Article

ICU Readmission Prediction with Correlation Enhanced Multi-task Learning

Ke Niu [1,*], Su Pei [1], Xueping Peng [2], You Lu [1] and Jingni Zeng [1]

[1] Beijing Information Science and Technology University, Beijing, China; {niuke, peisu1128, luyou04, zjn981223}@bistu.edu.cn
[2] Australian Artificial Intelligence Institute, Faculty of Engineering and Information Technology, University of Technology Sydney, Australia; xueping.peng@uts.edu.au
* Correspondence: niuke@bistu.edu.cn

Citation: Ke, N.; Su, P.; Xueping, P.; You L.;Jingni Z. ICU Readmission Prediction with Correlation Enhanced Multi-task Learning. *Journal Not Specified* **2021**, *1*, 0. https://doi.org/

Received:
Accepted:
Published:

Abstract: Prediction for Intensive Care Unit (ICU) readmission is conducive to assisting doctors to make clinical decisions and reducing the risk of relapse after discharge. Recently, ICU readmission prediction task has been address by multiple prediction sub-tasks which are related to each other and reflect the patient's ICU readmission risk in different time periods. However, these existing approaches train each sub-task independently, which prevents the models from using complementary information between these sub-tasks to improve prediction performance. In this paper, we propose correlation enhanced Multi-task learning with Pearson and RNN-based Neural Ordinary Differential Equations Model (MP-ROM). In order to enhance the learning of general features and avoid the local optima in single-task training, we construct the Shared-Bottom structure of multi-task learning, which enables multiple tasks to share model structure and parameters. To this end, we add the task correlation score, which is calculated by Pearson correlation calculation, in the Shared-Bottom structure to make advantage of the association between sub-tasks. We also conduct extensive experiments on MIMIC-III dataset, and the experimental results show that our proposed MP-ROM achieves the highest average precision, which demonstrates that the proposed Shared-Bottom structure to enhance task association can further improve the predictive performance of ICU readmission risk prediction.

Keywords: Machine Learning, Multi-task Learning, Electronic Health Record, ICU Readmission Prediction

1. Introduction

With the outbreak of COVID-19 in 2019, the patients with critical conditions have increased rapidly to admit to ICU which aggravates the medical resource shortage. The health of these patients would be getting worse and worse if they admitted to ICU multiple times within a short period. It is necessary to predict the risk of ICU readmission in advance [1] and assist doctors to provide early intervention for patients, which can relieve the shortage of medical resources and improve patients' wellbeing. Recently, with the continuous development of medical informatization, a large amount of patient data, such as Electronic Health Record (EHR) [2], has been generated [3,4]. In EHR, a large number of patient data from ICU contains valuable information for ICU readmission prediction [5], which has been a hot research topic [6,7].

With the widespread application of machine learning (ML) and the development of the clinical prediction research, ML-based models have become a new clinical research trend, including k-nearest neighbor (K-NN)[8], multi-layer perceptron[9], support vector machine (SVM)[10], random forest (RF)[7,11,12], decision tree (DT)[13] and artificial neural network (ANN)[14]. As a branch of ML, the deep learning (DL) based method has attracted great attention due to its end-to-end characteristics. DL-based methods such as long short-term memory (LSTM)[15], convolutional neural networks (CNN)[16,17], and trajectory-based deep learning (TADEL)[18] have been successfully applied to the prediction of readmission and have achieved good prediction results.

In deep learning based readmission prediction tasks, most clinical data are composed of sequential data. The memory capacity of recurrent neural networks (RNN) gives it an advantage in processing sequence data[20], which makes RNN widely used in readmission prediction[19]. However, there are lots of time sequential data sampled at irregular time intervals, such as diagnosis and procedure codes, which can not be processed by RNN. In order to solve this problem, Barbieri et al. propose to use neural ordinary differential equations (ODE) to process time sequential data on the basis of RNN to obtain continuous patient treatment process representation[21], and achieve better ICU readmission prediction results. It only predicts the risk of ICU readmission within a fixed period of time. For example, in the 30-day ICU readmission study, patients who re-enter the ICU 5 days or 20 days after leaving the ICU ward will be regarded as a positive sample. These two situations (ICU readmit within 5 or 20 days) show that the urgency of readmission in different periods has predictive value, while merely predicting the urgency of readmission risk within 30 days fails to reflect the ICU readmission risk in a shorter period of time (such as 5 days) satisfactorily. In order to analyze the risk of readmission in different time periods, Denise et al. predicted the risk of readmission in two time periods (15 and 30 days)[22,23]. However, there is complementary information between the ICU readmission prediction tasks in the two time periods in this study, including task-related parts that can help to learn common feature representations and task-related parts that help to escape from local optima [24,25]. Independently training each task prevents the model from using the complementary information to improve prediction performance.

To make better use of the complementary information between ICU readmission sub-tasks to improve prediction performance, we propose a multi-task learning based model shared bottom components to simultaneously learn multiple ICU readmission sub-tasks in different time periods. Specifically, the model shares the ODE-based RNN module to learn multiple prediction tasks at the same time, and employs the dynamic weighting method to optimize the training loss of the multi-task learning model[26,27]. To explore the relationship between multiple tasks to strengthen the task relationship in order to improve the prediction effect of the model, we introduce a task correlation approach, which is inspired by Pearson correlation coefficient used to reflect the similarity between two variables. It is suitable for calculation of binary similarity, such as calculating the similarity between labels of binary classification tasks, and can be used to quantify the degree of similarity between tasks correlation[28]. However, compared with the existing method that simply uses task correlation as an indicator of task relationship, we propose to integrate task correlation as additional hidden information into the multi-task learning model to improve model performance. Thus, out proposed model exploit the task relevance score represented by the Pearson correlation coefficient and the output information of the model sharing layer fused in the shallow layer of the model as a supplement of additional information.

