Multi-Modal Learning for Inpatient Length of Stay Prediction

Junde Chen[1], Yuxin Wen[1]*, Michael Pokojovy[2], Tzu-Liang (Bill) Tseng[3],
Peter McCaffrey[4], Alexander Vo[4], Eric Walser[4], Scott Moen[4]

Abstract: Predicting inpatient length of stay (LoS) is important for hospitals aiming to improve service efficiency and enhance management capabilities. Patient medical records are strongly associated with LoS. However, due to diverse modalities, heterogeneity, and complexity of data, it becomes challenging to effectively leverage these heterogeneous data to put forth a predictive model that can accurately predict LoS. To address the challenge, this study aims to establish a novel data-fusion model, termed as DF-Mdl, to integrate heterogeneous clinical data for predicting the LoS of inpatients between hospital discharge and admission. Multi-modal data such as demographic data, clinical notes, laboratory test results, and medical images are utilized in our proposed methodology with individual "basic" sub-models separately applied to each different data modality. Specifically, a convolutional neural network (CNN) model, which we termed CRXMDL, is designed for chest X-ray (CXR) image data, two long short-term memory networks are used to extract features from long text data, and a novel attention-embedded 1D convolutional neural network is developed to extract useful information from numerical data. Finally, these basic models are integrated to form a new data-fusion model (DF-Mdl) for inpatient LoS prediction. The proposed method attains the best R^2 and E_{VAR} values of 0.6039 and 0.6042 among competitors for the LoS prediction on the Medical Information Mart for Intensive Care (MIMIC)-IV test dataset. Empirical evidence suggests better performance compared with other state-of-the-art (SOTA) methods, which demonstrates the effectiveness and feasibility of the proposed approach.

Keywords: Chest X-ray images; data-fusion model; length of stay prediction; multi-modal learning.

1. Introduction

Hospital inpatient length of stay (LoS) is a critical indicator that not only measures the efficiency of healthcare facilities but also aids in managing resource utilization effectively. LoS is defined as the time duration that a patient is required to stay in a hospital or other healthcare facility for treatment. Prolonged LoS exposes patients to potential harm associated with hospitalization, such as hospital-acquired

*Corresponding author
[1] Dale E. and Sarah Ann Fowler School of Engineering, Chapman University, Orange, CA 92866, USA.
[2] Department of Mathematics and Statistics, Old Dominion University, Norfolk, VA 23529, USA.
[3] Department of Industrial, Manufacturing and Systems Engineering, The University of Texas at El Paso, El Paso, TX 79968, USA.
[4] University of Texas Medical Branch, Galveston, TX 77550, USA.

infections and adverse drug reactions [1]. In the US, every single extra hour of patient's transfer delay is associated with an adjusted 3% increase in the likelihood of inpatient mortality [2]. From the point of view of a healthcare provider, economic pressures to deliver efficient and accessible care are unprecedentedly high [3]. Healthcare providers face significant economic pressures to deliver efficient and accessible care. The COVID-19 pandemic has exacerbated these challenges to healthcare systems [4, 5]. Consequently, controlling hospital costs and improving service efficiency have become paramount, especially during resource-constrained situations such as the COVID-19 pandemic. Therefore, there is both an urgent need for and practical significance in developing new systems that can automatically predict the length of stay for inpatients.

The healthcare industry is experiencing a transformative phase of digital innovation, which has led to an accumulation of vast patient records in clinical practice. These massive quantities of data hold the promise of supporting a wide range of medical and healthcare applications in clinical decision support, disease surveillance, and population health management. However, to date, most of the successful models that learn from data in healthcare are "unimodal," i.e., they solely focus on a single data modality, ignoring the multi-modal nature of human perception [6]. Owing to the great potential for extracting meaningful insights and learning essential features from data, machine learning (ML) methods offer unprecedented opportunities to improve patient and clinical outcomes. Some notable works include a static input-based regression tree model for LoS prediction [7], an integrated multiple Artificial Neural Network (ANN) model for predicting LoS [8], and a Hierarchical Attention Network model for LoS and mortality predictions [9], etc. Nevertheless, so far, most existing ML models in healthcare mainly rely on a single data modality [6]. Though multi-modal learning has been extensively explored in the context of audio-visual applications [10] and natural image datasets [11, 12], it has been rarely applied in the healthcare domain [13]. Multi-modal learning in healthcare remains relatively unexplored, primarily due to the irregular sampling and diverse formats in which patient data is stored, ranging from diagnostic codes and laboratory results to medical imaging and unstructured texts [13, 14]. The complex and dynamic nature of healthcare data imposes significant challenges for efficient modeling. Consequently, LoS prediction, despite its importance, has received limited attention in the literature due to the inherent difficulty of the task [15]. This study aims to establish a novel data-fusion model that fuses heterogeneous clinical data to predict patient LoS. A multi-modal dataset combining demographic data, clinical notes, laboratory test results, and medical images is used in our methodology. To maximize the utilization of available data, different data modalities are first separately treated with the aid of respective basic sub-models. For numerical data, a novel attention-embedded 1D convolutional neural

network (CNN) is developed to extract meaningful features. By converting the text to sequence data, two long short-term memory (LSTM) networks are used to predict LoS. For chest X-ray image data, a CNN model, which we termed CRXMDL, is designed to predict the inpatient LoS, where the InceptionResnet V2 architecture is used as the backbone and three convolution blocks comprised of 32, 16, and eight 3 × 3 filters, a max pooling (MAP) layer, a flatten layer, and a completely-linked (CL) layer are embedded into the networks to compute specific features of CXR images. Finally, these individual basic models are integrated to generate a new data-fusion model (DF-Mdl), in which two fully connected layers with neuron numbers of 64 and 32 are incorporated to transform vector dimensions, and a CL prediction layer is used to produce the final prediction of inpatient LoS.

The key contributions of this study can be briefly protocoled as follows:

- We develop a data-fusion model that combines multiple "basic" predictive models using various data modalities, including clinical notes, laboratory test results, demographic information, and medical images to predict inpatient LoS. Notably, the LoS prediction task is approached as a regression problem aiming to predict the (expected) number of days a patient will stay in the hospital rather than assigning a patient into one of, say, two classes.
- CXR image data are utilized in our scheme. We developed a basic model referred to as CRXMDL that uses the InceptionResnet V2 architecture as the backbone; and three convolution blocks, a MAP layer, a flatten layer, and a CL layer are embedded into the networks to predict inpatient LoS.
- We convert long text blocks (in natural language) into sequence data after performing segmentation on them, and then design two LSTM networks to infer sequence-dependent feature representations.
- We upgrade the traditional squeeze-and-excitation (SE) block used in the field of computer vision to a 1D SE-block and embed it with spatial attention into a 1D-CNN for numerical data-based feature extraction.

The remainder of this article is organized as follows. Section 2 provides a literature review. Section 3 presents methodology development. The overall architecture of the proposed DF-Mdl along with technical details are discussed in this section. Through a real-world case study, Section 4 demonstrates the efficacy of the proposed approach. Finally, Section 5 concludes the paper and provides some possible future research directions.

2. Related work

Clinical practice produces large amounts of data of various types, such as clinical notes, laboratory test results, measurements of vital signs, and medical images [16]. The unprecedented proliferation of medical data has prompted substantial efforts toward the development of various data-driven models for LoS prediction [17]. Existing data-driven LoS models can be broadly grouped into two categories: classification models and regression models [18-23]. In classification models, the aim is to group the LoS into multiple classes, e.g., short, medium, and long stay, based on the number of days that the patient stays in the hospital. Thompson et al. [24] introduced the prolonged length of stay (PLOS) attribute as a binary nominal variable based on a given threshold and obtained attractive classification results. Similarly, by converting the LoS prediction to a binary classification task, Bednarski et al. [25] introduced a temporal convolutional network for clinical LoS prediction. Their proposed method outperformed other state-of-the-art (SOTA) methods, such as Gated Recurrent Unit (GRU) network and random forest (RF) method. Harutyunyan et al. [26] reduced the LoS prediction to a classification problem with 10 classes/buckets and achieved SOTA performance for forecasting LoS on the MIMIC-III dataset. The aforecited literature indicates the usefulness of classification models in LoS prediction. However, several studies have revealed that the LoS distributions are highly right-skewed [9, 27, 28]. This skewness implies that the dataset becomes imbalanced with a limited number of cases exhibiting long LoS. This imbalance misleads the performance evaluation as classes of long LoS tend to be treated as outliers. Therefore, interpreting the LoS task as a regression problem is a more appropriate and informative way of balancing the dataset as it focuses on the actual number of LoS days (possibly, fractional) and not classes [29]. Ma et al. [27] trained three decision tree regression models, including bagging, AdaBoost, and RF to predict inpatients' LoS. The bagging method achieved the best test results with a root mean square error (*RMSE*) of 0.296, R^2 of 0.831, and an accuracy rate of 0.723. Using four ML algorithms, Hasan et al. [30] performed LoS prediction on the MIMIC-III dataset. The extreme gradient boosting (XGBoost) regressor produced the best results with *RMSE* of 1.2. and R^2 of 0.86. In a similar study, Boff et al. [31] applied five ML algorithms to predict LoS, including multiple linear regression, RF, support vector regression, ridge regression, and partial least squares algorithms. The RF achieved top performance in their experiment with an R^2 of 0.6567 and a mean absolute error (*MAE*) of 3.51 days.

