

Received November 9, 2019, accepted December 15, 2019, date of publication January 17, 2020, date of current version January 27, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2967417

IDDSAM: An Integrated Disease Diagnosis and Severity Assessment Model for Intensive Care Units

ZHENKUN SHI^{ID 1,2,3}, WANLI ZUO^{ID 1,2}, SHINING LIANG^{ID 1,2}, XIANGLIN ZUO^{ID 1,2}, LIN YUE^{ID 5}, AND XUE LI^{ID 3,4}, (Member, IEEE)

¹Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun 130012, China

²College of Computer Science and Technology, Jilin University, Changchun 130012, China

³School of Information Technology and Electrical Engineering, The University of Queensland, St Lucia, QLD 4072, Australia

⁴Neusoft Institute of Information, Dalian Neusoft University of Information, Dalian 116081, China

⁵School of Information Science and Technology, Northeast Normal University, Changchun 130024, China

Corresponding authors: Xianglin Zuo (295224837@qq.com) and Zhenkun Shi (shizk14@mails.jlu.edu.cn)

This work was supported in part by the Nature Science Foundation of Jilin Province under Grant 20180101330JC and Grant 20190302029GX, in part by the Fundamental Research Funds for the Central Universities under Grant 2412017QD028, in part by the China Postdoctoral Science Foundation under Grant 2017M621192, in part by the Scientific and Technological Development Program of Jilin Province under Grant 20180520022JH and Grant 20190302109GX, and in part by the China Scholarship Council under Grant 201706170617.

ABSTRACT People are admitted to intensive care units (ICUs) because they need complete support for failing organ systems, constant monitoring, routine nursing care, and treatment. A critical or intensive illness is different from conventional or chronic diseases that most people are likely to have previously encountered. Such an illness is often unexpected and without warnings and can suddenly strike the previously fit. High levels of treatment and support are generally required to prevent life-threatening complications for the patients. Two of the most noticeable actions during an ICU stay are disease diagnosis and severity assessment of the patients. Unlike the majority of previous approaches where diagnosis and severity assessment are studied separately, we treat these actions as two tasks in an integrated procedure that clinicians must be able to quickly and accurately conduct such that patients are given the best possible chance for therapeutic success. In this paper, we propose an integrated disease diagnosis and severity assessment model (IDDSAM) to diagnose and assess diseases. Moreover, accompanying the prediction, we also provide an evidence-based explanation. IDDSAM is a multisource multitask model that is based on an attention mechanism and utilizes shareable information from laboratory tests, bedside monitoring, and complications to support patients' severity assessment and in-hospital disease diagnoses. We use 50,430 ICU cases consisting of 46,520 patients from 50 kinds of diseases over nine classifications to evaluate our proposed model. The experimental results demonstrated that our model outperforms the existing state-of-the-art mortality and diagnosis prediction framework by 3.79% on average in terms of accuracy for the mortality prediction tasks and by 14.51% on average for the diagnosis tasks.

INDEX TERMS Healthcare, data mining, disease diagnosis, mortality prediction, multisource multitask learning.

I. INTRODUCTION

The intensive care unit (ICU) is a special ward found in some hospitals, and a person is likely to be admitted to the ICU if they are in critical condition and require constant observation and specialized care. Intensive care refers to the specialized

treatment given to patients who are acutely unwell and require critical medical care [1]. The ICU is one of the most critically important operational environments in a hospital. To properly care for patients admitted to ICUs, clinicians need to quickly evaluate the severity and obtain the diagnosis in a remarkably short period.

In recent years, considerable effort has focused on establishing computer-aided systems or tools to reduce the burden

The associate editor coordinating the review of this manuscript and approving it for publication was Zhan Bu^{ID}.

of clinicians [1]–[9]. The traditional methods for this clinical practice are mainly focused on four groups: disease diagnosis, intensive care phenotype classification, forecasting length of stay, and mortality prediction [8]. Diagnosis is fundamental to the practice of medicine, and mastering diagnosis is central to both becoming and practicing as a doctor. Moreover, the diagnosis process is central to the practice of medicine and has, to date, received focused medical and computational science attention, where many have argued its importance [10], [11]. When used in the clinic or medicine, phenotype refers to the presentation of a disease, and a clinical phenotype would be the presentation of a disease in a given individual [12]. Forecasting the length of stay can benefit ICU resource management and reduce clinical costs. Mortality prediction for patients in the ICU is crucial for assessing the severity of illness and adjudicating the value of novel treatments, interventions, and healthcare policies.

Due to the lack of sufficient clinical data, the majority of the present works have studied these problems separately. In other words, they treat each of these clinical practices as an independent procedure. For example, in the task of disease diagnosis, clinicians and researchers have mainly focused on developing models to predict specific diseases. Jiri Polivka Jr *et al.* proposed predicting brain metastatic disease among breast cancer patients [13]. Patankar [4] attempted to detect breast cancer through a data mining approach. Long *et al.* diagnosed heart disease [14] by using the *IT2FLS* model. Nilashi *et al.* [15] used the neuro-fuzzy technique for hepatitis disease diagnosis. However, daily medical practices involve a complex mixture of scenarios and need different prediction models to address the hundreds to thousands of diseases [16]. Developing and deploying specialized models one by one is not the best approach.

Fortunately, in recent years, with the widespread adoption of Electronic Health Records (EHRs) in clinical practice, utilizing bedside data to conduct computer-aided diagnoses and evaluate the mortality risk of a patient has become possible. This will significantly benefit the ICU disease diagnosis, mortality prediction, patient care, and community services. Moreover, the human body as organic entities and different systems are closely connected, and no diseases are isolated. Therefore, it is feasible to develop a unified model to address these problems together.

Taking the above concerns into consideration, in this paper, we proposed an integrated disease diagnosis and severity assessment model (IDDSAM) to conduct ICU disease diagnoses with severity evaluations. We treated disease diagnosis and mortality prediction as an interrelated process in clinical practice. In an integrated view, severity assessment and the disease diagnosis are provided to the intensivists at the same time. We use multisource multitask attention [17] techniques in our model. Here, the sources come from different clinical measurements and medical treatments, and the tasks refer to disease diagnosis and mortality prediction. The detailed description will be provided in the **Problem Definition** section. Based on our previous work for disease

diagnosis [1], [18], this work is the first time utilizing a shared vector space among different tasks containing both disease diagnosis and mortality prediction to enhance healthcare performance. Meanwhile, we also provide explanations of how the results are obtained. In this way, further actions can be taken more faithfully because the clinicians are familiar with the prediction process.

In this paper, we address diagnosis and mortality prediction as a combination of a unified procedure. In the view of aggregate health data streams, we incorporate bedside monitoring, real-time diagnosis, and spatial clinical treatment together and treat disease diagnosis and illness severity prediction as a progressive process. In other words, we provide diagnosis and mortality prediction at every time window, and the clinicians can customize the result.

- **An Integrated Perspective for Disease Formulation.** We formulated the ICU disease diagnosis and mortality prediction as a unified multisource and multitask learning problem, where sources correspond to medical treatment and clinical measurements and the tasks correspond to disease diagnosis and mortality prediction. Our model is able to handle different kinds of diseases over all disease categories and provide mortality risk along with the diagnosis in a straightforward manner.
- **Real-time Diagnosis and Mortality Prediction.** We treat the disease diagnosis and the mortality prediction as a gradual process over the observations along with the sequential measurement and treatment accompanied by the complications.
- **An End-to-end Model for Disease Diagnosis and Mortality Assessment.** IDDSAM is an explicitly designed model that integrates window alignment, input embedding and attention mechanisms with focal loss techniques.
- **Comprehensive Experimental Analysis of the Proposed Model.** We conduct our experiment on a real-world MMIC-III benchmark dataset on 50 diseases over nine categories, which covers most commonly diagnosed diseases. The results demonstrate that our method is effective, competitive, and can achieve state-of-the-art performance.

II. RELATED WORK

A computer-aided diagnosis system provides an assessment of a disease using clinical information or in combination with other relevant diagnostic data and is used by clinicians as decision support in developing their diagnoses [19]. In the ICU scenario, automatic disease diagnosis prediction using the available clinical data can support clinicians in making quick decisions such that they can take further actions to save lives. In recent years, many researchers have worked on different methods [16], [20], [21] to predict different kinds of diseases, such as brain metastatic disease [13], heart disease [14] and sepsis [22]. Existing disease prediction methods can be roughly divided into two categories: clinical-based diagnosis [13], [21], [23] and data-based

diagnosis [14], [16], [24], [25]. Most existing clinical-based diagnoses require a profound knowledge of medicine, and most of them are focused on a certain field, such as specific diseases caused by specific germs [26]. Until the past few years, most of the techniques for computer-aided disease diagnosis were based on traditional machine learning and statistical techniques such as logistic regression [3], support vector machines (SVMs) [27], random forests (RFs) [28] and decision tree (DT) [2], [29], [30]. Recently, deep learning techniques have achieved great success in many domains through deep hierarchical feature construction and capturing long-range dependencies in an effective manner [31]. Given the rise in the popularity of deep learning approaches and the increasingly vast amount of clinical electronic data, there has also been an increase in the number of publications applying deep learning to disease diagnosis tasks [15], [16], [32], [33], which yield better performance than traditional methods and require less time-consuming preprocessing and feature engineering.

