-Darzi, "Length of stay-based patient flow models: Recent developments and future directions," *Health Care Manage. Sci.*, vol. 8, no. 3, pp. 213–220, 2005.
[21] L. Alexandre, R. S. Costa, L. L. Santos, and R. Henriques, "Mining pre-surgical patterns able to discriminate post-surgical outcomes in the oncological domain," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 7, pp. 2421–2434, Jul. 2021.
[22] Z. Hu, H. Qiu, L. Wang, and M. Shen, "Network analytics and machine learning for predicting length of stay in elderly patients with chronic diseases at point of admission," *BMC Med. Informat. Decis. Mak.*, vol. 22, no. 1, pp. 1–15, 2022.
[23] A. J. Triana, R. Vyas, A. S. Shah, and V. Tiwari, "Predicting length of stay of coronary artery bypass grafting patients using machine learning," *J. Surg. Res.*, vol. 264, pp. 68–75, 2021.
[24] S. Bacchi et al., "Prediction of general medical admission length of stay with natural language processing and deep learning: A pilot study," *Intern. Emerg. Med.*, vol. 15, no. 6, pp. 989–995, 2020.
[25] A. H. Marshall, S. I. McClean, C. Shapcott, I. Hastie, and P. H. Millard, "Developing a Bayesian belief network for the management of geriatric hospital care," *Health Care Manage. Sci.*, vol. 4, no. 1, pp. 25–30, 2001.
[26] L. Garg, S. I. McClean, M. Barton, B. J. Meenan, and K. Fullerton, "Intelligent patient management and resource planning for complex, heterogeneous, and stochastic healthcare systems," *IEEE Trans. Syst., Man, Cybern. - Part A: Syst. Hum.*, vol. 42, no. 6, pp. 1332–1345, Nov. 2012.
[27] S. El-Sappagh, T. Abuhmed, S. R. Islam, and K. S. Kwak, "Multimodal multitask deep learning model for Alzheimer's disease progression detection based on time series data," *Neurocomputing*, vol. 412, pp. 197–215, 2020.
[28] S. El-Sappagh, H. Saleh, F. Ali, E. Amer, and T. Abuhmed, "Two-stage deep learning model for Alzheimer's disease detection and prediction of the mild cognitive impairment time," *Neural Comput. Appl.*, vol. 34, pp. 1–23, 2022.
[29] N. El-Rashidy et al., "Sepsis prediction in intensive care unit based on genetic feature optimization and stacked deep ensemble learning," *Neural Comput. Appl.*, vol. 34, no. 5, pp. 3603–3632, 2022.
[30] L. Garg, S. McCLEAN, B. J. Meenan, and P. Millard, "Phase-type survival trees and mixed distribution survival trees for clustering patients' hospital length of stay," *Informatica*, vol. 22, no. 1, pp. 57–72, 2011.
[31] M. Fackrell, "Modelling healthcare systems with phase-type distributions," *Health Care Manage. Sci.*, vol. 12, no. 1, pp. 11–26, 2009.
[32] H. Jung, J. Kimball, T. Receveur, A. H. Gazi, E. Agdeppa, and O. Inan, "Estimation of tidal volume using load cells on a hospital bed," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 7, pp. 3330–3341, Jul. 2022.
[33] E. Fridriksdottir and A. G. Bonomi, "Accelerometer-based human activity recognition for patient monitoring using a deep neural network," *Sensors*, vol. 20, no. 22, 2020, Art. no. 6424.
[34] P. Sharma, F. E. Shamout, V. Abrol, and D. Clifton, "Data pre-processing using neural processes for modelling personalised vital-sign time-series data," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 4, pp. 1528–1537, Apr. 2022.
[35] K. Carey and M.-Y. Lin, "Hospital length of stay and readmission: An early investigation," *Med. Care Res. Rev.*, vol. 71, no. 1, pp. 99–111, 2014.
[36] A. J. Kind, M. A. Smith, J. R. Frytak, and M. D. Finch, "Bouncing back: Patterns and predictors of complicated transitions 30 days after hospitalization for acute ischemic stroke: (see editorial comments by Dr. Coleman on pp 467–469)," *J. Amer. Geriatr. Soc.*, vol. 55, no. 3, pp. 365–373, 2007.
[37] A. Artetxe, A. Beristain, and M. Grana, "Predictive models for hospital readmission risk: A systematic review of methods," *Comput. Methods Programs Biomed.*, vol. 164, pp. 49–64, 2018.
[38] M. Grzyb et al., "Multi-task cox proportional hazard model for predicting risk of unplanned hospital readmission," in *Proc. IEEE Syst. Inf. Eng. Des. Symp.*, 2017, pp. 265–270.