Nevertheless, for both classification and regression models, the research so far has mainly focused on studying one type of data source only. It is expected that the integration of heterogeneous medical data modalities (unstructured clinical notes, laboratory test results, measurements of vital signs, demographic

information, etc.) has the potential to improve prediction accuracy and enhance knowledge discovery from the data. Indeed, Zhang et al. [32] found that integrating heterogeneous data types across electronic health records (EHRs) helps improve the performance of prediction models and reduce errors. In their model, they only combined structured data and unstructured texts. Incorporating more data modalities such as image data may prove helpful if each data modality contains incomplete but complementary information [13, 33]. By integrating four modalities of data, in particular, chest X-ray images and tabular data, Soenksen et al. [33] predicted whether a patient is going to be discharged within the next 48h and converted the LoS prediction to a binary classification problem. Although some promising performance was obtained and image data were used, the class imbalance problem likely had an adverse effect on the performance of their model. Hayat et al. [13] proposed a MedFuse model to fuse the CXR images and clinical time-series data. However, the LoS prediction task was not considered in their work. Besides, the two-stage training strategy increases the computational complexity of the model. Al-Dailami et al. [15] introduced a multi-scale feature fusion model for predicting LoS in ICU and extracted multiple data modalities, such as demographic data, medical history, lab test results, and label information. But the image data were not utilized in their method. To overcome this shortcoming, we establish a data-fusion model that incorporates multiple individual predictive models using diverse data modalities such as clinical notes, laboratory test results, demographic information, and medical images. Section 3 describes the specific procedure underlying our approach.

3. Methodology

3.1 General outline

Compared with single modality based LoS prediction, our approach utilized multiple input sources to improve the prediction performance by integrating multiple heterogeneous data. Fig. 1 displays the overall flowchart of the proposed procedure. First, a novel data fusion architecture extracts meaningful patterns by fusing heterogeneous data, including text, numerical data, and CXR image data. Data pre-processing and cleaning are performed on the extracted data. Specifically, numerical data are normalized before being passed into the model, and the category variables are converted into dummy/indicator variables via One-Hot encoding. The text data are segmented and converted into the sequence by implementing the Python *text.Tokenizer* command. As for CXR image data, they are stored in the MIMIC-CXR-JPG database in JPG format with structured labels, where the dimensions of CXR images are uniformly resized to a fixed size of 512×512 pixels for fitting the models. The pre-processed data are consequently fused and utilized as input for the proposed fused model. The resulting observations are split into training and validation sets to fit the proposed model while tuning respective hyper-

parameters to select the best architecture, and the remaining data are used as a test set to evaluate the final model performance. The details of each given step are presented in subsequent sections.

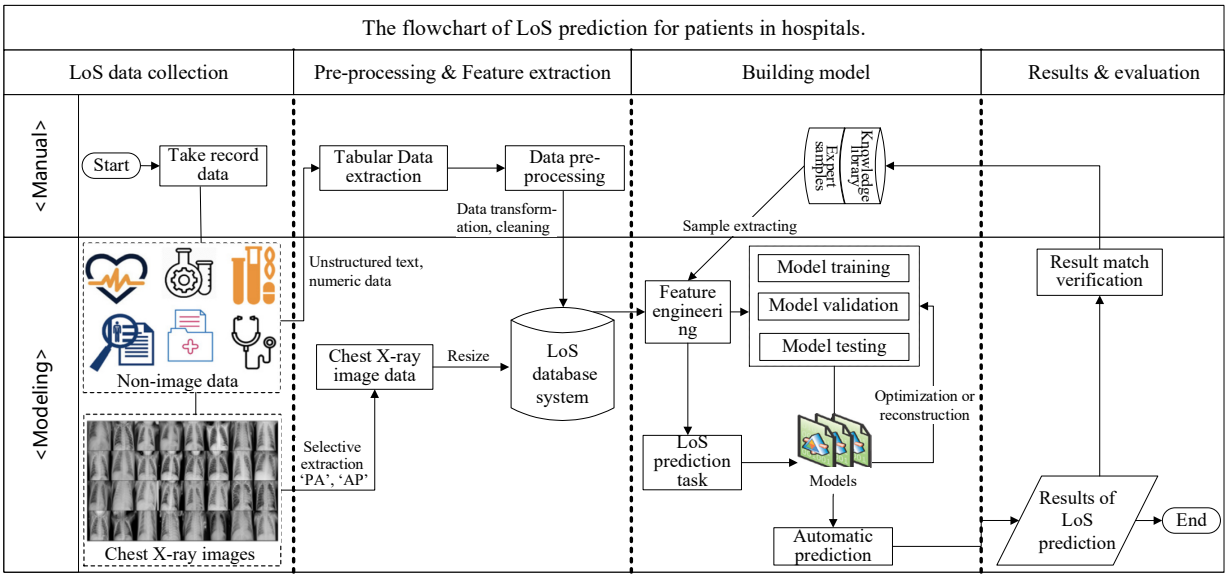


Fig. 1. Overall flowchart of the proposed LoS prediction framework.

3.2 Data fusion model

We consider three different modalities, including numerical variables (demographics, medical exam, etc.), unstructured text data, and CXR image data. Note that the numerical variables include discrete and continuous data. For unstructured text data (i.e., natural language text, such as nursing documentation and physician notes), the word embedding technique can map words or phrases from vocabulary to a corresponding vector of continuous values. However, directly modeling sequential notes using word embedding can be time-consuming. It may be impractical since clinical notes are usually quite lengthy and involve multiple timestamps. Moreover, the length of different documents varies. To address this challenge, we use a tokenizer to implement the work segmentation for long texts, such as the *long_title* field in *d_icd_procedures* and *d_icd_diagnoses* tables, Then, the *Text2Seq* function is used to transform the text data to sequence variables. To capture the dependencies among sequence variables, two long short-term memory networks are designed to take the output of *Text2Seq* for inferring sequence-dependent feature representations. Here, the hyper-parameter of the perceptron number is set to 4, with a ReLu activation function and l_2 regularization. Besides, note that other categorical variables, such as insurance, admission type, marital status, and gender, are converted into numerical variables of model input dictionary by implementing the one-hot encoding. For CXR image data, we design a convolutional neural network model referred to as CRXMDL to train on the data using InceptionResnet V2 [34] as a backbone and then embed three convolution blocks comprised of 32, 16, and eight 3 × 3 filters, max pooling layer, flatten layer, and a fully connected layer to calculate specific features of CXR images. As

for numerical data, a novel attention-embedded 1D convolutional neural network (Att-1DCNN) is developed to predict inpatient LoS. Using 32 small filters with the size of 3, two cascaded one-dimensional (1D) convolution layers followed by a max pooling layer are used to extract high-level features. In particular, an improved 1D SE-block [35] is incorporated into the network for feature adaptive calibration, and following the enhanced 1D SE-block, a spatial attention (SA) mechanism is added in the network to quantify the significance of spatial point features. In this manner, the features obtained by the enhanced 1D SE block and spatial attention are fused to generate the output of the Att-1DCNN for numerical data-based LoS prediction. Finally, combining the Att-1DCNN, LSTM, and CRXMDL, the multiple basic models are integrated to generate a new data-fusion model, for which two densely connected layers with neuron numbers of 64 and 32 are embedded into the networks to change the vector dimensions, and a fully connected (FC) prediction layer is used for the final prediction of inpatient LoS.