Mortality risk prediction has a long history in the medical domain, where life tables and statistical inference have been used to predict life expectancy for patients [34]. Most of the existing mortality prediction methods are based on scoring systems, as we mentioned in the introduction. A recent study [35] determined that up to the end of 2012, only approximately 10%-15% of US ICU patients used these types of scoring systems. More recently, RNNs have provided new effective paradigms for us to enable end-to-end learning from massive data. Harutyunyan *et al.* [8] and Song *et al.* [36] used LSTM and an attention model, respectively, to predict in-hospital mortality and provide state-of-the-art performance.

However, all of these studies treat disease diagnosis and mortality prediction as a self-governed task, and all of these methods are designed for a specific disease based on either the intensive use of domain-specific knowledge or the advantage of advanced statistical methods. Specifically, studies have been conducted on Alzheimer's disease [25], heart disease [14], chronic kidney disease [37], diabetes mellitus [38], and abdominal aortic aneurysm [39]. Moreover, these models have been developed to anticipate needs and focused on specialized predictive models that predict a limited set of diseases. However, the day-to-day clinical practice involves an unscheduled and heterogeneous mix of scenarios and needs hundreds to thousands of different prediction models. It is impractical to develop and deploy specialized models one by one [16]. Therefore, it is important to develop a unified model that can be applied for the majority of diseases and that can provide the mortality risk with the disease diagnosis. This is an elegant application of multitask learning, and each disease can be treated as a single learning task. Moreover, the diagnosis and the mortality prediction can also be treated as different learning tasks. Note that many approaches to multitask learning (ML) in the literature address a similar setting: they assume that all tasks are associated with a single output, e.g., the multiclass MNIST

dataset is typically cast as 10 binary classification tasks. More recent approaches address a more realistic, heterogeneous setting in which each task corresponds to a unique set of outputs [40]. We cannot simply apply their approaches to our situation because multiple clinical observations and multiple medical treatments cannot be integrated into the existing frameworks.

By considering the aforementioned problems, in this paper, we propose IDDSAM to simultaneously diagnose the disease and predict the mortality. IDDSAM is a multisource multitask model. A significant advantage of multisource multitask learning [41] is that multiple sources can effectively increase the sample size that we are using to train the model (because the samples of some kinds of disease are very small and not enough for learning; see Table 1). In IDDSAM, we first identified meaningful patient cohorts by the International Classification of Diseases (ICD) code and then used these cohorts as learning tasks. All tasks share a common relevant feature subset (e.g., temperature, gender, weight, and respiratory rate) with a different weight parameter on each task and use a unique feature subset (e.g., inosinic acid, hematocrit, PH, and oxygen concentration) for different tasks. To handle time between different sources, a window-alignment operation is conducted before learning. In addition, we use two attention layers to capture the correlations. Finally, we use a gated recurrent unit (GRU) to fuse the above-selected features from each modality to estimate multiple regression and classification variables.

III. PROPOSED FRAMEWORK

A. PROBLEM STATEMENT

For a given ICU stay length of T hours and a collection of diagnostic results $R_t(t \in T)$ with a collection of mortality labels $M_t(t \in T)$, we defined the clinical observation as follows:

$$O(t) = \begin{cases} \langle R_t, M_t \rangle, & \text{if } R_t \notin \emptyset \text{ and } M_t \notin \emptyset \\ \langle R_t, 0 \rangle, & \text{if } R_t \notin \emptyset \text{ and } M_t \in \emptyset \\ \langle 0, M_t \rangle, & \text{if } R_t \in \emptyset \text{ and } M_t \notin \emptyset \\ \langle 0, 0 \rangle, & \text{otherwise} \end{cases} \quad (1)$$

where $O(t)$ is the vector of bedside observations at time t . $O(t) = P_a^i \Theta Q_b^j$, where P_a^i represents the i -th clinical measurement at time a , Q_b^j represents the j -th medical treatment at time b , Θ is a window alignment operation between P_a^i and Q_b^j , R_t represents the diagnostic result at time t , and M_t is the mortality risk at time t . Our objective is to generate a sequence-level disease prediction at each sequence step. The type of prediction depends on the specific task and can be denoted as a discrete scalar vector R_t^i for the multitask classification. As all tasks are at least somewhat noisy, when training a model $Task_i$, we expect to learn a good representation for $Task_i$ that ideally ignores the data-dependent noise and generalizes well. By sharing representations between related tasks, we can enable our model to generalize better on our primary task.

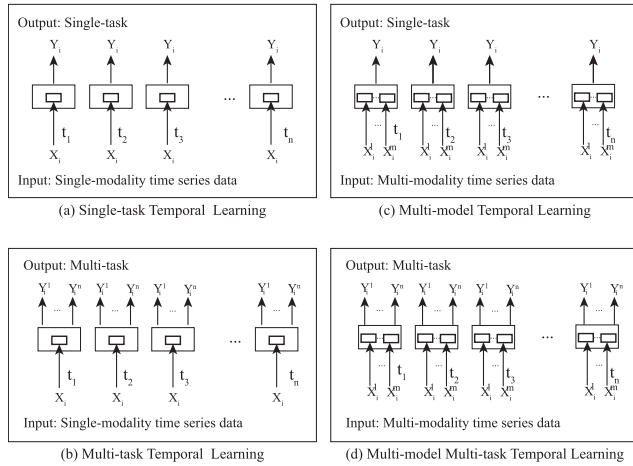


FIGURE 1. The framework of multimodal multitask temporal learning.

B. MULTIMODAL MULTITASK TEMPORAL LEARNING FRAMEWORK FOR TEMPORAL DATA

Inspired by Daoqiang Zhang and Dinggang Shen's work [25], we treat disease diagnosis and mortality prediction as a unified sequential multimodal multitask (SM3T) learning problem. Multimodal refers to the clinical measurements and medical treatments. The tasks represent the diagnosis and mortality prediction. The framework can simultaneously learn multiple tasks from multimodal temporal data. Fig. 1 illustrates the proposed SM3T method and a comparison with the existing learning methods.

Fig. 1(a) is single-modality single-task temporal learning. Each subject has only one modality of data represented as x_i at each time step, and each subject corresponds to only one task denoted as Y_i . This is the most commonly used learning method. Fig. 1(b) is single-modality multitask temporal learning. The input is similar to single-task temporal learning, but each object corresponds to multiple tasks denoted as $Y_i^1, Y_i^2, \dots, Y_i^n, n > 1$. Fig. 1(c) is multimodality single-task temporal learning. Each subject has multiple modalities of data represented as $x_i^1, x_i^2, x_i^3, \dots, x_i^n, n > 1$ at each time step, and each subject corresponds to only one task denoted as Y_i . Fig. 1(d) is multimodality multitask temporal learning. Each subject has multiple modalities of data represented as $x_i^1, x_i^2, x_i^3, \dots, x_i^n, n > 1$ at each time step, and each subject corresponds to multiple tasks denoted as $Y_i^1, Y_i^2, Y_i^3, \dots, Y_i^n, n > 1$.

Similar to Zhang and Shen [25], we can formally define the SM3T learning as follows. Given N training subjects over T time span and each with M modalities of data, represented as:

$$x_i^t = \{x_i^t(1), x_i^t(2), \dots, x_i^t(m), \dots, x_i^t(M)\},$$

$$i = 1, 2, \dots, N \quad (2)$$

our SM3T method jointly learns a series of models corresponding to Y different tasks denoted as:

$$Y_i = \{y_i^t(1), y_i^t(2), \dots, y_i^t(j), \dots, y_i^t(Y)\},$$

$$j = 1, 2, \dots, N \quad (3)$$

Note that SM3T is a general learning framework, and we implement it through an attention framework as shown in Fig. 2. We will provide detailed descriptions for each action in the SM3T framework.