The overall structure of the model is the multi-task learning shared bottom structure, which internally shares the RNN module based on neural ODE. The model allows the interaction of relevant parts between sub-tasks, which strengthens the learning of general feature representation and the interaction of irrelevant parts between tasks to avoid the local optima. The loss function of the model is dynamically weighted and optimized, which allows the neural network to adjust the weight of the loss of different tasks according to different task learning stages, learning difficulties and learning effects during training, optimize the dynamic weighting module of the model training rate, and improve the prediction performance. The task relevance score represented by the Pearson correlation coefficient added to the model is used as additional information to strengthen the relevance between tasks and improve the predictive ability of the underlying sharing model. We conducted experiments on the public Medical Information Mart Critical Care

III (MIMIC-III)[29] dataset and compared with other multi-task learning models. The results show that the proposed model outperforms the baseline models on the prediction of ICU readmission. Our main contributions are summarized as follows:

- We propose a method to use quantitative task relevance scores as additional information to strengthen the inter-task relevance in the multi-task learning model. The method calculates the Pearson correlation coefficient between the 30-day prediction task and other prediction subtasks as a quantitative task relevance score, and then uses the relevance score as hidden information to merge with the output of the model sharing layer at a shallow level to improve the correlation between prediction tasks.
- We propose MP-ROM, which is a multi-task learning model that integrates task relevance. The RNN module based on neural ODE is shared at the bottom of the model, and the training loss of multiple tasks is dynamically weighted and optimized.
- Extensive experiments on real data prove that our proposed model can more accurately predict the patient's future ICU readmission risk compared with other multi-task learning models.

The paper is structured as follows: Section 2 includes related work. In Section 3, the details of our model are explained. Next, in Section 4, we demonstrate the experimental results conducted on real data set. Finally, we summarize our work in Section 5.

2. Related work

In this section, we review the related work about clinical prediction research based on EHR and deep learning. And construction and optimization for multi-task learning and feature correlation calculation method researches are also reviewed.

2.1. Clinical prediction research based on EHR data and deep learning

Using EHR to clinical prediction is significant and difficult target with deep learning, which is limited by imbalance or incomplete data. Cho, K. et al. used recurrent neural network to process time sequence data in EHR[30], while the use of recurrent layers is associated with several drawbacks: interpretation of results is hampered by outputs being a nonlinear combination of current input and current memory state; lack of set-invariance[31]; and long training times due to difficulties in parallelizing these sequential algorithms. Choi, E. et al. proposed neural networks relying entirely on attention mechanisms as an alternative to overcome the limitations of recurrent neural networks[2], with improved accuracy when used for risk predictions. However, both of these model were not suitable to process time sequences sampled at irregular time intervals, such as the diagnosis and procedure codes contained in EHR.

When addressing this issue, a wide variety of approaches have been proposed. Choi et al. proposed simply adding or appending time-related information to the numerical vectors ("embeddings") used to represent timestamped codes[2]. Mozer et al. proposed modify the internal workings of recurrent cells using exponential time-decay functions. Chen et al. proposed neural ODE to describe how the embedding of a medical code changes over time[32]: codes related to a chronic condition, such as diabetes, will often maintain their relevance over years, whereas others may quickly become unimportant for prediction purposes. In the current study of ICU readmission, some studies used ODE to construct the patient dynamic process curve and obtain a robust representation of the patient treatment process[33].

2.2. Construction and optimization for multi-task learning

The Shared-Bottom mechanism (hard parameter sharing) is easy to implement and can effectively reduce model overfitting[34], but only provides security for closely related tasks. In order to construct a multi-task learning model for highly correlated prediction tasks, existing studies often use the Shared-Bottom model of multi-task learning: the hidden layer close to the input layer is shared as a whole, and the output result f(x) of the hidden layer is shared and input into their respective sub-networks, and finally each task gets one output[35]. Meyerson et al. proposed a soft sharing mechanism that allows each task to independently learn a network, but the network of each task can access the information in the corresponding network of other tasks[36]. Ruder et al. designed the model to do simple tasks at the lower level of the network and difficult tasks at the higher level to build a layered shared structure that relied on expert experience[37]. The MMoE sharing method[28] proposed by Jiaqi Ma et al. means that the bottom layer contains multiple Experts, and then based on the gating mechanism, different tasks will filter the output of different experts. On the basis of MMoE, Tang et al. proposed a Progressive Layered Extraction multi-task learning model[38], in which each task has an independent Expert while maintaining a shared Expert. At the same time, they also designed an asymmetric shared multi-tasking framework, in which the bottom modules of different tasks have their own corresponding output, but the output of some tasks will be used by other tasks, while some tasks will use their own unique output. Which part of the task uses the output of other tasks needs to be manually specified

Existing studies have found that the optimization of multi-task learning loss is also very important to improve model performance[27]. Different tasks have different degrees of difficulty in learning; and different tasks may be in different learning stages. Chen et al. proposed two types of Loss optimization[26], namely Label Loss and Gradient Loss. The optimization of Label Loss is dynamically weighted, which can be adjusted according to the learning stage of different tasks, the difficulty of learning, and even the learning effect. The w weight that changes in each batch is multiplied by the loss calculated by the real data label and network prediction label of each task. This is achieved by weighted summation of loss of different tasks.