Due to the heavily skewed LoS distribution, the prediction task is particularly challenging. In skewed data, the long tail may be treated as an outlier by ML models and adversely affect the model performance [7, 36]. Most of the recent LoS prediction studies tend to ignore this challenge, which may lead to underprediction of LoS and unreliable results [36]. To train the proposed DF-Mdl, the classical mean squared error (MSE) loss function is not an adequate choice as its robustness relies heavily on the Gaussian distribution assumption on the response variable. However, the LoS distribution is typically right-skewed and contains outliers, in which case using a suboptimal loss function may introduce bias and cause large prediction errors. Therefore, to alleviate this problem, the Huber loss function is employed, which furnishes a robust estimation of regression models and reduces outliers' contributions to the squared error loss, thereby limiting the impact on regression estimates [37]. The Huber-type customized loss function [38] used in our network is defined as

$$L(y_k - \hat{y}_k) = \begin{cases} (y_k - \hat{y}_k)^2, & \text{for } |y_k - \hat{y}_k| \le \theta, \\ \dfrac{2\theta^{2-p}}{p} |y_k - \hat{y}_k|^p - \dfrac{2-p}{p}\theta^2, & \text{otherwise.} \end{cases} \tag{1}$$

Here, $\hat{y}_k$ and y_k denote the predicted value and actual value, respectively, p and θ are two non-negative hyperparameters, where p controls the concavity and convexity of the loss function, while θ determines the turning point (p = 1 here). Fig. 2 displays the architecture of the proposed DF-Mdl. Below summarizes the core steps of the proposed algorithm:

1. Given the raw data from different modalities, the original dataset $\boldsymbol{D}$ is pre-processed into subsets $\boldsymbol{D}$=$\{\boldsymbol{D_1}, \boldsymbol{D_2}, \ldots, \boldsymbol{D_M}\}$, where M is the number of total data types. For each sub-dataset $\boldsymbol{D_m}$ of patient i, the data is denoted as $\boldsymbol{D}_{i,m} = \{x_{i,1}, x_{i,2}, \ldots, x_{i,L}\}$, where $x_{i,l}$ denotes the extracted feature for patient i,

L is the total number of features in sub-dataset $\boldsymbol{D}_{i,m}$, here $i = 1,2,\ldots,N, N$ is the total number of patients.

2. Basic models $H=\{H_1, H_2, \ldots, H_M\}$ are constructed separately to extract useful information from each modality. For each sub-model, the corresponding output z_m is obtained, which is denoted as $z_m = H_m(\boldsymbol{D}_m)$.

3. The outputs $\boldsymbol{Z} = \{z_1, z_2, \ldots, z_m\}$ from basic models are concatenated and fed into another network comprised of two densely connected layers with neuron numbers of 64 and 32, while a CL prediction layer is used for the final prediction of LoS. Table 1 depicts vital hyper-parameters in the proposed model.

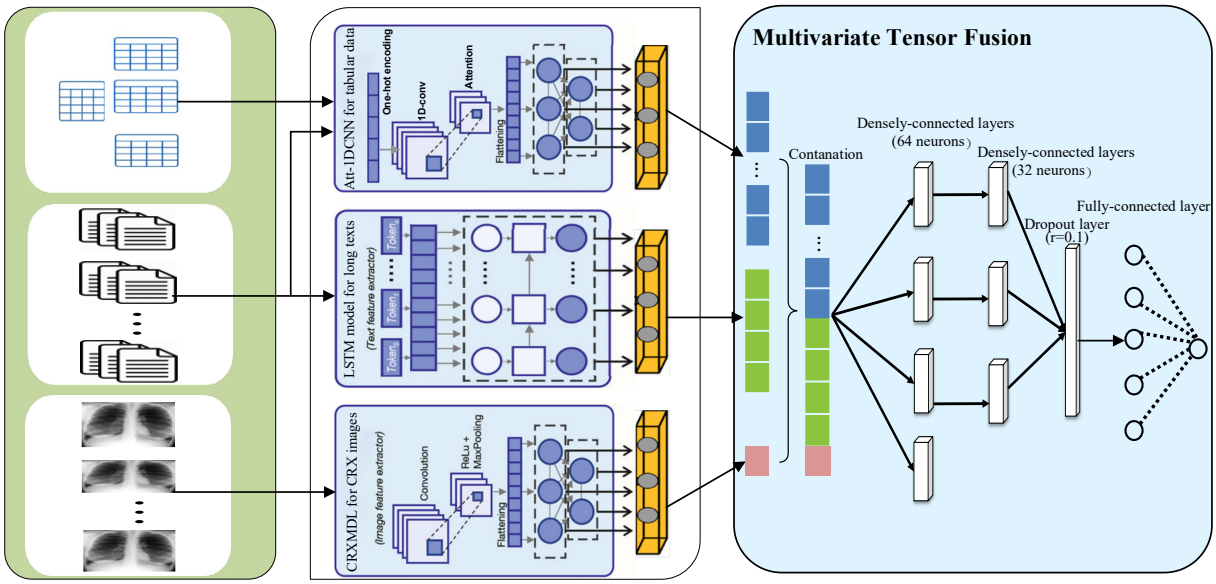


Fig. 2 The overall architecture of the proposed DF-Mdl.

Table 1. Essential hyperparameters in the proposed model.

Layer	Input shape	Filter number	Kernel Size	Output Shape	# of Parameters	Replications
InputLayer (numeric)	(None, 27, 1)	-	-	(None, 27, 1)	-	1
Conv1D	(None, 27, 1)	32	3	(None, 23, 32)	128+3104	2
MaxPooling1D	(None, 23, 32)	-	-	(None, 11, 32)	-	1
GlobalAveragePooling1D	(None, 11, 32)	-	-	(None, 32)	-	1
Conv1D	(None, 32, 1)	32	3	(None, 32, 1)	3	1
sigmoid	(None, 128,128,1)	-	-	(None, 128, 128, 1)	-	1
Multiply	(None, 11, 32)	-	-	(None, 1, 11, 32)	-	1
Add	(None, 11, 32)	-	-	(None, 1, 11, 32)	-	1
reducemax	(None, 11, 32)	-	-	(None, 32)	-	1
reducemean	(None, 32, 1)	32	3	(None, 32, 1)	3	1
Concatenate	(None, 1, 11, 1)	-	-	(None, 1, 11, 2)	-	1
Conv1D	(None, 1, 11, 2)	32	3	(None, 1, 11, 1)	3	1
sigmoid	(None, 1, 11, 1)	-	-	(None, 1, 11, 1)	-	1
InputLayer (text)	(None, 23)	-	-	(None, 23)	-	1
InputLayer (text)	(None, 18)	-	-	(None, 18)	-	1
Multiply	(None,1 11, 32)	-	-	(None, 1, 11, 32)	-	1
Embedding	(None, 23, 32)	-	-	(None, 1, 32)	16416+16928+192	3
Flatten	(None, 1, 11, 32)	-	-	(None, 352)	-	2

InputLayer (images)	[(None,128,128,1)]	-	-	[(None,128,128,1)]	-	1
LSTM	(None, 23, 32)	-	-	(None, 4)	592+592	2
Concatenate	(None, 32)	-	-	(None, 448)	-	2
InputLayer (categorical)	(None, 28)	-	-	(None, 28)	-	1
Flatten	(None, 1, 32)	-	-	(None, 32)	-	1
Sequential	(None,128,128,1)	-	-	(None, 1)	-	1
Dense	(None, 448)	-	-	(None, 32)	28736+2080	2
FC	(None, 32)	-	-	(None, 1)	33	1

4. Case study

To evaluate the performance of the proposed approach, we conducted an empirical study on the Medical Information Mart for Intensive Care (MIMIC)-IV v1.0 database. For benchmarking purposes, state-of-the-art LoS prediction methods were chosen for comparisons.

4.1. Experimental setup

The computations were primarily conducted using Python 3.7 deep learning framework utilizing widely used libraries, such as Keras, Scikit-learn (https://scikit-learn.org/stable/), Matplotlib, and Tensorflow, which are accelerated by a graphics processing unit (GPU). The hardware configuration used for operating the deep learning framework to perform LoS prediction was AMD EPYC 7502P 32-Core Central Processing Unit (CPU), 32 GB memory, NVIDIA RTX A6000 graphics processing unit (GPU). Training, validation, and testing were all carried out in a Linux series OS. The *Adam* optimizer was selected to train the model, with a batch size of 64 and a learning rate of 1.0×10^{-3}.