C. WINDOW ALIGNMENT

The framework contains multiple data sources, including medical treatments and clinical measurements. Medical treatments influence clinical measurements; however, medical treatments generally take some time to take effect. Therefore, it is inevitable that the prediction performance will be slowed by using the clinical measurements and the medical treatments at the same action time. Consequently, how to align the time window for when medical treatment was administered and the window when the clinical measurements were taken becomes vital. For example, a patient P_a developed a fever at time t_0 , and the body temperature at t_0 was bt_0 . To treat hypothermia, the doctor administered P_a some aspirin marked as mt_1 at t_1 , and at t_1 , the body temperature is bt_1 ; here, $bt_1 = bt_0$ because no action was taken before t_1 . Later, at t_2 , the temperature decreased from bt_1 to bt_2 . Under normal circumstances, we made an observation at t_1 and obtained two actions: clinical measurement bt_1 and medical treatment mt_1 . Then, we sent these features into a predictive model at the same time window t_1 . Clearly, this contradicts the common sense because mt_1 leads to bt_2 rather than bt_1 .

To solve this problem, in this work, we add a window alignment operation. Assume that $A_p^{k\circ}t_i$ represents the k -th clinical measurement for patient p at time step t_i and that $B_p^{k\circ}t_j$ represents the k -th medical treatment for patient p at time step t_j ; then, we can obtain a feature vector of the n -th disease, Φ^n . For Φ^n , in this work, we simply join A_p and B_p :

$$\Phi^n = \{A_p^{k\circ}t_i; B_p^{k\circ}t_j\} \quad (4)$$

At time t , Φ^n can be denoted as $\Phi_t^n = \phi_t$. Then, we can define the window alignment operation as follows:

$$\phi_t = w_1 A_p t_i + w_2 B_p t_j + b \quad (5)$$

where ϕ_t is the input feature vector for patient p at time window t . w_1 , w_2 , and b are learnable parameters. The time steps t_i for clinical measurements and t_j for medical treatments are aligned by mapping $A_p t_i$ and $B_p t_j$ to a unique time step ϕ_t .

This produces an acceptable result for our purposes. Moreover, in the same time window ϕ_t , according to the results of the experiments, compared with t_i , t_j generally has a time delay, which is consistent with prevailing medical sense.

D. DENSE LAYER

To balance the computational cost with prediction performance, the dimensionality of the data must be reduced before it is transferred to the next step in the process. Typically, the embeddings are simply concatenated at every step in the sequence. However, clinical features almost always suffer

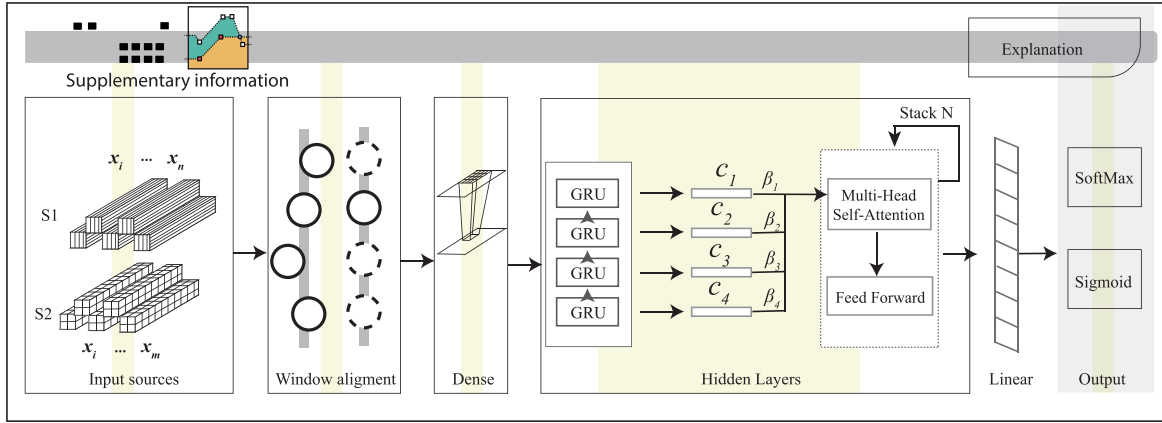


FIGURE 2. IDDSAM: an attention-based multisource and multitask prediction framework.

from the “curse of dimensionality”, which means that the representations are not suitable for learning or inference. Inspired by Trask et al.’s [42] work in natural language processing (NLP) and Song et al.’s [36] work in clinical data processing, we added a dense layer to unify and flatten the input features while retaining the information useful for interpretability. To prevent overfitting, we set the dropout to 0.30 in this paper.

E. THE GATED RECURRENT UNIT LAYER

The GRU layer takes the sequence of action $\{x_t\}_{t \geq 1}^T$ from the previous dense layer and then associates the p -th patient with a disease class label vector Y , and a mortality label vector Z along with the time span denotes the class label for the p -th patient with the n -th disease at time T . $Y_p^n(t)$ is set as follows:

$$Y_p^n(t) = \begin{cases} \text{diseaseID}, & \text{if diagnosis recorded at time } t \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

$$Z_p^n(t) = \begin{cases} 1, & \text{if patient alive at time } t \\ 0, & \text{otherwise.} \end{cases} \quad (7)$$

We created two T-dimensional response vectors for the p -th patient:

$$Y^{(p)} = (y_{p,1}, y_{p,2}, \dots, y_{p,p_t})^T \quad (8)$$

$$Z^{(p)} = (z_{p,1}, z_{p,2}, \dots, z_{p,p_t})^T \quad (9)$$

For the diagnosis of ICU patients, we adopted GRU and represent the posterior probability of the outcome that patient p has y -th disease as follows:

$$Pr[P_y^n(t) = 1 | \phi_h^p(t)] = \sigma(\omega^{(p)T} \phi_h^p(t)) \quad (10)$$

where $\phi(a)$ is the sigmoid function $\sigma(a) \equiv (1 + \exp(-a))^{-1}$ and $\omega^{(p)}$ is an $\alpha + \beta$ -dimensional model parameter vector for the p -th patient. Similar to diagnosis, the mortality prediction task’s posterior probability is:

$$Pr[P_z^n(t) = 1 | \phi_h^p(t)] = \text{Softmax}(\omega^{(p)T} \phi_h^p(t)) \quad (11)$$

To learn the mutual information of data resulting from the customization, we model for all diseases jointly such that we

can share the same vector space across the disease, which is very useful for diseases with fewer samples. We represent the trainable parameters of the GRU as $(Sa + Sb) \times TW \equiv [\omega^1, \omega^2, \dots, \omega^T]$.

F. MULTIHEAD ATTENTION AND FEED FORWARD

This attention layer is designed to capture the dependencies in the entire sequence. In ICUs, actions closer to the current position are more critical than those farther away. Additionally, only information in positions before the current position needs to be analyzed. Inspired by Vaswani et al. [17], we chose to use H-head attention to create multiple attention graphs. The resulting weighted representations are concatenated and linearly projected to obtain the final representation. Moreover, we also added several 1D convolutional sublayers with a kernel size of 2. Two of these 1D convolutional sublayers are used internally with rectified linear unit (ReLU) activation in between. Residual connections are used in these sublayers. In contrast to Song et al. [36] and Harutyunyan et al. [8], who only make mortality predictions once after a specific timestamp, DIMM makes a prediction with an interpretation at each timestamp. This is more helpful for ICU clinicians because they also need to know a patient’s risk of mortality at all times. The attention modules are stacked N times, and the final representations are used in the mortality risk prediction model.

G. LINEAR, SIGMOID AND SOFTMAX LAYERS

The linear layer is designed to obtain the logits from the unified output of the attention layer. The activation function used in this layer is ReLU. The last layer is preparing for the output based on different tasks. We use softmax to classify the different diseases, and sigmoid to mortality prediction. The loss function is cross-entropy (CE) with L_2 regularization:

$$\mathcal{L}_{ce} = -\frac{1}{N} \sum_{k=1}^K \sum_{i=1}^{N_k} c_k [y_i^k \log(\hat{y}_i^k) + (1 - y_i^k) \log(1 - \hat{y}_i^k)] + \lambda \|W\|_2, \quad (12)$$

where y_i^k and \hat{y}_i^k are the ground truth and prediction for the k -th class, respectively. Note that there are only two classes in

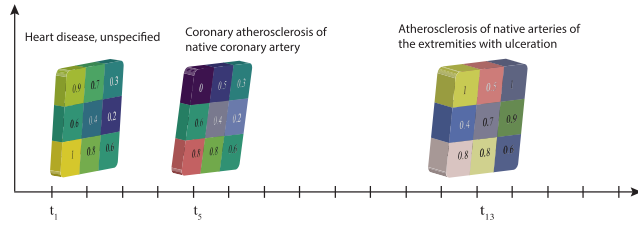


FIGURE 3. Showcase of result explanations.

mortality prediction and 50 classes in disease diagnosis; thus, the class distribution is generally imbalanced. In other words, positive observations (i.e., death records) are much fewer than negative ones. $\|W\|_2$ norm is the overall network weight, and λ helps tune the regularization strength.