2.3. Feature correlation calculation method

In machine learning, correlation calculation methods can be used to calculate the similarity degree between features and categories, so as to judge whether the extracted features and categories are positively correlated, negatively correlated or not correlated[39]. In current studies, Mu et al. used Pearson correlation coefficient as a new measure of feature quality to determine the optimal splitting attribute and splitting point in decision tree growth[40]. Y Lu et al. started from a thin network and gradually enlarged it, selectively shared tasks, and mined the correlation between those tasks (Pearson correlation coefficient) for sharing segmentation[41]. Spearman's correlation coefficient and Kendall's rank correlation coefficient are both rank correlation coefficients and statistical analysis indexes reflecting the degree of rank correlation[42]. In contrast, Pearson is a statistic used to reflect the similarity degree between bivariate variables[43]. In this study, the task labels we need to calculate are applicable to bivariate similarity calculation, so Pearson correlation coefficient calculation method is adopted to obtain the task correlation score.

3. Method

In this study, we propose an MP-ROM model using Pearson correlation to calculate multi-task correlation scores and RNN based Odes to predict the risk of ICU readmission at 5, 10, 15, and 30 days. The overall structure of the model is shown in Figure 1. There are three layers in the MP-ROM model, which are shared layer, fusion layer and output layer. Static data, dynamic data processed by ODE-based RNN in shared layer and task correlation obtained by label data processed by Pearson correlation calculation are integrated in fusion layer. In output layer, dynamic weighting of loss function is used for training optimization. Finally, four ICU readmission sub-tasks at 5, 10, 15 and 30 days are predicted. In MP-ROM, the ROM part uses this part to form the bottom shared part of multi-task learning, and obtains continuous representation of patient treatment process through neural ordinary differential equations to improve the accuracy

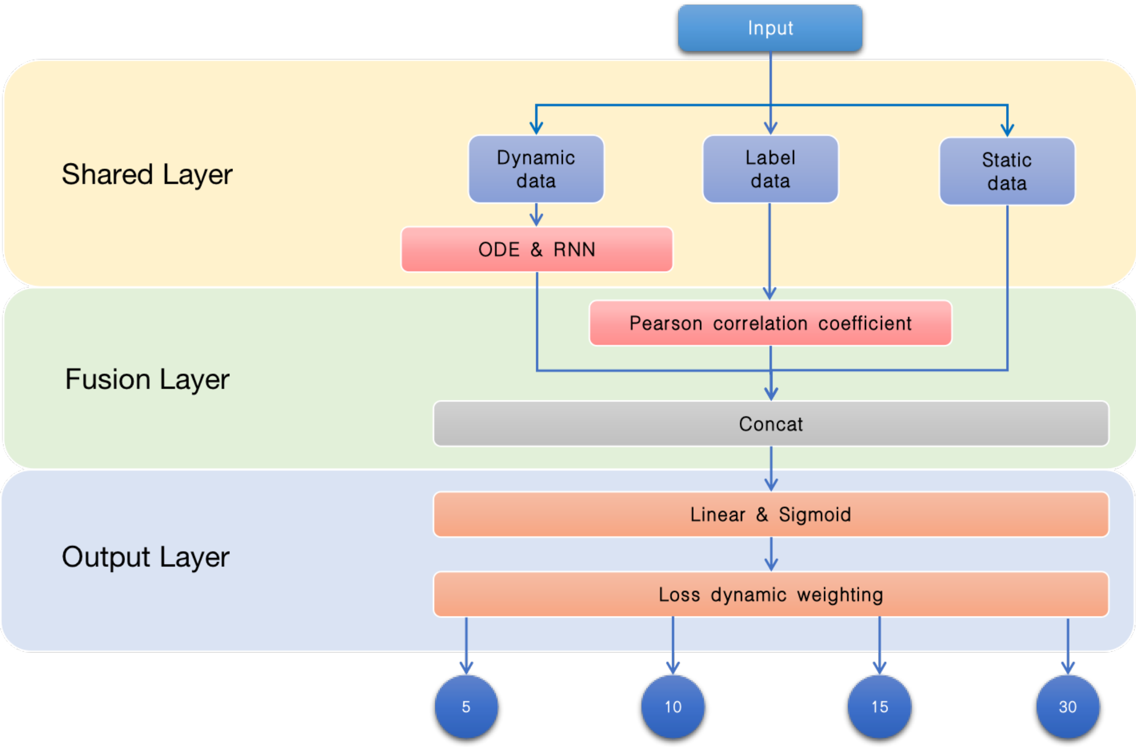


Figure 1. The structure diagram of MP-ROM model.

Algorithm 1 The MP-ROM model

Randomly initialize diagnoses and procedures, medications and vital signs, and related time information embedding matrix DP, CP;
repeat
  for visit v_t in X **do**
    Update DP, CP with ODE function (see Section 3.2);
    Calculate the task correlation score P ma P (see Section 3.3);
    Fusion state information, DP, CP and P;
    Make prediction y^t using Softmax function;
  end for
  Calculate the prediction loss L (see Section 3.5);
  Update parameters according to the gradient of L;
until convergence

---

of model prediction. In the MP part, Pearson correlation coefficient is used to calculate the similarity between the 5, 10 and 15 tasks and the 30-day prediction task labels to obtain the task similarity score, and then the similarity score is respectively fused with the ODE&RNN part. In this section, we will describe each part of the model in detail. Algorithm 1 describes the overall training process of the proposed MP-ROM.

Deep neural network structure is used to predict the risk of patient readmission to ICU. All task models share a similar structure at the bottom: the ODE module calculates dynamic data scores related to diagnostic codes, medication codes and vital signs codes; the task correlation score calculated by Pearson correlation coefficient needs to be associated with the bottom static variables and ODE&RNN parts respectively according to the prediction task. These data need to be correlated to the "Logistic regression layer" (i.e., the fully connected layer with sigmoid activation function). More information about each network component is reported in the following sections.

3.1. Multi-task learning framework

Compared with soft sharing and AS sharing mechanism, this study used the Shared-Bottom model which is more suitable for that the correlation among four prediction tasks is very strong. The model is fully shared in the ODE&RNN part, and task relevance is integrated in the fusion part. After layer of logistic regression, the training loss function is dynamically weighted and the prediction result of each prediction task is obtained.