4.2. MIMIC-IV v1.0 dataset

MIMIC-IV v1.0 dataset contains medical records for over 40,000 patients admitted to the intensive care unit (ICU) at the Beth Israel Deaconess Medical Center (BIDMC) between 2008 and 2019 [39], and it contains all documentation entered for each patient throughout their interaction with the hospital. MIMIC-IV v1.0 dataset includes a wide range of observations through time, such as free-text notes composed by clinicians, medication orders, laboratory test results, procedures, demographic information, and diagnosis codes. In addition, MIMIC-IV v1.0 was complemented by a further data source, MIMIC-CXR [40], which contains patients' chest radiographs in DICOM format with free-text radiology reports. All data records are deidentified, i.e., patient identifiers are removed according to the Health Insurance Portability and Accountability Act (HIPAA) Safe Harbor provision. The data can be accessed at the PhysioNet website (https://physionet.org/content/mimic-iv-demo/1.0/core/) upon signing the data use agreement. Following the cohort selection procedure recommended by [41], we extract a set of features. 511,741 rows of records are extracted from the tabular data of the MIMIC-IV v1.0 dataset. For chest X-ray images, only images with ViewPositions of "PA (posteroanterior)" or "AP (anteroposterior)" are

selected. As such, 170,934 CXR images are retained. Except for CXR images, a total of 52 attributes, including circulatory, endocrine, blood, infectious, injury, and mental, have been extracted from the following tables: ADMISSIONS, DIAGNOSIS_ICD, D_DIAGNOSIS_ICD PATIENTS, ICDSTAYS, PROCEDURES_ICD, and D_PROCEDURES_ICD. These tables record inpatients' demographic data and intensive care data at the BIDMC. For instance, the ADMISSIONS table provides information regarding patients' admission to the hospital, such as admission and discharge time, demographic information, and the source of admission. The DIAGNOSIS_ICD table contains diagnosis category information, which is divided into 17 categories within the ICD9 nomenclature, and the corresponding category names are summarized in the D_DIAGNOSIS_ICD table. The PATIENTS table provides inpatients' age and gender information, and the ICDSTAYS table contains the ICU data for each hospital admission. The PROCEDURES_ICD table provides the procedure code for inpatients, while corresponding procedure names are included in the D_PROCEDURES_ICD table. Then, data transformation, data cleaning, imputation of missing data, and other data preprocessing operations are performed on the original data. The admission and discharge times are converted into date-time type, which is useful for counting the days of inpatient LoS. The null values are replaced with 0, and irrelevant columns such as LANGUAGE, ADMIT_MIN, ADMISSION_LOCATION, and DOB are dropped from the original data. We also removed no longer needed columns like HADM_ID, ADMITTIME, DISCHTIME, EDREGTIME, and EDOUTTIME, and verified that there were no missing values in the dataset. Besides, the negative indicator values or outliers, such as negative LoS initiated by entry form error, were removed as they could skew prediction results. Beyond that, to accelerate model training and testing, we adjusted the dimensions of the CXR images so that all CXR images were uniformly resized to a fixed size of 512×512 pixels. Fig. 3 presents a flow diagram of preparing data on the MIMIC IV v1.0 database. Some representative observations extracted from the MIMIC-IV v1.0 database are summarized in Table 2.

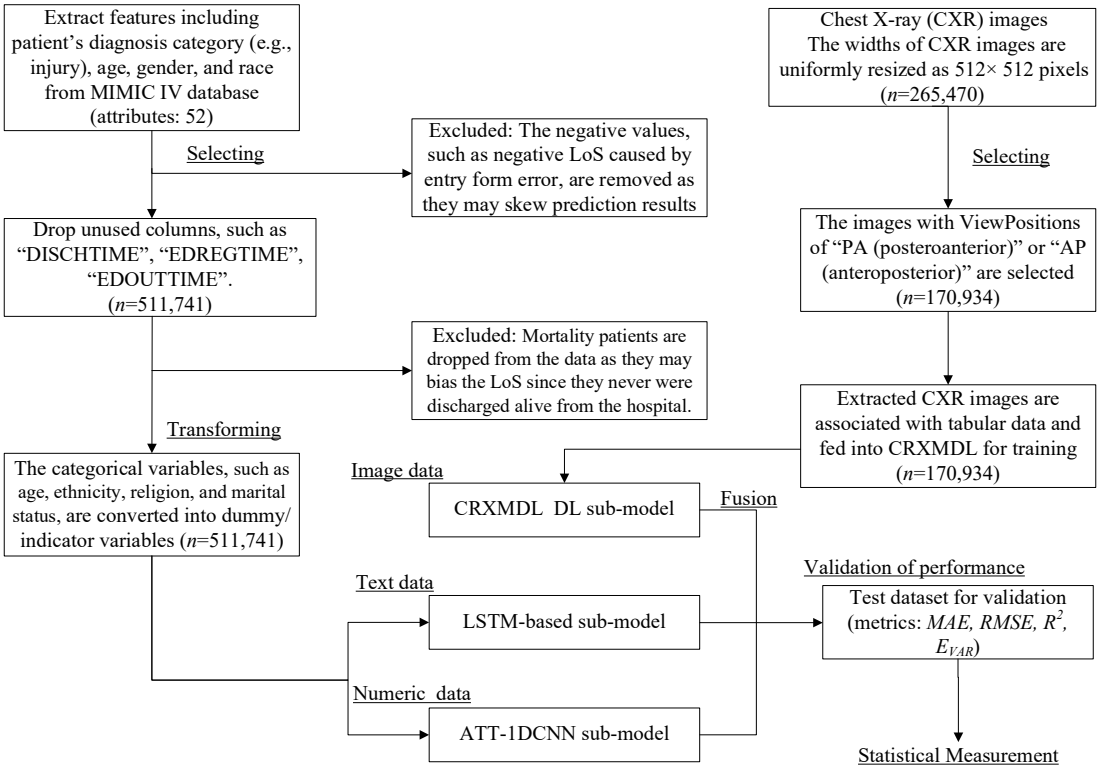
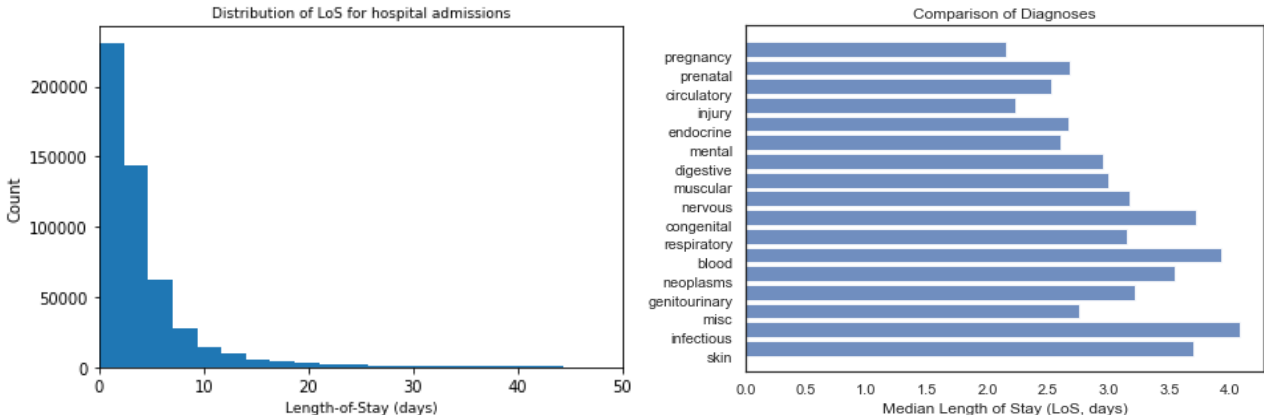


Fig. 3 Flow diagram of preparing the data in the MIMIC IV dataset.

Table 2 A representative sample from the MIMIC-IV v1.0 database.

Subject_id	LoS	Blood	Circulatory	Digestive	Endocrine	...	Injury	Nervous	Pregnancy	Prenatal	Respiratory	Gender
15124376	2.8541	0	0	0	0	...	2	0	2	0	0	1
11303384	2.0166	0	6	0	0	...	1	0	0	0	0	0
14689001	4.4548	0	3	0	1	...	2	1	0	1	0	0
14095761	1.5277	2	1	0	4	...	3	2	0	0	0	1
14095761	3.7437	1	1	1	3		1	2	0	0	0	1
11798821	4.6055	0	0	0	0		17	0	0	0	0	0
11212657	2.9201	0	0	0	0	...	2	0	0	0	0	0
17649604	2.9965	1	2	1	2	...	3	0	0	1	0	1
17649604	4.6041	1	6	1	2	...	3	0	0	0	2	1
16665229	11.975	0	0	0	0	...	18	0	0	0	0	1



(a) Distribution of inpatient LoS (b) Median LoS by diagnosis category

Fig. 4. The LoS distribution and median LoS by diagnosis category.

The LoS of each inpatient is defined as the time difference in days measured between hospital discharge and admission. Fig. 4(a) shows the distribution of inpatient LoS in the MIMIC-IV v1.0 dataset, while Fig. 4(b) displays the distribution of median LoS by diagnosis category. From Fig. 4(a), it can be seen that most LoS is under 10 days, and the median LoS is relatively large for some diagnosis categories, including infectious, blood, congenital, and skin, as shown in Fig. 4(b). Using the data-fusion model proposed in Section 3.2, we perform LoS prediction on the MIMIC-IV v1.0 dataset, and the usual splitting method into training, validation, and test sets refers to the tags provided in the MIMIC-CAR-2.0.0-SPLIT table. The frequently used metrics are employed to calibrate the models, and the specific calculation process and the results are described in subsequent sections.