Similar to Lin et al.'s work [43] and our previous work [1], focal loss is also introduced. The final state of the loss function can be defined as:

$$\mathcal{L} = (1 - e^{\mathcal{L}_{ce}})^{\gamma} \mathcal{L}_{ce}, \quad \gamma > 0 \quad (13)$$

where γ is a focusing parameter that smoothly adjusts the rate at which easy examples are down-weighted. When $\gamma = 0$, FL is equivalent to CE, and as γ is increased, the effect of the modulating factor is likewise increased.

H. EXPLANATION OF RESULTS

Explanations accompany the prediction results. We observed that a large number of patients in ICUs are diagnosed with 2-15 diseases (see Fig. 4), and we add related complications to the explanation process. In addition, we also add some supplementary information to the whole sequence, such as output events and clinical notes. All this information has interacted with the learning process. As shown in Fig. 3, different colors represent different clinical measurements or different medical treatments, the number on the small cube represents the contribution to the prediction, and the thickness of the large cube represents the mortality rate. With the abundance of clinical measurements and clinical treatments, we find that the diagnosis is more accurate and more specific. For example, the initial diagnosis at t_1 is unspecified heart disease, which is a very general diagnosis, but the diagnosis at t_5 , “coronary atherosclerosis of native coronary artery”, is more specific, and the final diagnosis “atherosclerosis of native arteries of the extremities with ulceration” at t_{13} is more accurate and contains more useful information. By providing this information to clinicians, they will be more likely to trust the prediction results.

Moreover, we observed that a large number of patients in ICUs are diagnosed with 2-15 diseases (see Fig. 4).

IV. EXPERIMENTAL

A. DATA DESCRIPTION

We use MIMIC III¹ to evaluate our proposed approach. MIMIC-III is a large open-access dataset of deidentified

¹Data available at <https://mimic.physionet.org/>

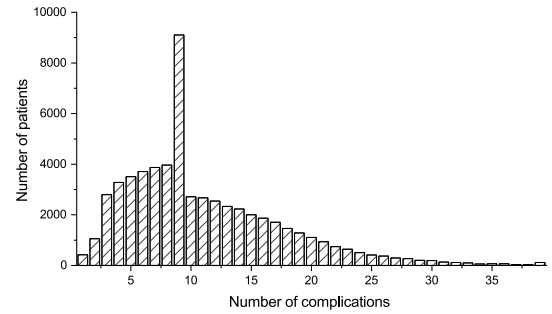


FIGURE 4. Complication distribution of patients in Medical Information Mart for Intensive Care (MIMIC-III).

patient records, and it has been widely used by 845 publications as of the end of August 2019. The data in MIMIC-III are associated with 46,520 distinct adult patients admitted to the ICU between 2001 and 2012, a total of 12 years [44]. By using MIMIC-III, researchers are able to reproduce and improve their studies through the open-source communities. In this study, each ICU stay was treated as an independent admission to acquire more samples. Table 1 presents a detailed description of the prediction tasks in our experiment. In our research, we use 50,430 ICU cases of 9 categories over 50 types of diseases as our data source. We grouped these samples by International Classification of Diseases (ICD) code, ICD-9 (version 2014). As shown in Fig. 4, most of the patients have multiple complications. In this study, all complications were collected based on the clinical monitoring process. We do not filter any patients, which is unlike the existing works. For clinical measurements, we obtained 129 features. We select all the features used in existing ICU scoring systems, and then we add some frequently occurring features. For medical treatments, we obtained 50 features. We select the top 50 features that occurred in both CareVue and MetaVision.

B. EXPERIMENTAL SETTINGS

Similar to our previous work [1], our experiment included over 40,000 patients. We first grouped all the ICU stay cases based on their diagnosed disease, and then we selected those groups that contained more than 1000 samples and obtained 50 different types of diseases. Based on the ICD9 manual, these samples are involved in nine categories. We treat the disease diagnosis as a multiple classification problem. We set the outcome as “true” if the prediction result is consistent with the labels during the diagnosis time window; otherwise, it is set as “false.” During the training procedure, results are only given if there are observations during this time window. Conversely, during the test procedure, we can provide a diagnosis at every time step or according to the customization of clinicians. The learning rate in this experiment is 0.001, and $\gamma = 2$. The epoch size that we set in this experiment is 30. The batch size in this experiment is 32, also with the ADAM optimizer. The dropout is set to 0.35. According to our experiment, the best performance achieved for the attention stack is 4. We fixed the training set, validation set,

TABLE 1. Detailed description of IDDSAM tasks for disease diagnosis and mortality prediction.

Category	ICD 9	Title	SampleSize	age	Mrate
Infectious, Parasitic	008.45	Clostridium difficile	2,672	69.07±24.31	64.09%
	038.9	U septicemia	5,787	69.11±32.13	65.23%
Neoplasms	197.0	Lung Bronchus	866	62.23±13.31	85.49%
	197.7	Liver, specified as secondary	926	64.63±17.47	86.36%
	198.5	Bone and bone marrow	984	63.59±12.77	86.85%
Endocrine, Nutritional, Metabolic Immunity	250.0	Diabetes mellitus no mention of complication	10,585	71.40±28.41	44.09%
	250.4	Diabetes with renal manifestations	1,574	69.26±20.04	63.75%
	250.6	Diabetes with neurological manifestations	1,793	70.02±26.25	51.32%
	263.9	Other and U protein–calorie malnutrition	2,258	65.95±26.35	56.95%
Blood, Blood-forming organs	280.0	Iron deficiency anemia 2 to blood loss C	1,346	68.34±25.88	56.24%
	280.9	Iron deficiency anemia, unspecified	1,992	67.38±39.21	37.53%
	285.1	Acute posthemorrhagic anemia	6,998	69.10±36.81	39.55%
	285.21	Anemia in chronic kidney disease	2,616	66.70±28.35	55.25%
	285.29	Anemia of other chronic illness	2,225	67.45±32.21	54.17%
	285.9	Anemia unspecified	8,253	67.90±34.13	39.77%
Circulatory	397.0	Diseases of tricuspid valve	1,286	77.26±40.76	55.19%
	401.9	Hypertension, unspecified	23,153	71.27±32.66	36.19%
	403.90	Hypertensive C kidney, U, stage I to IV	4,712	81.32±45.61	45.78%
	403.91	Hypertensive C kidney, U, stage V	3,756	65.27±19.49	67.80%
	410.71	Subendocardial infarction	4,474	74.17±30.51	50.26%
	411.1	Intermediate coronary syndrome	2,200	69.42±22.56	24.72%
	412	Old myocardial infarction	4,479	74.93±36.99	41.76%
	413.9	Other and unspecified angina pectoris	1,468	70.64±27.84	23.11%
	414.0	Coronary atherosclerosis	2,415	78.53±37.30	52.97%
	414.01	Of native coronary artery	14,585	73.24±32.09	34.91%
	414.8	Ischemic heart disease, chronic, other	1,526	74.54±28.52	57.68%
	431	Intracerebral hemorrhage	1,561	69.71±28.83	57.35%
	433.10	Occlusion and stenosis of carotid artery	1,109	75.77±30.39	37.48%
	434.91	Cerebral artery occlusion, U	907	69.41±28.22	53.09%
Respiratory	482.41	Methicillin susceptible pneumonia	1,297	64.56±22.81	64.61%
	486	Pneumonia, organism unspecified	7,779	68.51±32.89	58.3%
	491.21	Obstructive C bronchitis with exacerbation	1,851	72.91±24.79	66.03%
	493.20	Chronic obstructive asthma, unspecified	1,215	69.22±26.13	45.30%
	493.90	Asthma,unspecified type, unspecified	2,781	59.18±30.16	26.57%
Digestive	571.2	Alcoholic cirrhosis of liver	1,529	55.93±12.54	61.02%
	571.5	Cirrhosis of liver without mention of alcohol	1,820	60.29±16.73	59.80%
Genitourinary	584.5	Acute kidney failure with lesion	3,567	65.98±24.11	61.29%
	584.9	Acute kidney failure, unspecified	3,564	71.45±36.21	55.58%
	585.6	End stage renal disease	2,720	62.39±20.38	60.15%
	585.9	Chronic kidney disease, unspecified	4,942	79.01±41.90	50.21%
	600.00	Hypertrophy of prostate II	1,850	79.81±35.58	37.93%
Conditions originating in the perinatal period	765.18	Other preterm infants, 2,000-2,499 grams	621	0.03±0.03	0.22%
	765.19	Other preterm infants, 2,500 grams and over	557	0.02±0.02	0.04%
	765.27	33-34 completed weeks of gestation	545	0.04±0.03	0.11%
	765.28	35-36 completed weeks of gestation	642	0.02±0.02	0.23%
	769	Respiratory distress syndrome in newborn	511	0.10±0.09	3.06%
	770.6	Transitory tachypnea of newborn	535	0.02±0.03	0.13%
	770.81	Primary apnea of newborn	331	0.10±0.08	0.19%
	774.2	Neonatal jaundice with preterm delivery	1,021	0.08±0.08	1.11%
	774.6	Unspecified fetal and neonatal jaundice	514	0.02±0.04	0.47%

*Mrate stands for mortality rate; U stands for unspecified; C stands for chronic.