3.2. ODE-based RNN layer

There are a large volume of irregular sampling to obtain patient-related time information, including diagnostic and procedural codes, as well as medication and vital sign codes, which are mapped to the corresponding "embeddings". Diagnoses & procedures and medication & vital signs are processed separately, as they are measured on different time scales[19]. It is difficult for such information to be directly applied to the neural network, so we use the neural ordinary differential equation model proposed by Chen[32], which is very attractive to the processing of time sequences. In ODEs, the continuously defined dynamic information can be naturally incorporated into any data arriving at any given time[21]. So, we calculate the time-aware code embedded in the ODE dynamic simulation, and neural nodes are used to simulate the dynamic process of embedding. To make better use of the data's timestamp information and be able to make predictions at any point in time, the neural ODE models the time sequences as a continuous trajectory of change. Each trajectory is determined by the local initial state S_{t_0} and the potential dynamic global set shared by all time sequences. Given observation t_0, t_1, ..., t_n and initial state S_{t_0}, an ODE solver produces S_{t_1}, ..., S_{t_n}, which describe the latent state at each observation. We define this generative model formally through a sampling procedure:

$$
\begin{gathered}
s_{t_0} \sim p(s_{t_0}) \\
s_{t_1}, s_{t_2}, ..., s_{t_N} = \texttt{ODESlove}(s_{t_0}, f, \theta f, t_0, ..., t_N) \\
each x_{t_i} \sim p(x|s_{t_i}, \theta x)
\end{gathered} \tag{1}
$$

Function f is a time-invariant function that takes the value s at the current time step and outputs the gradient: $\partial s(t))/\partial t = f(s(t), \theta f)$. This function is parametrized using a neural network. Because f is time-invariant, given any latent state s(t), the entire latent trajectory is uniquely defined. Extrapolating this latent trajectory lets us make predictions arbitrarily far forwards or backwards in time.

In MP-ROM, we, using adjunction sensitivity to calculate gradients, treat the ODE approximately as a black box layer to process an irregularly sampled time sequences in the data. The adjunction sensitivity method, applicable to all ODE solvers, calculates the gradient by solving backward the second step of the augmented ODE which scales linearly with problem size, having lower storage costs and explicitly limiting numerical errors. After the ODE layer processes the time information, the processed information is passed to the RNN layer for further processing. Bidirectional RNN is used to overcome the drawback that the prediction accuracy decreases with the increase of sequence length, and to deal with the gradient disappearance.

3.3. Task relevance score calculation

In this study, the task labels we need to calculate are suitable for binary similarity calculation, so Pearson correlation coefficient calculation method is adopted to obtain the task correlation score.

Pearson correlation coefficient was calculated by the labels in each batch in the training task, and the calculation formula is as follows:

$$\rho_{X,Y} = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{(\sum X^2 - \frac{(\sum X)^2}{N})(\sum Y^2 - \frac{(\sum Y)^2}{N})}} \tag{2}$$

Among them, X and Y are the sample points of variables. In this study, X and Y are the prediction task labels divided at different times, and all calculated ratios are based on 30 days.

3.4. Logistic regression

The computed diagnoses & procedures and medications & vital signs scores are concatenated to the vector of static variables and passed to a fully connected layer with a sigmoid activation function. The output of the network corresponds to the risk of readmission to the ICU within 5-days, 10-days, 15-days and 30-days from discharge.

3.5. Objective fuction

The objective function guides the network parameter learning and representation learning through the back propagation between the predicted results of samples and the errors generated by real markers.

3.5.1. BCEWithLogitsLoss

In our experiment, we use the following BCEWithLogitsLoss function to calculate the predicted loss of the model:

$$l(x,y) = L = \{l_1, ..., l_N\}^T \tag{3}$$

where x represents the input and y represents the target label.

$$l_n = -\omega_n[y_n \cdot \log(\sigma(x_n)) + (1 - y_n) * \log(1 - \sigma(x_n))] \tag{4}$$

where ω is the weight of the item. Compared with Sigmoid with BCELoss, this loss function uses log-sum-exp trik, getting better numerical calculation stability.

3.5.2. Loss function weighting

Based on the design of single-task loss function, we dynamically weighted the loss function during the reverse propagation of fusion loss, so that the loss of each task could be adjusted according to the learning stage of different tasks, the difficulty of learning, and even the learning effect. The calculated loss of the real data label and network prediction label of each task is determined by the nature of the learning task, which is realized by weighted sum of the loss of different tasks:

$$L = \sum_i \omega_i(t) * L_i \tag{5}$$

4. Experiment

In this section, we performed several comparative experiments on mimic-III data sets of large public health electronic medical records to evaluate the performance of the proposed MP-ROM in predicting ICU readmission. This part includes three parts: data description, experimental design, results and discussion.

4.1. Data Description

4.1.1. Dataset

The algorithm was evaluated on publicly available MIMIC-III data set[1] (no ethical approval is required). The data set included unidentified health data for 61,532 intensive

[1] https://mimic.physionet.org/

care units and 46,476 intensive care patients at Beth Israel Deaconess Medical Center in Boston, Massachusetts, between 2001 and 2012. In our experiment, the supervised learning tasks consist of predicting whether the patient will be readmitted to the ICU within 5, 10, 15 and 30 days from discharge for a given ICU stay. Patients were excluded if they died during the ICU stay, or were not adults at the time of discharge, or died within 5, 10, 15 and 30 days from discharge without being readmitted to the ICU. The final data set comprised of 13,383 ICU stays, labelled as either positive (N= 10,775, 9,738, 9,091, 7,888) or negative (N= 10,775, 9,738, 9,091, 7,888) depending on whether a patient did or did not experience readmission within 5, 10, 15 and 30 days from discharge. To develop and evaluate algorithms, patients were randomly subdivided into training and validation (90%) and test sets (10%). This segmentation is based on patient identification not on ICU admission identification, to prevent information leakage between data sets.