4.3. Performance metrics

To evaluate the prediction accuracy of the proposed approach, four performance metrics including the mean absolute error (*MAE*), the root mean squared error (*RMSE*), the coefficient of determination *R*-squared (R^2), and the explained variance (E_{VAR}) score are used to facilitate the comparison with benchmark methods:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{2}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{3}$$

$$R^2 = 1 - \sum_{i=0}^{n}(y_i - \hat{y}_i)^2 \Big/ \sum_{i=1}^{n}(y_i - \bar{y})^2 \tag{4}$$

$$E_{VAR} = 1 - \mathrm{Var}(y_i - \hat{y}_i)/\mathrm{Var}(y_i) \tag{5}$$

where y_i and $\hat{y}_i$ stand for the actual value and predicted value, respectively, n denotes the number of predicted samples, and Var(·) is the (empirical) variance operator that calculates the sample variance of a dataset. For both the E_{VAR} and R^2, the ideal value is equal to 1, while greater values are worse for the indicators of *MAE* and *RMSE*.

4.4. Results

As described in Section 3.2, the proposed multimodal fusion model is used for LoS prediction on the MIMIC-IV v1.0 dataset. Fig. 5 shows the training performance of the proposed model. From Fig. 5, it can be seen that the validation accuracy of the proposed approach exceeds 0.6 after training for 30 epochs, indicating a solid performance of the proposed approach for LoS prediction. To demonstrate the superiority of the proposed model, six popular methods, including multilayer perceptron (MLP),

SVM, random forest (RF), extreme gradient boosting (XGBoost), VGG-style CNN (V-CNN), and one-dimensional convolutional neural network (1D-CNN), were selected for comparison purposes. Some hyperparameter settings include a mini-batch size set to 64, a learning rate of 1.0×10^{-3}, 30 epochs of training, and the *Adam* optimizer [42]. Tables 3 and 4 summarize the performance of different methods for LoS prediction. Further, to detect the potential presence of statistically significant differences among the methods studied, we applied the well-known Friedman test [43, 44]. After being ranked, the performance results on test data of different methods are presented in Table 4.

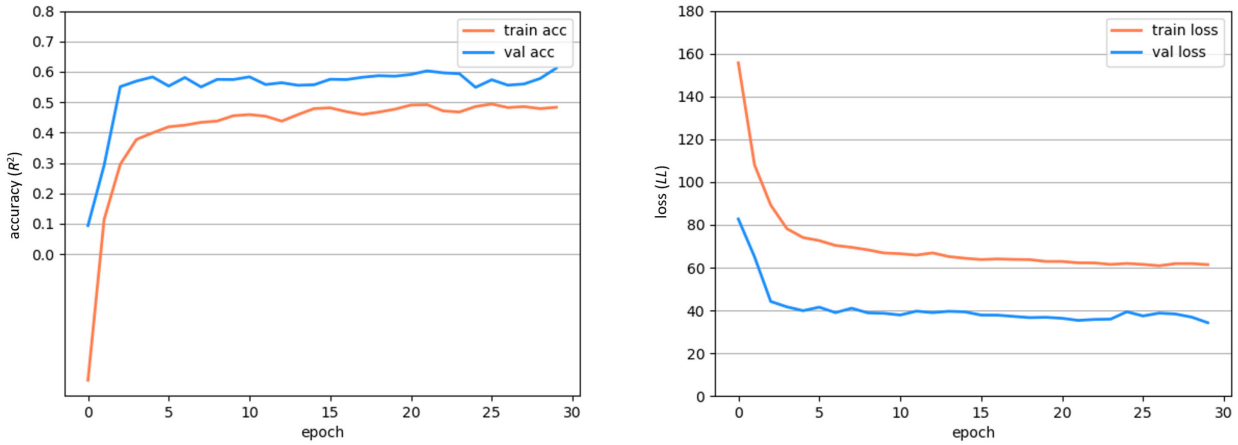


Fig. 5. The training performance of the proposed approach.

Table 3 Comparison of different methods on the training set.

Models	LoS prediction accuracy on the training set				Time (h:m:s)
	MAE↓	*RMSE*↓	R^2↑	E_{VAR}↑	
SVM	4.7965±0.1029	9.1562±0.0986	0.2333±0.0008	0.3063±0.0129	3:32:29
MLP	4.6935±0.0580	7.5337±0.1695	0.4809±0.0349	0.4809±0.0316	0:09:40
RF	4.4583±0.4331	7.7830±0.6892	0.4460±0.1161	0.4472±0.1154	0:05:22
XGBoost	4.4523±0.1096	7.3646±0.1800	0.5040±0.0355	0.5043±0.0345	**0:03:30**
V-CNN	4.5856±0.0704	7.7921±0.0174	0.4447±0.0025	0.4474±0.0007	0:19:17
1D-CNN	4.6845±0.0207	8.0269±0.3666	0.4107±0.0684	0.4114±0.0685	0:11:23
Proposed method	**4.2869**±0.1085	**7.1066**±0.0532	**0.5438**±0.0068	**0.5442**±0.0031	0:11:42

↑ indicates the higher is better, while ↓ is in the reverse.

Table 4 Comparison of different methods on the test set.

Models	LoS prediction accuracy on the test set				*Average rank*
	MAE↓	*RMSE*↓	R^2↑	E_{VAR}↑	
SVM	4.2208±0.0493 (6)	6.2714±0.0146 (7)	0.1822±0.1462 (7)	0.3104±0.1532 (6)	6.50
MLP	4.2267±0.4696 (7)	5.9387±0.9289 (6)	0.2667±0.0683 (6)	0.2732±0.0067 (7)	6.50
RF	3.6029±1.2172 (2)	5.3722±2.2683 (2)	0.3999±0.2118 (4)	0.4229±0.2464 (4)	3.00
XGBoost	3.7359±0.1278 (3)	**5.2653**±0.2089 (1)	0.4236±0.0670 (2)	0.4325±0.0619 (2)	2.00
V-CNN	**3.5692**±0.7810 (1)	5.4056±0.6650 (3)	0.3925±0.1587 (5)	0.4267±0.1465 (3)	3.00
1D-CNN	4.0150±0.2825 (5)	5.6920±0.2679 (5)	0.4024±0.0975 (3)	0.4190±0.0815 (5)	4.50
Proposed method	3.8682±0.2107 (4)	5.5311±0.2810 (4)	**0.6039**±0.0412 (1)	**0.6042**±0.0377 (1)	2.50

From Tables 3 and 4 it can be seen that the proposed approach attains the R^2 of 0.5438 and 0.6039, and the E_{VAR} of 0.5442 and 0.6042 on the training set and the test set, respectively, which are the best values when compared with other algorithms. As for *RMSE*, the proposed approach attains 7.1066 and 5.5311 on the training set and the test set, respectively, which outperforms most competitors, except for the methods of RF and XGBoost. Both RF and XGBoost are ensemble learning (EL) methods comprised of multiple decision tree algorithms (70 in our situation). Furthermore, we applied the Friedman statistical test to compare the average ranking of different methods. Let r_j^i be the ranking of the *i*-th method on the *j*-th metric, and thus the average ranking can be calculated by

$$R_i = \frac{1}{N}\sum_j r_j^i \tag{6}$$

where *N* denotes the number of measure indicators. The null hypothesis of the Friedman test is that the performance of the compared algorithms is the same, i.e., there is no distinguishable difference amongst the methods. Mathematically, the Friedman test statistic is expressed as

$$F_F = \frac{(N-1)x_F^2}{N(k-1)-x_F^2}. \tag{7}$$

Here, F_F follows the Fisher-Snedecor *F* distribution with degrees of freedom *k*-1 and (*k*-1)(*N*-1), *k* represents the number of compared methods, and x_F^2 is computed as

$$x_F^2 = \frac{12N}{k(k+1)}\left[\sum_i R_i^2 - \frac{k(k+1)^2}{4}\right]. \tag{8}$$

There are a total of 7 algorithms evaluated using 4 different metrics in the experiments to compare, and thus the *k* and *N* are assigned as 7 and 4, respectively. With the degrees of freedom of *k*-1 and (*k*-1)(*N*-1), the critical value of $F(k\text{-}1,(k\text{-}1)(N\text{-}1))$ is obtained as $F(6,18)=2.66$ referring to the probability distribution table (test size α=0.05). Consequently, since the observed $\hat{F}_F = 9.0$ is greater than 2.66, the null hypothesis of the Friedman test is rejected suggesting a statistically significant difference between the algorithms exists. Based on pairwise comparisons, the proposed approach has outperformed all the methods compared except for the ensemble learning method of XGBoost. Even though the most prominent methods were adopted, the proposed approach has produced competitive advantages for the prediction of inpatient LoS. The LoS prediction accuracy metrics *RMSE* and E_{VAR} for different methods on the training set and test set are depicted in Fig. 6. Fig. 7 displays a comparison between the predicted values using our proposed approach and the actual values (left) as well as their respective distributions (right).