*ICD9: stands for ICD-9-CM Diagnosis Codes (version 2014).

and test set to obtain relatively fair results for all the baselines, and the detailed information is listed in Table 2. Because the samples in the category of “Conditions originating in the perinatal period” are limited and the mortality rate is lower than other categories, we cannot obtain sufficient data for training; thus, we do not list the mortality prediction results here, and future work can focus on this point.

C. COMPARED METHODS

In this experiment, we use six methods as our baselines: logistic regression (LR) with L2 regularization, support vector machine (SVM), decision tree (DT), random forest (RF), GRU, and the state-of-the-art LSTM-based method [6] for both the diagnosis task and mortality prediction task. Because the results are similar, we only listed the best of the top two in our paper for each of the tasks. For the disease diagnosis task, the top two compared methods are RF and the state-of-the-art multitask channel-wise LSTM (MWLSTM). For the task of mortality prediction, the best two methods are SVM and MWLSTM. We also compared the mortality prediction tasks with the existing ICU score systems, and the result can be found in the supplementary.

D. EVALUATION METRICS

To provide a comparison among the aforementioned techniques, three evaluation techniques were used in the task of disease diagnosis, F1-measure, accuracy, and recall, and three evaluation methods are used in the task of mortality prediction, area under the receiver operating characteristic curve (AUROC) [45], area under precision-recall curve (AUPRC) [46], and accuracy. These evaluation techniques are defined as follows:

$$\text{Accuracy} = \frac{TN + TP}{FP + TP + FN + TN} \quad (14)$$

$$\text{F1-Measure} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Precision} + \text{Recall}} \quad (15)$$

$$\text{Recall} = \frac{TP}{FN + TP} \quad (16)$$

where TP and FP are the numbers of true positives and false negatives, respectively.

E. EXPERIMENTAL RESULTS AND DISCUSSION

Table 3 and Tab. 4 present the results of the disease diagnosis task and mortality prediction task, respectively. As shown, our model significantly outperformed the baseline methods in most of the tasks. Due to the cohort selection, we did not exclude any patients; thus, the prediction accuracy varies from 55.23% to 93.81%.

In the majority of the tasks, our model achieved the best performance: 129/150 in disease diagnosis, and 81/123 in mortality prediction. We find that the number of samples greatly influences the performance of disease diagnosis; more samples result in better performance. As mentioned

TABLE 2. Experimental settings for training, validation and test.

Task	Train	Validation	Test
008.45	1870	534	268
038.9	4,050	1,157	580
197.0	606	173	87
197.7	648	185	93
198.5	688	196	100
250.00	7,409	2,117	1,059
250.40	1,101	314	159
250.60	1,255	358	180
263.9	1,580	451	227
280.0	942	269	135
280.9	1,394	398	200
285.1	4,898	1,399	701
285.21	1,831	523	262
285.29	1,557	445	223
285.9	5,777	1,650	826
397.0	900	257	129
401.9	16,207	46,30	2,316
403.90	3,298	942	472
403.91	2,629	751	376
410.71	3,131	894	449
411.1	1,540	440	220
412	3,135	895	449
413.9	1,027	293	148
414.00	1,690	483	242
414.01	10,209	2,917	1,459
414.8	1,068	305	153
431	1,092	312	157
433.10	776	221	112
434.91	634	181	92
482.41	907	259	131
486	5,445	1,555	779
491.21	1,295	370	186
493.20	850	243	122
493.90	1,946	556	279
571.2	1,070	305	154
571.5	1,274	364	182
584.5	2,496	713	358
584.9	2,494	712	358
585.6	1,904	544	272
585.9	3,459	988	495
600.00	1,295	370	185
765.18	434	124	63
765.19	389	111	57
765.27	381	109	55
765.28	449	128	65
769	357	102	52
770.6	374	107	54
770.81	231	66	34
774.2	714	204	103
774.6	359	102	53

TABLE 3. Performance evaluation on each diagnosis task.

Cat.	Tasks	Random Forest			MWLSTM			IDDSAM (Ours)		
		F1	ACC	Recall	F1	ACC	Recall	F1	ACC	Recall
001-139	008.45	0.5671	0.5861	0.2164	0.5412	0.5248	0.4925	0.8123	0.6840	1.0000
	038.9	0.5671	0.5861	0.2164	0.5412	0.5248	0.5397	0.7587	0.6840	1.0000
140-239	197.0	0.4641	0.4714	0.1034	0.5656	0.6714	0.0230	0.7919	0.6714	0.9067
	197.7	0.4878	0.4964	0.1828	0.5238	0.5107	0.5806	0.7570	0.6357	0.8663
	198.5	0.4831	0.4750	0.3600	0.4972	0.4893	0.3800	0.6012	0.5214	0.5667
240-279	250.00	0.6205	0.6589	0.8754	0.6164	0.6552	0.8725	0.7612	0.6533	0.0531
	250.40	0.8307	0.8454	0.0377	0.8746	0.8904	0.2013	0.9485	0.9021	1.0000
	250.60	0.8190	0.8461	0.0278	0.8361	0.8110	0.2389	0.9413	0.8892	1.0000
	263.9	0.7956	0.8282	0.0841	0.8038	0.8110	0.2389	0.9252	0.8608	1.0000
280-289	280.0	0.9008	0.9143	0.0074	0.6364	0.5136	0.2963	0.9704	0.9425	1.0000
	280.9	0.8504	0.8666	0.0100	0.8697	0.9020	0.0150	0.9557	0.9152	1.0000
	285.1	0.5592	0.5533	0.3039	0.5371	0.5200	0.3852	0.8204	0.6995	0.9897
	285.21	0.8252	0.8303	0.1756	0.8396	0.8815	0.0382	0.9409	0.8883	1.0000
	285.29	0.8311	0.8457	0.0135	0.8391	0.8551	0.0405	0.9503	0.9054	1.0000
	285.9	0.5530	0.5448	0.4673	0.5348	0.6441	0.0472	0.5852	0.4996	0.5447
390-459	397.0	0.8344	0.8404	0.1473	0.8476	0.8851	0.0310	0.9453	0.8962	0.9991
	401.9	0.4554	0.4677	0.6239	0.5301	0.5244	0.5484	0.8012	0.6137	0.2380
	403.90	0.6414	0.6369	0.3708	0.6549	0.6346	0.6970	0.8065	0.6838	0.9027
	403.91	0.7374	0.7400	0.3697	0.7108	0.7869	0.0532	0.8777	0.7824	0.9956
	410.71	0.6152	0.6208	0.2205	0.5925	0.5653	0.6325	0.8293	0.7159	0.9314
	411.1	0.7519	0.7398	0.4000	0.7501	0.8132	0.0364	0.8559	0.7534	0.8887
	412	0.6176	0.6151	0.2717	0.6423	0.7463	0.0156	0.8122	0.6924	0.8951
	413.9	0.8091	0.8140	0.1554	0.7001	0.6377	0.4527	0.9345	0.8771	0.9937
	414.00	0.6722	0.6576	0.2479	0.7451	0.7981	0.0992	0.7307	0.6081	0.6588
	414.01	0.4554	0.4677	0.2195	0.5301	0.5244	0.4863	0.7606	0.6137	1.0000
	414.8	0.7880	0.7933	0.0980	0.7586	0.7271	0.2941	0.9327	0.8739	0.9955
	431	0.8698	0.8691	0.4904	0.8497	0.8436	0.4777	0.9317	0.8723	0.9954
	433.10	0.8502	0.8619	0.0714	0.7955	0.7614	0.1786	0.9532	0.9106	1.0000
	434.91	0.8818	0.8962	0.0652	0.8911	0.9234	0.0109	0.9623	0.9274	1.0000
460-519	482.41	0.8619	0.8858	0.0382	0.8116	0.7762	0.3282	0.9542	0.5091	1.0000
	486	0.4746	0.5050	0.7356	0.5292	0.5297	0.4917	0.8546	0.5210	0.0292
	491.21	0.7916	0.8130	0.0430	0.8122	0.8357	0.0968	0.9338	0.8758	1.0000
	493.20	0.8631	0.8844	0.0082	0.8593	0.8591	0.1393	0.9575	0.9185	1.0000
	493.90	0.7452	0.7669	0.1900	0.7241	0.7208	0.2832	0.8972	0.8136	1.0000
520-579	571.2	0.4682	0.4673	0.4610	0.5067	0.5446	0.2468	0.6866	0.5625	0.8846
	571.5	0.4682	0.4673	0.4725	0.5067	0.5446	0.7967	0.4111	0.5625	0.3377
580-629	584.5	0.7258	0.6949	0.2507	0.8028	0.8356	0.0958	0.9257	0.8618	1.0000
	584.9	0.4568	0.5017	0.2101	0.5052	0.5344	0.2883	0.6421	0.4945	0.0545
	585.6	0.8632	0.8605	0.3860	0.7604	0.7093	0.5000	0.9441	0.8941	1.0000
	585.9	0.6155	0.5744	0.3596	0.6258	0.5876	0.3495	0.8933	0.8072	1.0000
	600.00	0.8750	0.8784	0.0973	0.8844	0.8958	0.0865	0.9637	0.9299	1.0000
760-779	765.18	0.7842	0.7880	0.0635	0.8171	0.8405	0.0794	0.9361	0.8799	0.9979
	765.19	0.7947	0.7955	0.0179	0.8406	0.8480	0.1786	0.9446	0.8949	1.0000
	765.27	0.7994	0.7955	0.0545	0.8480	0.8968	0.0000	0.9446	0.8949	0.9979
	765.28	0.7636	0.7636	0.0308	0.6846	0.6210	0.4000	0.9340	0.8762	0.9979
	769	0.8311	0.8311	0.1346	0.8493	0.8405	0.3269	0.9487	0.9024	1.0000
	770.6	0.8588	0.8612	0.2778	0.8551	0.8893	0.0556	0.9476	0.9006	1.0000
	770.81	0.8754	0.8780	0.0000	0.9054	0.9362	0.0000	0.9680	0.9381	1.0000
	774.2	0.6398	0.6304	0.1078	0.7203	0.8030	0.0000	0.6174	0.4953	0.5035
	774.6	0.9012	0.9043	0.4423	0.7480	0.6848	0.6731	0.9497	0.9043	1.0000