4.1.2. Data Pre-processing

The data of a patient can be represented as a set of significant static variables and timestamped codes. In our research, static variables included the patient's gender, age, ethnicity, insurance type, marital status, the previous location of the patient prior to arriving at the ICU, and whether the patient was admitted for elective surgery. Both length of ICU stay and length of hospital stay prior to ICU admission were recorded. An additional static variable was given by the number of ICU admissions in the year preceding the considered index ICU stay. Data types of timestamped codes included international classification of diseases and related health problems (ICD-9) diagnoses and procedures codes, prescribed medications and patient vital signs. Overall, the models were trained using 21 static variables, 992 unique ICD-9 diagnostic codes, 298 unique ICD-9 program codes, 586 unique medication therapy codes, and 32 codes related to vital signs. The insertion size was 12 for diagnosis and treatment and 10 for medication and vital signs. The record for each patient contained up to 552 ICD-9 diagnose/procedure codes, as well as 392 medications and vital signs codes related to the current ICU admission.

4.2. Experiment Setup

We wanted to determine whether the model will perform good or poorly for ICU readmission within different time points. To do so, we designed some experiments to verify the predictive effect of the model on readmission by compare with baseline models.

4.2.1. Normal Baseline Models

To verify the predictive performance of the proposed MP-ROM, we compared it with the following three methods:

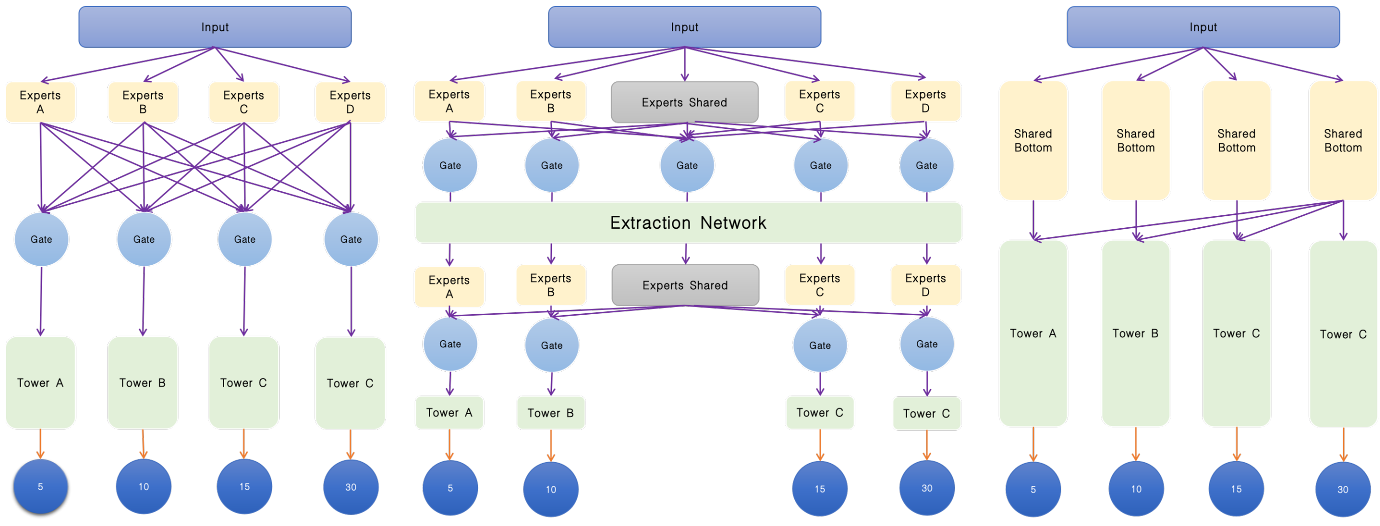
- M-ROM+: Multi-task learning Shared-Bottom model (Loss dynamic weighting).
- M-ROM: Multi-task learning Shared-Bottom model.
- ROM: Single task model.

4.2.2. Multi-task Shared Baseline Models

To verify the Shared-Bottom structure of the proposed MP-ROM is more suitable for these prediction tasks, we compare it with the following three models:

- MMoE: Filter shared features of tasks through double-layer gating mechanism[28].
- PLE: Progressive Layered Extraction considers interactions between different Experts, which is a combination of Customized Sharing and multi-level MMoE[38].
- AS: Uneven Sharing, using 30-day predictive task results as additional information for other predictive tasks.

The shared structure of the three models is shown in Figure 2.



(a) MMoE (b) PLE (c) AS

Figure 2. The structure of the three models.

4.2.3. Model Ablation Study Design

In order to verify the influence of each part of the model on the prediction of the model, we deal with the structure of different modules of the model. For dynamic data of diagnosis, prescription and event information sequences, ordinary differential equation can effectively improve the accuracy of the model. For task correlation processing, the task correlation score obtained by Pearson correlation calculation can effectively improve the accuracy of each task prediction module in the upper layer of the model. In order to verify the results of model ablation, our experiment uses Pearson correlation score and neural ordinary differential equations respectively to predict on the basis of multi-task framework.

- MP-ROM: The model proposed in this study.

- M-ROM: The fusion part of task relevance score was removed from this research model.

- M-RNN: The fusion part of ODE and task relevance score is removed in this research model.

We change the model combination to prove that the use of each part of the model structure has a good effect on improving the prediction accuracy.

4.2.4. Data Ablation Study Design

In order to obtain the results of data ablation, the model dealt with the detailed structure of different types of data. For dynamic data of diagnosis, prescription information sequence and event information sequence, ordinary differential equation could effectively improve the accuracy of the model. To validate the results of data ablation, our experiment used static data, dynamic data and graph attention information constructed using diagnostic code, which in turn contained sequences of diagnostic information and program code, as well as sequences of prescription information and event information.