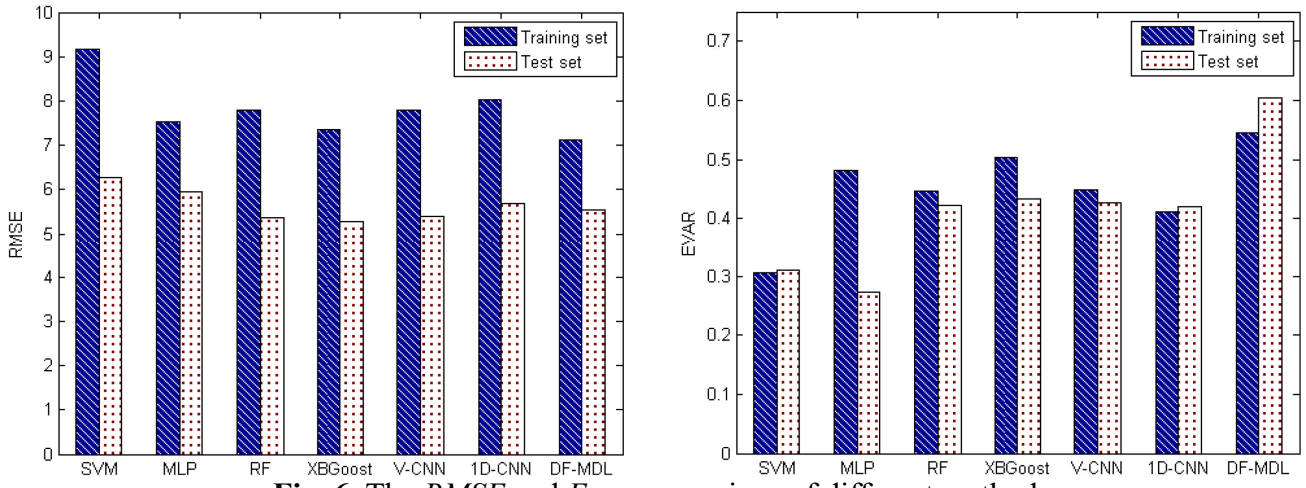


Fig. 6. The *RMSE* and E_{VAR} comparison of different methods.

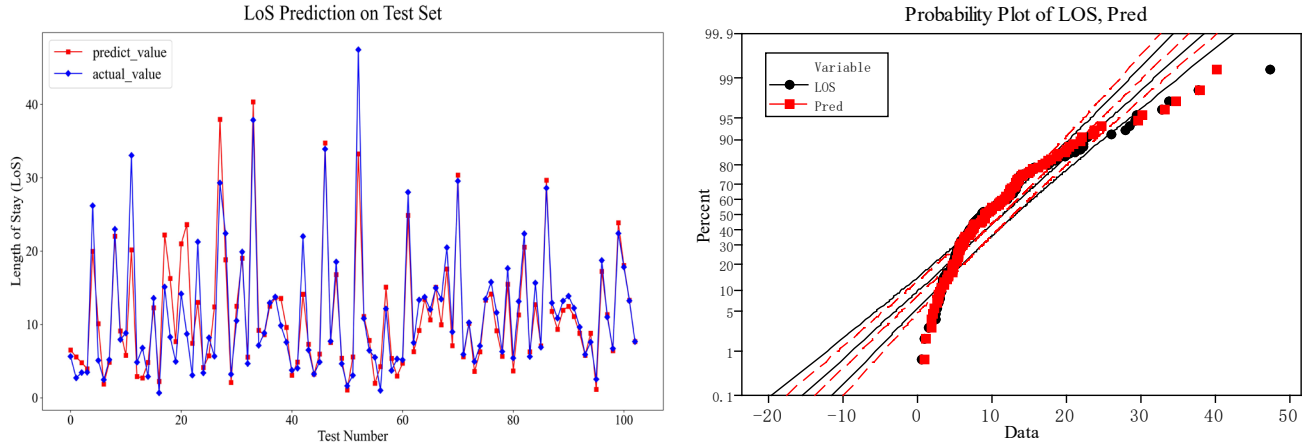


Fig. 7. Prediction performance of the proposed approach.

From Fig. 6 it can be visualized that the E_{VAR} values of the proposed method are the highest while the training *RMSE* is the lowest, which reveals the efficacy of the proposed method. Also, as shown in Fig. 7, except for some isolated cases, the plotted predicted values approximately coincide with the actual ones, and their distributions are very close, indicating the effectiveness of the proposed approach. On the other hand, the runtime of the proposed approach for 30 epochs of training is around 11 minutes, which is lower than the average runtime of competing methods. Thus, it can be concluded that the proposed approach exhibits competitive advantages in accuracy and efficiency relative to other state-of-the-art methods at predicting LoS. Moreover, we compared the performance of our methods to the results reported in the existing literature, as shown in Table 5. From Table 5 it can be seen that the proposed approach is on par and even outperforms most of the existing methods on the MIMIC-IV v1.0 dataset.

Table 5 Comparison results with recent literature [15].

#	References	Year	Description	$RMSE^{\downarrow}$	$R^{2\uparrow}$
1	Zimmerman et al. [45]	2006	Mean	7.23	0.00
2	Zimmerman et al. [45]	2006	Median	7.71	-0.14
3	Harutyunyan et al. [26]	2019	LSTM	6.61	0.28
4	Harutyunyan et al. [26]	2019	Multi-Channel LSTM (MC-LSTM)	6.20	0.26
5	Vaswani et al. [46]	2017	Transformer	6.18	0.27
6	Rocheteau et al. [41]	2021	Temporal Pointwise Convolution (TPC)	4.90	0.54
7	Al-Dailami et al. [15]	2022	Temporal Dilated Separable Convolution with Context-Aware Feature Fusion (TDSC-CAFF)	**4.30**	**0.64**
8	This study	2023	Data-Fusion Model (DF-Mdl)	5.53	0.60

4.5. Ablation study

An ablation study is further performed for our model to analyze the efficacy of the fused dataset. In the first ablation experiment, we estimate the effect of the CRXMDL using chest X-ray images to predict LoS. To this end, we remove the CRXMDL from the framework and use the model without X-ray images fed to perform LoS prediction. We notice a significant decrease in the test accuracy of the ablated model, where the *MAE* and *RMSE* increase to 4.3861 and 5.8599 (increase of 0.5179 and 0.3288), and the R^2 and E_{VAR} drop to 0.5554 and 0.5570 (decrease by 0.0485 and 0.0472), respectively. Though the efficacy of this ablated model is still better than that of the competing methods, it suffers a notable decline compared with that of the "full" DF-Mdl. This ablation experiment demonstrates that removing the CRXMDL significantly impacts the performance relative to the aggregated CRXMDL of the proposed approach. In the second ablation experiment, we remove the newly added modules from the proposed DF-Mdl architecture. Initially, we remove the optimized 1D SE block from the network to investigate the performance of the proposed approach. We notice a minor decrease in the result of the ablated model, where the test *MAE* and *RMSE* of the ablated model increase to 4.1525 and 5.7031 (increase of 0.2843 and 0.1720), respectively. Subsequently, the entire attention mechanism, including the enhanced 1D SE-block and SA module, is removed from the network. A significant drop in accuracy occurs in this ablation model. The test R^2 and E_{VAR} drop to 0.5514 and 0.5549 (decrease of 0.0525 and 0.0493), and the test *MAE* and *RMSE* rise to 4.1254 and 5.8861 (increase of 0.2572 and 0.355), respectively. This ablation experiment demonstrates that removing the attention mechanism has a significant impact on the performance compared to the enhanced 1D SE-block and SA module integrated into the proposed approach. Moreover, in the third ablation experiment, we investigate the performance and complexity of the model using the simple MSE loss function in place of the proposed loss function. We notice a significant decrease in the test accuracy of the ablated model, where R^2 and E_{VAR} drop to 0.5482 and 0.5485 (both decreased by 0.0557), while the *MAE* and *RMSE* increase to 4.3677 and 5.9072 (increase of 0.4995 and 0.3761), respectively. Further, the runtime of the ablated model is

12 minutes and is very close to that of the proposed method, which indicates that there is no increase in the complexity of the model using the proposed loss function. Thus, this ablation experiment demonstrates that using the customized loss function improves the performance of the model without increasing computational complexity. Table 6 summarizes the comparison results of the ablation experiments.

Table 6 The comparison results of ablation experiments.