*Cat. stands for category.

TABLE 4. Performance evaluation on each mortality prediction task.

Cat.	Tasks	SVM			MWLSTM			IDDSAM (Ours)		
		AUROC	AUPRC	ACC	AUROC	AUPRC	ACC	AUROC	AUPRC	ACC
001-139	008.45	0.5214	0.8187	0.8090	0.7560	0.9292	0.8561	0.7870	0.9383	0.8833
	038.9	0.5398	0.6956	0.6904	0.7259	0.8719	0.5741	0.8277	0.9415	0.8948
140-239	197.0	0.5489	0.5242	0.5581	0.7298	0.8701	0.7816	0.7837	0.8877	0.8046
	197.7	0.5511	0.6016	0.5407	0.6147	0.7992	0.6559	0.6147	0.7992	0.6559
	198.5	0.5591	0.6464	0.6102	0.6918	0.8531	0.7300	0.6918	0.8531	0.7300
240-279	250.00	0.6524	0.8923	0.8374	0.7074	0.9503	0.8225	0.6798	0.9508	0.8971
	250.40	0.5912	0.8175	0.7750	0.6101	0.9328	0.8365	0.6390	0.9360	0.8302
	250.60	0.5347	0.8344	0.8178	0.6919	0.9566	0.8833	0.7539	0.9647	0.8944
	263.9	0.5266	0.8125	0.8006	0.5981	0.8968	0.5595	0.6073	0.9223	0.8238
280-289	280.0	0.5375	0.8474	0.8327	0.6272	0.9374	0.8222	0.5946	0.9345	0.8370
	280.9	0.6229	0.8902	0.8589	0.7038	0.9713	0.7700	0.6427	0.9700	0.9200
	285.1	0.5820	0.8792	0.8662	0.7107	0.9630	0.9001	0.6452	0.9564	0.9058
	285.21	0.5107	0.8327	0.8245	0.5887	0.9468	0.8931	0.6448	0.9516	0.8779
	285.29	0.5409	0.8493	0.8335	0.5773	0.9496	0.8700	0.6347	0.9557	0.8789
	285.9	0.5214	0.8532	0.8493	0.7192	0.9671	0.9286	0.7137	0.9665	0.9274
390-459	397.0	0.6906	0.8807	0.8726	0.5706	0.9340	0.8217	0.5571	0.9305	0.7984
	401.9	0.7554	0.9392	0.8651	0.7379	0.9650	0.9111	0.6607	0.9564	0.9089
	403.9	0.6020	0.8630	0.8438	0.6495	0.9467	0.8919	0.6800	0.9503	0.8919
	403.91	0.5906	0.8099	0.7753	0.7133	0.9451	0.8298	0.7269	0.9505	0.8856
	410.71	0.5139	0.8053	0.7979	0.7582	0.9626	0.9154	0.7885	0.9674	0.9332
	411.1	0.6302	0.9567	0.9469	0.6040	0.9834	0.9545	0.5898	0.9824	0.9273
	412	0.6017	0.9064	0.8801	0.6969	0.9629	0.9220	0.7316	0.9666	0.9243
	413.9	0.7601	0.9495	0.9341	0.6631	0.9789	0.9527	0.7490	0.9834	0.9189
	414.00	0.5236	0.8498	0.8379	0.7499	0.9605	0.9174	0.8642	0.9778	0.9504
	414.01	0.6503	0.9113	0.8686	0.7236	0.9719	0.9225	0.6305	0.9637	0.9184
	414.8	0.6431	0.8681	0.8075	0.6483	0.9442	0.8889	0.6329	0.9404	0.8627
	431	0.5654	0.7297	0.7109	0.7042	0.8692	0.6624	0.8446	0.9332	0.8726
	433.10	0.6521	0.8757	0.8122	0.5082	0.9393	0.8393	0.5031	0.9382	0.8304
	434.91	0.5195	0.7240	0.6926	0.8211	0.9412	0.8152	0.8566	0.9539	0.8478
460-519	482.41	0.5315	0.7713	0.7476	0.6447	0.9280	0.8092	0.7021	0.9399	0.8473
	486	0.5962	0.8210	0.7740	0.6551	0.9291	0.8652	0.7033	0.9361	0.8652
	491.21	0.5302	0.7683	0.7477	0.6776	0.9425	0.8656	0.7006	0.9468	0.8817
	493.20	0.5252	0.8412	0.8304	0.8003	0.9757	0.8852	0.8479	0.9817	0.9098
	493.90	0.6235	0.9387	0.9128	0.6416	0.9611	0.4875	0.6265	0.9683	0.9319
520-579	571.2	0.6058	0.7373	0.7371	0.7194	0.9217	0.8571	0.7505	0.9267	0.8506
	571.5	0.5771	0.7157	0.7028	0.7342	0.9267	0.8516	0.7539	0.9311	0.8571
580-629	584.5	0.6738	0.8188	0.7546	0.7562	0.9259	0.8338	0.7805	0.9348	0.8704
	584.9	0.6320	0.8242	0.7656	0.7319	0.5375	0.7109	0.7300	0.6009	0.5523
	585.6	0.5559	0.8503	0.8282	0.7228	0.9517	0.8272	0.7013	0.9512	0.8787
	585.9	0.5981	0.8504	0.8193	0.7336	0.9555	0.9010	0.6917	0.9496	0.8949
	600.00	0.5537	0.8869	0.8698	0.7106	0.9578	0.9081	0.7138	0.9585	0.9135

*Cat. stands for category.

in the Data Description section, mortality prediction did not obtain any output in categories 760-799. From the results, MWLSTM and IDDSAM outperformed RF and SVM, suggesting that deep learning methods are more powerful when handling these kinds of tasks.