- MP-ROM: using all data to predict, including static information, diagnoses & procedures information and medication & vital sign scores information.

- SD: using static information and diagnoses & procedures information to predict.

- SC: using static information and medication & vital sign scores information to predict.

- STAT: using static information to predict.

We changed the data combination to prove that the use of each part of information has a good effect on improving the prediction accuracy.

4.2.5. Validation Tasks

The task we selected to evaluate the performance of our proposed model was to predict ICU readmission.

ICU readmission is an indicator of the quality of care. We predicted unplanned admissions within 5, 10, 15 or 30 days following a discharge from an indexed visit. A visit is considered as a readmission if admission date is within 5, 10, 15 or 30 days after discharge of an eligible indexed visit.

4.2.6. Evaluation metrics

The five evaluation metrics used are:

- Average precision: is computed by the area under the two-dimensional curve of precision and recall
- AUROC: Area Under the Receiver Operating Characteristic Curve
- F1: uses a harmonic average combined with recall rate and accuracy
- Sensitivity: is computed by maximising Youden's J statistic[44].
- Specificity: is computed by maximising Youden's J statistic.

Average precision may reflect algorithmic performance on imbalanced data sets better than on AUROC as it does not reward true negatives[45]. The F_1-Score is maximised over different threshold values on risk predictions. Average Precision(AP), F_1-Score, sensitivity (TPR) and specificity (TNR) performance metrics are used to test the accuracy of the methods. We can describe precision, F_1, sensitivity and specificity as:

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

$$F1 = \frac{2TP}{2TP + FP + FN} \tag{7}$$

$$TPR = \frac{TP}{TP + FN} \tag{8}$$

$$TNR = \frac{TN}{FP + TN} \tag{9}$$

where true positive, positive, true negative, false positive and false negative are denoted as TP, TN, FP and FN respectively.

4.2.7. Implementation

To compare several neural network architecture's classification accuracy in training, the maximum likelihood estimation of network parameters were obtained using log-loss cost function in the training data. The RNN and graph attention layer were embedded by dropout of 0.5, and the Adam optimizer with random gradient descent was used (batch size was 128 and the learning rate was 0.001)[46]. Considering the imbalance between classes, the proportionally increased misclassification overhead was allocated to fewer classes[47]. The training ended after 80 epochs because the over-fitting of the training data became apparent, with the additional training of epochs (based on the average accuracy of the validation data).

4.3. Results and Discussion

4.3.1. Overall Performance

Table 1 reports the average accuracy, AUROC, F_1-SCORE, sensitivity and specificity of single-task, multi-task, multi-task loss optimization and our proposed model for the 30-day prediction task. The average precision of multi-task learning model is 0.647, which is 1.1% higher than that of single-task model. However, under the multi-task framework, the generalization ability of the model was improved, and the overall sensitivity was decreased, and the AUROC and F_1 scores were slightly decreased by

Table 1. Summary statistics for the different baseline algorithms used to predict readmission within 30 days of discharge from the intensive care unit.

	ROM	M-ROM	M-ROM+	MP-ROM
Average Precision	0.638	0.647	0.661	**0.666**
AUROC	0.715	0.713	0.723	**0.726**
F1	0.627	0.625	0.628	**0.635**
Sensitivity	**0.62**	0.582	0.534	0.588
Specificity	0.72	0.765	**0.82**	0.77