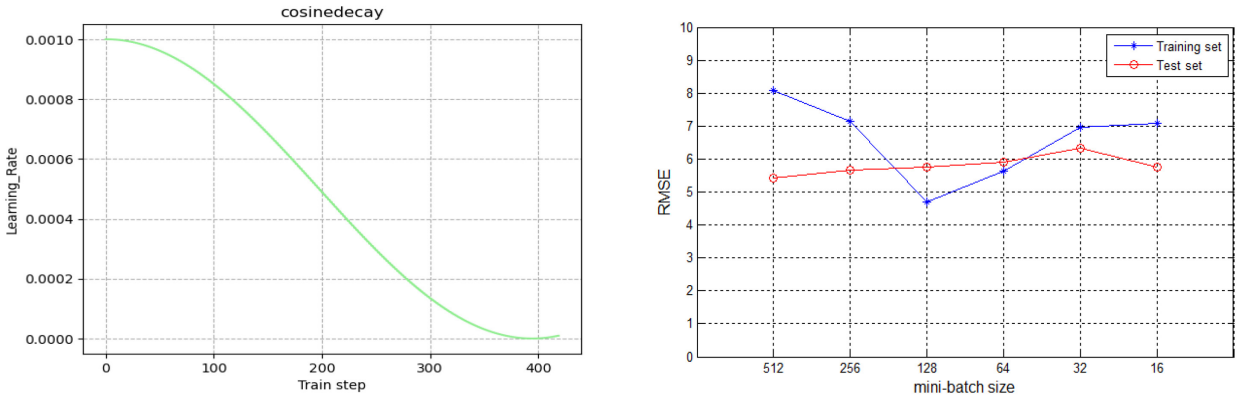
Ablation approach	LoS prediction accuracy on the training set				LoS prediction accuracy on the test set				Time (h:m:s)
	$MAE^{\downarrow}$	$RMSE^{\downarrow}$	$R^{2\uparrow}$	$E_{VAR}{}^{\uparrow}$	$MAE^{\downarrow}$	$RMSE^{\downarrow}$	$R^{2\uparrow}$	$E_{VAR}{}^{\uparrow}$	
Delete CRXMDL	4.3362	7.1641	0.5364	0.5390	4.3861	5.8599	0.5554	0.5570	0:01:23
Delete 1D SE-block	4.2439	7.2273	**0.5282**	**0.5299**	4.1525	5.7031	0.5789	0.5805	0:11:44
Delete SE+SA module	**4.1788**	7.1400	0.5396	0.5409	4.1254	5.8861	**0.5514**	**0.5549**	0:11:41
Replace the customized loss function with MSE	4.3001	7.1823	0.5341	0.5342	4.3677	5.9072	0.5482	0.5485	0:12:00
This study	4.2869	**7.1066**	0.5438	0.5442	**3.8682**	**5.5311**	0.6039	0.6042	0:11:49

4.6. Parameter Finetuning

In this section, we further study the effect of parameter finetuning on the proposed model. We finetune the essential hyperparameters, including the mini-batch size and the learning rate in the *Adam* optimizer and random search, based on cosine decay. The range of the mini-batch size hyperparameter is set as $(|B|) \in \{512, 256, 128, 64, 32, 16\}$. The initial learning rate is set as $(lr_0) = 0.001$, which drops to the minimum value of 0 following the cosine decay, i.e., the learning rate changes periodically according to the cosine function. The formula is expressed as follows

$$lr_s = 0.5 \cdot lr_0(1 + \cos(s \cdot \pi/S)) \tag{9}$$

where lr_0 is the initial learning rate, s refers to the current step, and S is the number of steps in which the learning rate decays to 0. We train our model using hyperparameters from these sets for 30 epochs on the publicly available MIMIC-IV v1.0 dataset with default splits. Fig. 8(a) plots the cosine descent curve of the learning rate. The prediction performance of the proposed method with different hyperparameter settings is summarized in Table 7, while Fig. 8(b) shows the changes in the *RMSE* metric. From Table 7, we can observe that a minor improvement is obtained when the mini-batch size is assigned as 512 along with the dynamic learning rate implemented by cosine decay. The *MAE* and *RMSE* on the test set drop to 3.7718 and 5.4183, which are a decrease of 0.0964 and 0.1128, respectively, compared to those before finetuning. Also, the R^2 and E_{VAR} rise to 0.6198 and 0.6260, i.e., an increase of 0.0159 and 0.0218, respectively, over the proposed method before finetuning. Other parameter settings do not lead to better effects. Therefore, these experimental results demonstrate that parameter finetuning can lead to slight improvements.



(a) Descent curve of learning rate (b) *RMSE* under different batch size level

Fig. 8. Fine-tuning of learning rate and mini-batch size hyperparameters.

Table 7 The fine-tuning performance of the proposed approach.

Batch-sizes	LoS prediction accuracy on the training set				LoS prediction accuracy on the test set				Time (h:m:s)
	$MAE^{\downarrow}$	$RMSE^{\downarrow}$	$R^{2\uparrow}$	$E_{VAR}{}^{\uparrow}$	$MAE^{\downarrow}$	$RMSE^{\downarrow}$	$R^{2\uparrow}$	$E_{VAR}{}^{\uparrow}$	
512	4.7631	8.0859	0.4095	0.4118	**3.7718**	**5.4183**	**0.6198**	**0.6260**	**0:03:11**
256	4.4664	7.1420	0.5393	0.5398	4.1088	5.6779	0.5826	0.5891	0:04:30
128	**3.2901**	**4.6877**	**0.8015**	**0.8085**	4.3601	5.7473	0.5723	0.5799	0:07:12
64	3.8645	5.6198	0.7147	0.7147	3.9988	5.9151	0.5470	0.5471	0:12:57
32	4.2196	6.9706	0.5612	0.5612	4.6905	6.3410	0.4794	0.4904	0:25:04
16	4.2817	7.0775	0.5476	0.5476	4.0982	5.7648	0.5697	0.5714	0:48:55

5. Discussion and conclusions

LoS prediction is a crucial task for hospitals to achieve and maintain resource efficiency and high level of treatment quality. Clinical practices in hospitals have generated diverse data modalities, including but not limited to, vitals, medical images, laboratory test results, clinical notes, etc. Utilizing this information effectively in a unified model can lead to more efficient resource allocation, higher diagnostic/prognostic accuracy, and better informed clinical decision making. For example, knowing the LoS days for a prospective patient would aid hospitals in predicting limitations in bed availability and also would allow hospitals to prioritize certain diagnostic testing in anticipation that a patient would otherwise have an extended hospital stay. To this end, this study proposed a novel data-fusion model to integrate heterogeneous clinical data for predicting inpatient LoS. Multi-modal data are utilized as part of the proposed method with the sub-models (basic models) established for individual data modalities. Specifically, a novel attention-embedded 1D convolutional neural network is developed to predict inpatient LoS using numerical data. By converting the text to sequence data, two long short-term memory networks are used for text-based LoS prediction. As for CXR image data, a convolutional neural network model, which we termed CRXMDL, is designed to predict LoS. Finally, these basic models are integrated to generate a unified data-fusion model, where two fully connected layers with

the neuron numbers of 64 and 32 are incorporated to change the vector dimensions, and a CL prediction layer is used for the final prediction of inpatient LoS. Experimental findings demonstrate the superiority of the proposed approach compared to other state-of-the-art methods. As the experimental materials used in this work are publicly accessible datasets with deidentified patient health information, this study poses neither privacy nor confidentiality concerns for the patients the data were collected from. Since the model solely relies on clinical variables, the risk of potential ethical concerns such as systematic biases are also reduced allowing for implementing the model across various public health organizations.

In our experiments, the proposed method has produced attractive results and proven to be promising. However, it is essential to recognize that the approach has certain limitations, leaving room for enhancements and refinement. First, the three predictive sub-models can be further enhanced by introducing more complex but efficient structures like octave convolution, spatial pyramid pooling, dilated and causal convolutions, etc. In our future work, we intend to explore other promising architectures to improve model performance. Second, despite promising performance, it is worth noting that the issue of LoS data skewness still requires more investigation. A potential solution is to combine statistical approaches that take the skewness into account. Incorporating such methods could further enhance the model's capability of better quantifying the statistical uncertainty associated with LoS prediction in lieu of the point prediction discussed in this paper. Lastly, due to concerns about transparency, accountability, and the perceived "black box" nature of deep learning algorithms, healthcare stakeholders, including clinicians, administrators, and patients, may express reservations about adopting deep learning techniques. We will focus on improving the transparency of models, increasing diversity in training data, and implementing fairness-aware techniques to make models more robust and interpretable, thereby fostering trust between medical professionals and algorithmic recommendations. In addition, we are committed to delivering a roadmap for future improvement that not only meets technical standards but also addresses emerging challenges in terms of legal compliance, ease of use, transferability, stability, safety, and security.

Acknowledgments

This research is partially funded by the National Science Foundation [IIS-2246158, DMS-2210929 and DUE-2216396]; Department of Education [#P116S210004]; and National Institute on Minority Health and Health Disparities (NIMHD) [#U54MD007592].

Conflict of Interests

The authors declare no conflicts of interest.