[39] M. Grana, J. M. Lopez-Guede, J. Irazusta, I. Labayen, and A. Besga, "Modelling hospital readmissions under frailty conditions for healthy aging," *Expert Syst.*, vol. 37, no. 2, 2020, Art. no. e12437.
[40] L. Turgeman and J. H. May, "A mixed-ensemble model for hospital readmission," *Artif. Intell. Med.*, vol. 72, pp. 72–82, 2016.
[41] M. Deschepper, K. Eeckloo, D. Vogelaers, and W. Waegeman, "A hospital wide predictive model for unplanned readmission using hierarchical ICD data," *Comput. Methods Programs Biomed.*, vol. 173, pp. 177–183, 2019.
[42] H. Wang, Z. Cui, Y. Chen, M. Avidan, A. B. Abdallah, and A. Kronzer, "Predicting hospital readmission via cost-sensitive deep learning," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 15, no. 6, pp. 1968–1978, Nov./Dec. 2018.
[43] O. Banos, J.-M. Galvez, M. Damas, H. Pomares, and I. Rojas, "Window size impact in human activity recognition," *Sensors*, vol. 14, no. 4, pp. 6474–6499, 2014.
[44] J. Han, J. Pei, and M. Kamber, *Data Mining: Concepts and Techniques*. New York, NY, USA: Elsevier, 2011.
[45] M. Zhu et al., "Class weights random forest algorithm for processing class imbalanced medical data," *IEEE Access*, vol. 6, pp. 4641–4652, 2018.
[46] N. Japkowicz, "The Class Imbalance Problem: Significance and Strategies," in *Proc. Int. Conf. Artif. Intell.*, Citeseer, 2000, vol. 56, pp. 111–117.
[47] X. Zeng and G. Luo, "Progressive sampling-based Bayesian optimization for efficient and automatic machine learning model selection," *Health Inf. Sci. Syst.*, vol. 5, no. 1, pp. 1–21, 2017.
[48] Z. Zhao, R. Anand, and M. Wang, "Maximum relevance and minimum redundancy feature selection methods for a marketing machine learning platform," in *Proc. IEEE Int. Conf. Data Sci. Adv. Analytics*, pp. 442–452, 2019.
[49] C. K. Ryrsø et al., "The impact of physical training on length of hospital stay and physical function in patients hospitalized with community-acquired pneumonia: Protocol for a randomized controlled trial," *Trials*, vol. 22, no. 1, pp. 1–14, 2021.
[50] G. Harerimana, J. W. Kim, and B. Jang, "A deep attention model to forecast the length of stay and the in-hospital mortality right on admission from ICD codes and demographic data," *J. Biomed. Informat.*, vol. 118, 2021, Art. no. 103778.
[51] C. Guo, M. Liu, and M. Lu, "A dynamic ensemble learning algorithm based on k-means for ICU mortality prediction," *Appl. Soft Comput.*, vol. 103, 2021, Art. no. 107166.
[52] P. V. Staziaki et al., "Machine learning combining CT findings and clinical parameters improves prediction of length of stay and ICU admission in torso trauma," *Eur. Radiol.*, vol. 31, no. 7, pp. 5434–5441, 2021.
[53] N. Hou et al., "Predicting 30-days mortality for MIMIC-III patients with sepsis-3: A machine learning approach using XGboost," *J. Transl. Med.*, vol. 18, no. 1, pp. 1–14, 2020.
[54] N. El-Rashidy, S. El-Sappagh, T. Abuhmed, S. Abdelrazek, and H. M. El-Bakry, "Intensive care unit mortality prediction: An improved patient-specific stacking ensemble model," *IEEE Access*, vol. 8, pp. 133541–133564, 2020.
[55] W. Caicedo-Torres and J. Gutierrez, "ISEEU: Visually interpretable deep learning for mortality prediction inside the ICU," *J. Biomed. Informat.*, vol. 98, 2019, Art. no. 103269.
[56] K. Lin, Y. Hu, and G. Kong, "Predicting in-hospital mortality of patients with acute kidney injury in the ICU using random forest model," *Int. J. Med. Informat.*, vol. 125, pp. 55–61, 2019.
[57] R. Davoodi and M. H. Moradi, "Mortality prediction in intensive care units (ICUS) using a deep rule-based fuzzy classifier," *J. Biomed. Informat.*, vol. 79, pp. 48–59, 2018.
[58] T. Waring, K. Gross, R. Soucier, and R. ZuWallack, "Measured physical activity and 30-day rehospitalization in heart failure patients," *J. Cardiopulmonary Rehabil. Prevention*, vol. 37, no. 2, pp. 124–129, 2017.
[59] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997. [Online]. Available: https://doi.org/10.1162/neco.1997.9.8.1735