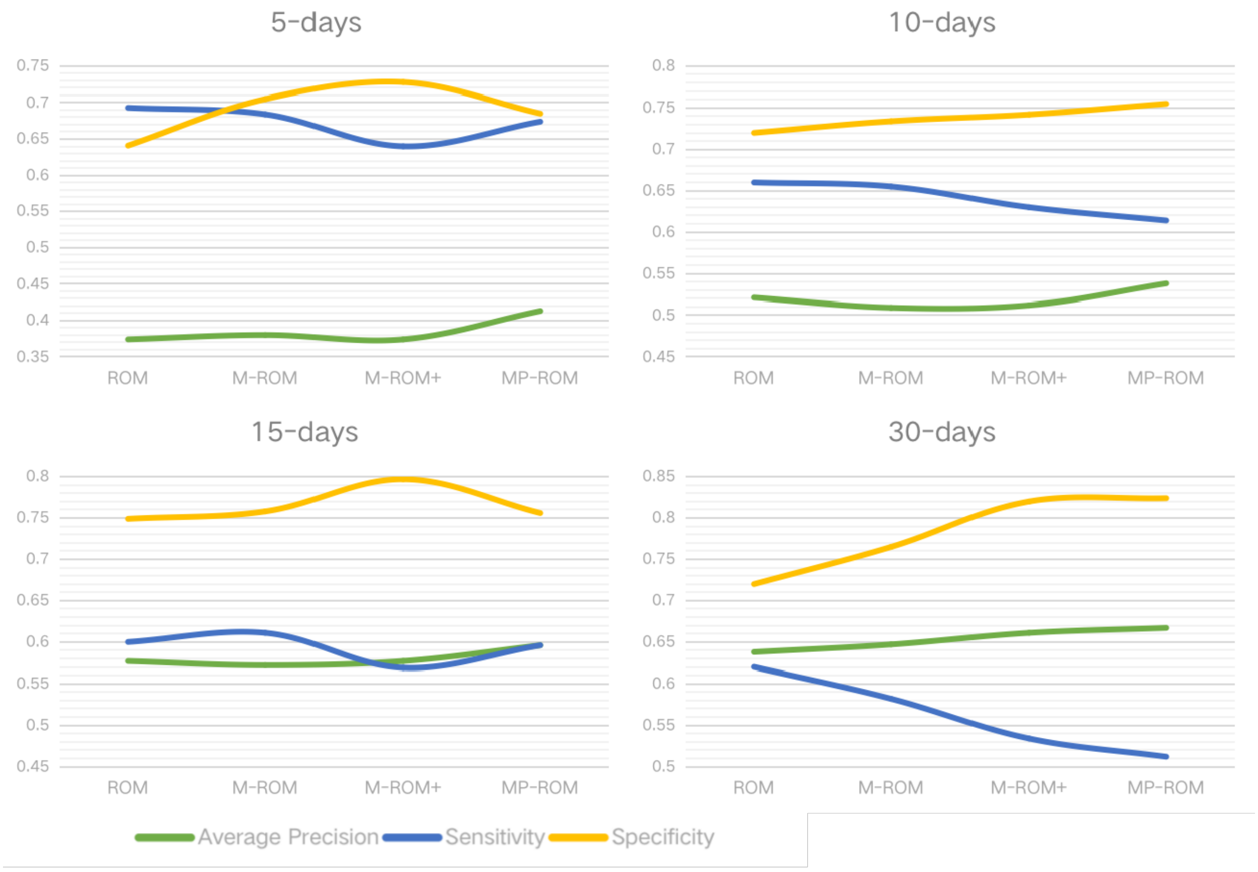


Figure 3. 5-days, 10-days, 15-days and 30-days tasks prediction results of baseline models.

0.2%, respectively. The negative effect of sensitivity on the model was reduced due to a good increase in specificity of 4.5%. Overall, the prediction accuracy of multi-task is slightly better than that of single-task model. Compared with the multi-task loss dynamic weighting model, the specificity of the multi-task loss dynamic weighting model was significantly improved, reaching 0.82, which weakened the negative effect of the sensitivity decreased by 4.8% under the influence of weighting, resulting in better improvement of the average precision and AUROC after optimization, which increased by 1.4% and 1% respectively. It can be concluded that the prediction accuracy of dynamic weight of multi-task loss is obviously higher than that of multi-task learning model. Compared with the multi-task loss dynamic weighting and the proposed model in this study, the average precision of the proposed model in this study is 0.666 and AUROC is highest at 0.726, while F_1 had the highest score of 0.735. The sensitivity of the model was 5.4 % higher than that of the multi-task model with loss optimization. In general, the prediction accuracy of the proposed model is significantly better than that of the multi-task loss dynamic weighting model.

Table 2. Summary statistics for the different multi-task algorithms used to predict readmission within 30 days of discharge from the intensive care unit.

	MMoE	PLE	AS	MP-ROM
Average Precision	0.653	0.653	0.662	**0.666**
AUROC	0.719	0.725	0.724	**0.726**
F1	0.63	0.631	0.629	**0.635**
Sensitivity	0.552	**0.607**	0.565	0.588
Specificity	**0.794**	0.736	0.785	0.77

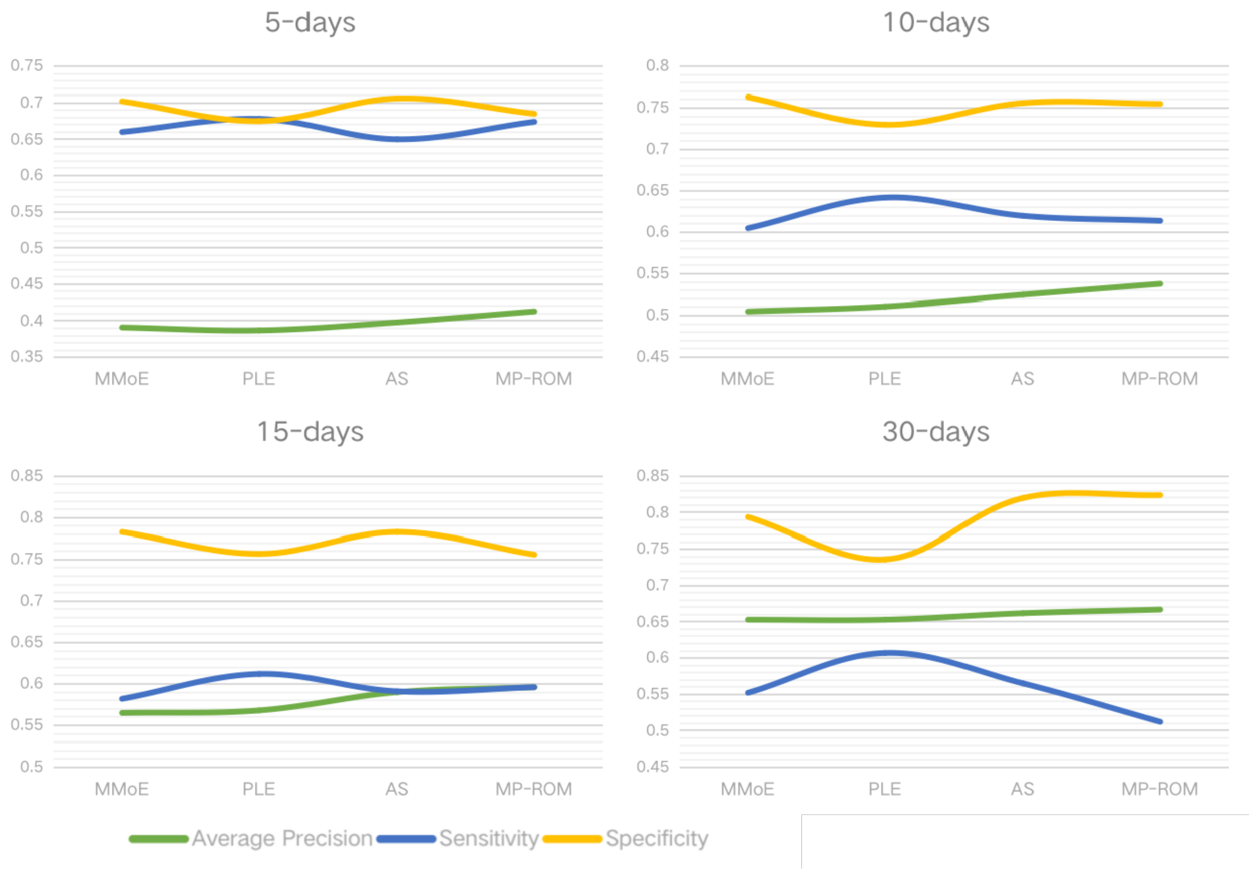


Figure 4. 5-days, 10-days 15-days and 30-days tasks prediction results of multi-task baseline models.