The difference in prediction performance between categories is more evident than that within categories. This means

that in the same category, tasks can share more information than those between categories. Additionally, this suggests that multisource multitasks can help to improve the prediction performance. Under the measurement of accuracy, the result can decrease by an average of 3.43 percent between categories. The difficult task for diagnosis in IDDSAM is in category 3, “Endocrine, Nutritional, Metabolic, and

Immunity”, and the easiest task is “Conditions originating in the perinatal period” in category 9. The reason is that the diversities between category 9 and the others are greater; however, the diversities between category 3 and the others are relatively smaller. In the same group, the disease diagnosis and mortality prediction performances are similar, indicating that the relevance within the same system is much higher; this is consistent with the common medical sense. For the severity assessment, ablation studies are conducted to verify the defectiveness of the multisource, and the results suggested that an average F1 score of 3.6 can be achieved by multiple sources cooperating than by each single source. In other words, IDDSAM can share the task vector space among different data sources and prediction tasks in the hidden states. Comparing with ICU score systems, IDDSAM outperforms SOFA and SAPS II by 23.19% and 25.78% on average in terms of accuracy, respectively. Overall, IDDSAM can significantly improve performance for both tasks. Although not all tasks are improved by multisource and multitask learning, for most of the tasks, IDDSAM substantially improved the performance for both tasks. Moreover, the performance of IDDSAM can be continually improved by more training samples, and specific optimizations can be conducted to improve the specific tasks.

V. CONCLUSION AND FUTURE WORK

In this work, we proposed a multisource multitask model named IDDSAM for the disease diagnosis and severity assessment of ICU patients. We treated the disease diagnosis and mortality prediction as an integrated multisource multitask classification problem. We adapt the attention mechanism and window alignment operation to improve the prediction performance and use focal loss to solve the imbalance problem. We use spatial information of the clinical measurements and the clinical treatments for real-time modeling such that this model can be used in a real-time clinical scenario. The significance of our proposed model can be summarized as follows:

- 1) **We considered the diversity of complications.** On the one hand, this meets the medical situation in which no disease is isolated. On the other hand, diversity among different conditions exists. Therefore, the criteria for diagnosis and treatment and even severity assessment should be similar but different. The proposed multisource multitask model IDDSAM is explicitly designed for these situations.
- 2) **We considered the sequential diagnosis relationship.** By introducing the window alignment operation and the attention layer, we simulated the ICU handling process for admitted patients and captured the interaction information within and among the ICU stay process.
- 3) **Solved the imbalance problem.** By considering the samples, a considerable difference exists among different tasks. For example, the “Diabetes mellitus no mention of complication” has 10585 samples. However, the “Unspecified fetal and neonatal jaundice” has only

514 samples. Therefore, if we train the model without any precautionary measures, the prediction result would prefer the majority ones. To solve this problem, we introduce a focal loss function into our loss process.

- 4) **Give out explanations.** We provide evidence-based explanations for the clinicians with related diseases and the diagnosis of trajectory.

A comprehensive experiment was conducted using 47855 ICU admissions. These admissions include 50 different types of diseases and nine categories. The results of our experiment are promising and improved the performance by approximately 15% in terms of all the evaluation metrics for all baselines. The results provide strong evidence about the robustness and accuracy of IDDSAM. However, due to the differences among different diseases, the performance of some tasks is still difficult to improve. Moreover, how to use these diagnoses and severity results in further clinical actions to treat ICU patients remains a challenge. Therefore, there are many future works in this field.

REFERENCES

- [1] Z. Shi, W. Zuo, W. Chen, L. Yue, Y. Hao, and S. Liang, “DMMAM: Deep multi-source multi-task attention model for intensive care unit diagnosis,” in *Proc. Int. Conf. Database Syst. Adv. Appl.* Springer, 2019, pp. 53–69. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-18579-4_4
- [2] A. T. Azar and S. M. El-Metwally, “Decision tree classifiers for automated medical diagnosis,” *Neural Comput. Appl.*, vol. 23, nos. 7–8, pp. 2387–2403, Dec. 2013.
- [3] T. Chandler, R. Pralle, J. Dórea, S. Pooock, G. Oetzel, R. Fourdraine, and H. White, “Predicting hyperketonemia by logistic and linear regression using test-day milk and performance variables in early-lactation Holstein and Jersey cows,” *J. Dairy Sci.*, vol. 101, no. 3, pp. 2476–2491, Mar. 2018.
- [4] R. Patankar, “A survey on computer-aided breast cancer detection using mammograms,” *Nat. J. Comput. Appl. Sci.*, vol. 2, no. 1, pp. 1–6, 2019.
- [5] L. P. Zimmerman, P. A. Reyfman, A. D. Smith, Z. Zeng, A. Kho, L. N. Sanchez-Pinto, and Y. Luo, “Early prediction of acute kidney injury following ICU admission using a multivariate panel of physiological measurements,” *BMC Med. Informat. Decis. Making*, vol. 19, no. 1, p. 16, 2019.
- [6] C. Lin, Y. Zhangy, J. Ivy, M. Capan, R. Arnold, J. M. Huddleston, and M. Chi, “Early diagnosis and prediction of sepsis shock by combining static and dynamic information using convolutional-lstm,” in *Proc. IEEE Int. Conf. Healthcare Informat. (ICHI)*, Jun. 2018, pp. 219–228.
- [7] L. Bonomi and X. Jiang, “A mortality study for ICU patients using bursty medical events,” in *Proc. IEEE 33rd Int. Conf. Data Eng. (ICDE)*, Apr. 2017, pp. 1533–1540.
- [8] H. Harutyunyan, H. Khachatrian, D. C. Kale, G. Ver Steeg, and A. Galstyan, “Multitask learning and benchmarking with clinical time series data,” *Sci. Data*, vol. 6, no. 1, p. 96, 2019.
- [9] M. Alhussein, “Monitoring Parkinson’s disease in smart cities,” *IEEE Access*, vol. 5, pp. 19835–19841, 2017.
- [10] M. Blaxter, “Diagnosis as category and process: The case of alcoholism,” *Social Sci. Med. A, Med. Psychol. Med. Sociol.*, vol. 12, pp. 9–17, Jan. 1978.
- [11] P. Brown, “Naming and framing: The social construction of diagnosis and illness,” *J. Health Social Behav.*, vol. 35, pp. 34–52, 1995.
- [12] R. H. Scheuermann, W. Ceusters, and B. Smith, “Toward an ontological treatment of disease and diagnosis,” *Summit Transl. Bioinformat.*, vol. 2009, p. 116, Mar. 2009.
- [13] J. Polivka, M. Kralickova, J. Polivka, C. Kaiser, W. Kuhn, and O. Golubnitschaja, “Mystery of the brain metastatic disease in breast cancer patients: Improved patient stratification, disease prediction and targeted prevention on the horizon?” *EPMA J.*, vol. 8, no. 2, pp. 119–127, Jun. 2017.
- [14] N. C. Long, P. Meesad, and H. Unger, “A highly accurate firefly based algorithm for heart disease prediction,” *Expert Syst. Appl.*, vol. 42, no. 21, pp. 8221–8231, Nov. 2015.