CrediT authorship contribution statement

Junde Chen: Conceptualization, Methodology, Validation, Data curation, Visualization, Writing-original draft. **Yuxin Wen**: Conceptualization, Supervision, Methodology, Writing – review & editing. **Michael Pokojovy:** Interpretation of data, Supervision, Methodology – statistical consulting, Writing – review & editing. **Tzu-Liang (Bill) Tseng**: Supervision, Writing – review & editing. **Peter McCaffrey:** Interpretation of data, Clinical interpretation, Writing – review & editing. **Alexander Vo**: Clinical interpretation and support. **Eric Walser**: Clinical interpretation and support. **Scott Moen**: Clinical interpretation and support.

References

[1] Levin, Scott, et al. "Machine-learning-based hospital discharge predictions can support multidisciplinary rounds and decrease hospital length-of-stay." *BMJ Innovations* 7.2 (2021).

[2] Churpek, Matthew M., et al. "Association between intensive care unit transfer delay and hospital mortality: a multicenter investigation." *Journal of Hospital Medicine* 11.11 (2016): 757-762.

[3] American Hospital Association. "AHA hospital statistics: fast facts on US hospitals." *American Hospital Association, available at: www/aha/org (accessed May 31, 2017)* (2017).

[4] Lapidus, Nathanael, et al. "Biased and unbiased estimation of the average length of stay in intensive care units in the Covid-19 pandemic." *Annals of Intensive Care* 10.1 (2020): 135.

[5] Ma, Xin, et al. "Length of stay prediction for ICU patients using individualized single classification algorithm." *Computer Methods and Programs in Biomedicine* 186 (2020): 105224.

[6] Huang, Shih-Cheng, et al. "Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines." *NPJ Digital Medicine* 3.1 (2020): 136.

[7] Turgeman, Lior, Jerrold H. May, and Roberta Sciulli. "Insights from a machine learning model for predicting the hospital Length of Stay (LOS) at the time of admission." *Expert Systems with Applications* 78 (2017): 376-385.

[8] Fu, Zhiwei, et al. "Predicting the Length of Stay of Patients in Hospitals." *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2021.

[9] Harerimana, Gaspard, Jong Wook Kim, and Beakcheol 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." *Journal of Biomedical Informatics* 118 (2021): 103778.

[10] Joze, Hamid Reza Vaezi, et al. "MMTM: Multimodal transfer module for CNN fusion." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020.

[11] Zellers, Rowan, et al. "Merlot: Multimodal neural script knowledge models." *Advances in Neural Information Processing Systems* 34 (2021): 23634-23651.

[12] Hayat, Nasir, et al. "Synthesizing the unseen for zero-shot object detection." *Proceedings of the Asian Conference on Computer Vision*. 2020.

[13] Hayat, Nasir, Krzysztof J. Geras, and Farah E. Shamout. "MedFuse: Multi-modal fusion with clinical time-series data and chest X-ray images." *Machine Learning for Healthcare Conference*. PMLR, 2022.

[14] Ngiam, Jiquan, et al. "Multimodal deep learning." *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*. 2011.

[15] Al-Dailami, Abdulrahman, Hulin Kuang, and Jianxin Wang. "Predicting length of stay in ICU and mortality with temporal dilated separable convolution and context-aware feature fusion." *Computers in Biology and Medicine* 151 (2022): 106278.

[16] Asri, Hiba, et al. "Big data in healthcare: challenges and opportunities." *2015 International Conference on Cloud Technologies and Applications (CloudTech)*. IEEE, 2015.

[17] Wang, Junwen, et al. "Reducing length of stay in emergency department: A simulation study at a community hospital." *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 42.6 (2012): 1314-1322.

[18] Awad, Aya, Mohamed Bader–El–Den, and James McNicholas. "Patient length of stay and mortality prediction: a survey." *Health Services Management Research* 30.2 (2017): 105-120.

[19] Tanuja, S., Dinesh U. Acharya, and K. R. Shailesh. "Comparison of different data mining techniques to predict hospital length of stay." *Journal of Pharmaceutical and Biomedical Sciences* 7.7 (2011).

[20] Vekaria, Bindu, et al. "Hospital length of stay for COVID-19 patients: Data-driven methods for forward planning." *BMC Infectious Diseases* 21.1 (2021): 1-15.

[21] Tong, Erica Y., et al. "Multi-site evaluation of partnered pharmacist medication charting and in-hospital length of stay." *British Journal of Clinical Pharmacology* 86.2 (2020): 285-290.

[22] Lequertier, Vincent, et al. "Hospital length of stay prediction methods: a systematic review." *Medical Care* 59.10 (2021): 929-938.

[23] Lu, Zhonghua, et al. “The effect of high-flow nasal oxygen therapy on postoperative pulmonary complications and hospital length of stay in postoperative patients: a systematic review and meta-analysis.” *Journal of Intensive Care Medicine* 35.10 (2020): 1129-1140.

[24] Thompson, Brandon, Karim O. Elish, and Robert Steele. "Machine learning-based prediction of prolonged length of stay in newborns." *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2018.

[25] Bednarski, Bryan P., et al. "Temporal convolutional networks and data rebalancing for clinical length of stay and mortality prediction." *Scientific Reports* 12.1 (2022): 21247.

[26] Harutyunyan, Hrayr, et al. “Multitask learning and benchmarking with clinical time series data.” *Scientific Data* 6.1 (2019): 96.

[27] Ma, Fei, et al. “Length-of-stay prediction for pediatric patients with respiratory diseases using decision tree methods.” *IEEE Journal of Biomedical and Health Informatics* 24.9 (2020): 2651-2662.

[28] Xu, Zhenhui, et al. "Predicting in-hospital length of stay: a two-stage modeling approach to account for highly skewed data." *BMC Medical Informatics and Decision Making* 22.1 (2022): 110.

[29] Muhlestein, Whitney E., et al. “Predicting inpatient length of stay after brain tumor surgery: developing machine learning ensembles to improve predictive performance.” *Neurosurgery* 85.3 (2019): 384-393.

[30] Hasan, Md Nahid, et al. "Prediction of Length-of-stay at Intensive Care Unit (ICU) Using Machine Learning based on MIMIC-III Database." *2023 IEEE Conference on Artificial Intelligence (CAI)*. IEEE, 2023.

[31] Boff Medeiros, Natália, et al. "Predicting the length-of-stay of pediatric patients using machine learning algorithms." *International Journal of Production Research* (2023): 1-14.

[32] Zhang, Dongdong, et al. “Combining structured and unstructured data for predictive models: a deep learning approach.” *BMC Medical Informatics and Decision Making* 20.1 (2020): 1-11.

[33] Soenksen, Luis R., et al. “Integrated multimodal artificial intelligence framework for healthcare applications.” *NPJ Digital Medicine* 5.1 (2022): 149.

[34] Szegedy, Christian, et al. “Inception-v4, Inception-Resnet and the impact of residual connections on learning.” *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 31. No. 1. 2017.

[35] Hu, Jie, Li Shen, and Gang Sun. “Squeeze-and-excitation networks.” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018.

[36] Naemi, Amin, et al. “Quantifying the impact of addressing data challenges in prediction of length of stay.” *BMC Medical Informatics and Decision Making* 21 (2021): 1-13.

[37] Esmaeili, Ashkan, and Farokh Marvasti. “A novel approach to quantized matrix completion using huber loss measure.” *IEEE Signal Processing Letters* 26.2 (2019): 337-341.

[38] Peter J. Huber. “Robust estimation of a location parameter.” *The Annals of Mathematical Statistics*, 35.1 (1964):73–101.

[39] Johnson, Alistair, et al. “MIMIC-IV.” *PhysioNet. Available online at: https://physionet. org/content/mimiciv/1.0/(accessed August 23, 2021)* (2020).

[40] Johnson, Alistair EW, et al. “MIMIC-CXR-JPG, a large publicly available database of labeled chest radiographs.” *arXiv preprint arXiv:1901.07042* (2019).

[41] Rocheteau, Emma, Pietro Liò, and Stephanie Hyland. “Temporal pointwise convolutional networks for length of stay prediction in the intensive care unit.” *Proceedings of the Conference on Health, Inference, and Learning*. 2021.

[42] Kingma, D., and J. Ba. “Adam: A method for stochastic optimization in: Proceedings of the 3rd International Conference for Learning Representations (ICLR’15).” *San Diego* 500 (2015).

[43] Pereira, Dulce G., Anabela Afonso, and Fátima Melo Medeiros. "Overview of Friedman’s test and post-hoc analysis." *Communications in Statistics-Simulation and Computation* 44.10 (2015): 2636-2653.

[44] Liu, Jie, and Yubo Xu. "T-Friedman test: a new statistical test for multiple comparison with an adjustable conservativeness measure." *International Journal of Computational Intelligence Systems* 15.1 (2022): 29.

[45] Zimmerman, Jack E., et al. “Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today’s critically ill patients.” *Critical Care Medicine* 34.5 (2006): 1297-1310.

[46] Vaswani, Ashish, et al. “Attention is all you need.” *Advances in Neural Information Processing Systems* 30 (2017).