- [15] M. Nilashi, H. Ahmadi, L. Shahmoradi, O. Ibrahim, and E. Akbari, "A predictive method for hepatitis disease diagnosis using ensembles of neuro-fuzzy technique," *J. Infection Public Health*, vol. 12, no. 1, pp. 13–20, Jan. 2019.
- [16] E. Choi, M. T. Bahadori, A. Schuetz, W. F. Stewart, and J. Sun, "Doctor AI: Predicting clinical events via recurrent neural networks," in *Proc. Mach. Learn. Healthcare Conf.*, 2016, pp. 301–318.
- [17] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.
- [18] Z. Shi, W. Chen, S. Liang, W. Zuo, L. Yue, and S. Wang, "Deep interpretable mortality model for intensive care unit risk prediction," in *Proc. Int. Conf. Adv. Data Mining Appl.* Springer, 2019, pp. 617–631.
- [19] N. Petrick, B. Sahiner, S. G. Armato, III, A. Bert, L. Correale, S. Delsanto, M. T. Freedman, D. Fryd, D. Gur, and L. Hadjiiski, "Evaluation of computer-aided detection and diagnosis systems^(a)," *Med. Phys.*, vol. 40, no. 8, p. 087001, 2013.
- [20] J.-W. Perng, I.-H. Kao, Y.-W. Chen, Y.-H. Lai, C.-M. Su, S.-C. Hung, M. S. Lee, and C.-T. Kung, "Analysis of the 72-h mortality of emergency room septic patients based on a deep belief network," *IEEE Access*, vol. 6, pp. 76820–76830, 2018.
- [21] S. H. Sicherer and H. A. Sampson, "Food allergy: A review and update on epidemiology, pathogenesis, diagnosis, prevention, and management," *J. Allergy Clin. Immunol.*, vol. 141, no. 1, pp. 41–58, Jan. 2018.
- [22] X. Wang, Z. Wang, J. Weng, C. Wen, H. Chen, and X. Wang, "A new effective machine learning framework for sepsis diagnosis," *IEEE Access*, vol. 6, pp. 48300–48310, 2018.
- [23] L. Detemmerman, S. Olivier, V. Bours, and F. Boemer, "Innovative PCR without DNA extraction for African sickle cell disease diagnosis," *Hematology*, vol. 23, no. 3, pp. 181–186, Mar. 2018.
- [24] P. Croft, D. G. Altman, J. J. Deeks, K. M. Dunn, A. D. Hay, H. Hemingway, L. LeResche, G. Peat, P. Perel, S. E. Petersen, R. D. Riley, I. Roberts, M. Sharpe, R. J. Stevens, D. A. Van Der Windt, M. V. Korff, and A. Timmis, "The science of clinical practice: Disease diagnosis or patient prognosis? Evidence about 'what is likely to happen' should shape clinical practice," *BMC Med.*, vol. 13, no. 1, p. 20, 2015.
- [25] D. Zhang and D. Shen, "Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease," *NeuroImage*, vol. 59, no. 2, pp. 895–907, Jan. 2012.
- [26] R. A. Toledo, N. Burnichon, A. Cascon, D. E. Benn, J.-P. Bayley, J. Welander, C. M. Tops, H. Firth, T. Dwight, T. Ercolino, M. Mannelli, G. Opocher, R. Clifton-Bligh, O. Gimm, E. R. Maher, M. Robledo, A.-P. Gimenez-Roqueplo, and P. L. M. Dahia, "Consensus Statement on next-generation-sequencing-based diagnostic testing of hereditary pheochromocytomas and paragangliomas," *Natural Rev. Endocrinol.*, vol. 13, no. 4, pp. 233–247, Apr. 2017.
- [27] A. Subasi, "Classification of EMG signals using PSO optimized SVM for diagnosis of neuromuscular disorders," *Comput. Biol. Med.*, vol. 43, no. 5, pp. 576–586, Jun. 2013.
- [28] C. Nguyen, Y. Wang, and H. N. Nguyen, "Random forest classifier combined with feature selection for breast cancer diagnosis and prognostic," *JBISe*, vol. 06, no. 5, pp. 551–560, 2013.
- [29] Y. Hao, W. Zuo, Z. Shi, L. Yue, S. Xue, and F. He, "Prognosis of thyroid disease using MS-Apriori improved decision tree," in *Proc. Int. Conf. Knowl. Sci., Eng. Manage.* Springer, 2018, pp. 452–460. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-99365-2_40
- [30] Z. Shi, W. Zuo, W. Chen, L. Yue, J. Han, and L. Feng, "User relation prediction based on matrix factorization and hybrid particle swarm optimization," in *Proc. 26th Int. Conf. World Wide Web Companion*, 2017, pp. 1335–1341.
- [31] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep Learning*, vol. 1. Cambridge, MA, USA: MIT Press, 2016.
- [32] M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, "Disease prediction by machine learning over big data from healthcare communities," *IEEE Access*, vol. 5, pp. 8869–8879, 2017.
- [33] Z. Che, S. Purushotham, R. Khemani, and Y. Liu, "Interpretable deep models for ICU outcome prediction," in *Proc. AMIA Annu. Symp.*, 2016, p. 371.
- [34] C. L. Chiang, *The Life Table and Its Applications*. New York, NY, USA: Academic, 1984. [Online]. Available: <https://www.sciencedirect.com/book/9780125139304/life-table-techniques-and-their-applications>
- [35] M. J. Breslow and O. Badawi, "Severity scoring in the critically ill: Part 1—Interpretation and accuracy of outcome prediction scoring systems," *Chest*, vol. 141, no. 1, pp. 245–252, 2012.
- [36] H. Song, D. Rajan, J. J. Thiagarajan, and A. Spanias, "Attend and diagnose: Clinical time series analysis using attention models," in *Proc. 32nd AAAI Conf. Artif. Intell.*, 2018.
- [37] N. Tangri, L. A. Stevens, J. Griffith, H. Tighiouart, O. Djurdjev, D. Naimark, A. Levin, and A. S. Levey, "A predictive model for progression of chronic kidney disease to kidney failure," *J. Amer. Med. Assoc.*, vol. 305, no. 15, pp. 1553–1559, 2011.
- [38] W. De Winter, J. DeJongh, T. Post, B. Ploeger, R. Urquhart, I. Moules, D. Eckland, and M. Danhof, "A mechanism-based disease progression model for comparison of long-term effects of pioglitazone, metformin and gliclazide on disease processes underlying type 2 diabetes mellitus," *J. Pharmacokinetics Pharmacodyn.*, vol. 33, no. 3, pp. 313–343, 2006.
- [39] M. J. Johnson and A. S. Willsky, "Bayesian nonparametric hidden semi-Markov models," *J. Mach. Learn. Res.*, vol. 14, no. 1, pp. 673–701, 2013.
- [40] S. Ruder, "An overview of multi-task learning in deep neural networks," 2017, *arXiv:1706.05098*. [Online]. Available: <https://arxiv.org/abs/1706.05098>
- [41] D. Xu, Y. Shi, I. W. Tsang, Y.-S. Ong, C. Gong, and X. Shen, "A survey on multi-output learning," 2019, *arXiv:1901.00248*. [Online]. Available: <https://arxiv.org/abs/1901.00248>
- [42] A. Trask, D. Gilmore, and M. Russell, "Modeling order in neural word embeddings at scale," 2015, *arXiv:1506.02338*. [Online]. Available: <https://arxiv.org/abs/1506.02338>
- [43] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, Feb. 2020.
- [44] A. E. W. Johnson, T. J. Pollard, L. Shen, L.-W. H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. A. Celi, and R. G. Mark, "MIMIC-III, a freely accessible critical care database," *Sci. Data*, vol. 3, p. 160035, May 2016.
- [45] V. Bewick, L. Cheek, and J. Ball, "Statistics review 13: Receiver operating characteristic curves," *Crit. Care*, vol. 8, no. 6, p. 508, 2004.
- [46] B. Ozenne, F. Subtil, and D. Maucourt-Boulch, "The Precision-recall curve overcame the optimism of the receiver operating characteristic curve in rare diseases," *J. Clin. Epidemiol.*, vol. 68, no. 8, pp. 855–859, Aug. 2015.



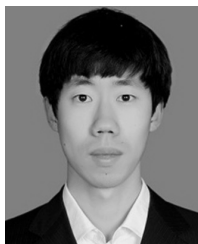
ZHENKUN SHI received the B.Sc. degree from the College of Computer Science and Technology, Agricultural University of Hebei, China. He is currently pursuing the Ph.D. degree with the Key Laboratory of Symbol Computation and Knowledge Engineering of the Ministry of Education, Jilin University, Changchun, China. He is also the joint Ph.D. student with the University of Queensland. His research interests include clinical data mining, social computing, and deep learning.



WANLI ZUO received the B.Sc., M.S., and Ph.D. degrees from the College of Computer Science and Technology, Jilin University, Changchun, China. He is currently a Professor with Jilin University. He has published more than 160 journal articles and conference papers. His research interests include data mining, information retrieval, natural language processing, and machine learning.



SHINING LIANG received the B.Sc. degree from the College of Computer Science and Technology, Jilin University, Changchun, China, where he is currently pursuing the Ph.D. degree with the Key Laboratory of Symbol Computation and Knowledge Engineering, Ministry of Education. His research interests include natural language processing, causality mining, deep learning, and clinical data mining.



XIANGLIN ZUO received the B.Sc. degree from Jilin University, China, in 2015, where he is currently pursuing the Ph.D. degree in computer science with the College of Computer Science and Technology. His main research interests include machine learning, deep learning, and social network analysis.



LIN YUE received the B.S. and M.S. degrees from Northeast Normal University, the Ph.D. degree from the Key Laboratory of Symbolic Computation and Knowledge Engineering, Ministry of Education, Jilin University, China, all in computer science, and the joint Ph.D. degree from The University of Queensland, Australia. She is currently a Postdoctoral Research Fellow with Northeast Normal University. Her current main research interests include data mining, natural language processing, and machine learning.



XUE LI (Member, IEEE) received the Ph.D. degree in information systems from the Queensland University of Technology, in 1997. He is currently a Full Professor with the Department of Computer and Software, Dalian Neusoft University of Information. He is also a Full Professor with the School of Information Technology and Electrical Engineering, The University of Queensland, Australia. He is also an Adjunct Professor with Chongqing University, China. His research areas are big data analytics, pattern recognition, and intelligent information systems. He is a member of the ACM.

• • •