Figure 3 shows the variation trend of average accuracy, sensitivity and specificity of single-task, multi-task, multi-task loss optimization and our proposed model in the prediction tasks of 5, 10, 15 and 30 days. It can be seen that in different prediction tasks, the three evaluation indicators of the model have similar characteristics. The average accuracy of our model is better than the baseline model in all tasks. The specificity is better than the sensitivity in the comparison of different improved models, which is conducive to the tendency of more specificity's improvement in clinical prediction. Under the 30-day prediction task, the sensitivity and specificity of our model fluctuates greatly compared with the baseline model. This is because the large increase in the proportion of positive labels causes a decrease in the accuracy of positive judgments and an increase in the accuracy of negative judgments.

4.3.2. Multi-task Model Comparison Results

Figure 4 shows the variation trend of the average precision, sensitivity and specificity results of MMoE, PLE, AS models and of our proposed models in the 5, 10, 15 and 30-day prediction tasks. It can be seen that among different prediction tasks, the

Table 3. 5, 10, 15, 30 days prediction task model ablation model and proposed model comparison.

Algorithm	Task	AP	AUROC	F1
M-RNN	5-days	0.347	0.673	0.404
M-ROM	prediction	0.374	0.719	0.458
MP-ROM	task	**0.405**	**0.737**	**0.463**
M-RNN	10-days	0.473	0.688	0.505
M-ROM	prediction	0.512	0.724	0.543
MP-ROM	task	**0.538**	**0.743**	**0.557**
M-RNN	15-days	0.55	0.691	0.543
M-ROM	prediction	0.577	0.723	0.575
MP-ROM	task	**0.598**	**0.736**	**0.586**
M-RNN	30-days	0.625	0.695	0.615
M-ROM	prediction	0.661	0.723	0.628
MP-ROM	task	**0.666**	**0.726**	**0.635**

MP-ROM model proposed by us has the highest average precision. Because MMoE and PLE are more suitable for the model with complex task relationship, compared with the Shared-Bottom mechanism of MP-ROM application, the data processed by multi-level gated process will cause additional exposure loss and loss of valuable information, resulting in the multi-task sharing mode with lower prediction accuracy than the Shared-Bottom. Besides, the most models' sensitivity is lower than its specificity due to clinical prediction model is conducive to the tendency of more specificity's improvement.

4.3.3. Model Ablation Results

Table 3 reports that, the average prediction accuracy of the MP-ROM model was higher than that of the model ablation model during the 5, 10, 15 and 30-day prediction tasks. The results showed that RNN alone was not ideal for processing dynamic data, because dynamic data were distributed at a certain time point in the treatment process, and RNN alone could not obtain robust representation of patient treatment process. After processing dynamic information with neural ODE, the prediction effect can be further improved by introducing Pearson calculation of task relation, which indicates that task-correlation score can further improve the prediction performance by strengthening task association.

4.3.4. Data Ablation Results

Table 4 reports the STAT, SC, SD models for data ablation and the average accuracy, AUROC and F1-score of our proposed model (data complete) in the 5-days, 10-days, 15-days and 30-days prediction task. The average precision, AUROC and F1-score of MP-ROM were better than those of other data ablation models, suggesting that the use of all data is more conducive to the prediction of readmission. The MP-ROM improved AUROC and F1-scores compared to the average precision of SD model, and the SC model improved AUROC and F1-score compared to the STAT model, suggesting that medical event and diagnostic sequence data played a positive role in the prediction of ICU readmission within 5, 10, 15 and 30 days in this model. Compared with the SC model, the average precision, AUROC and F1-score of MP-ROM were significantly improved. And the prediction results of SD model were significantly improved compared with STAT model, suggesting that the diagnostic and prescription information sequence data played a positive role. It was proved that diagnostic and prescription information sequence data played a greater role in the model than medical events and diagnostic sequence data.

Analysis of the results showed that static patient information alone was the worst predictor because static information was only a partial static attribute of patients admit-

Table 4. 5, 10, 15, 30 days prediction task data ablation model and proposed model comparison.

Algorithm	Task	AP	AUROC	F1
STAT	5-days prediction task	0.251	0.578	0.344
SC		0.311	0.635	0.375
SD		0.363	0.719	0.445
MP-ROM		**0.405**	**0.737**	**0.463**
STAT	10-days prediction task	0.365	0.593	0.45
SC		0.42	0.632	0.459
SD		0.509	0.728	0.538
MP-ROM		**0.538**	**0.743**	**0.557**
STAT	15-days prediction task	0.423	0.599	0.51
SC		0.486	0.634	0.512
SD		0.574	0.727	0.573
MP-ROM		**0.598**	**0.736**	**0.586**
STAT	30-days prediction task	0.512	0.594	0.593
SC		0.56	0.63	0.597
SD		0.661	0.724	0.629
MP-ROM		**0.666**	**0.726**	**0.635**

ted to ICU and was poorly correlated with readmission. Patients' diagnostic information had the greatest impact on the predictive task, because these diagnostic codes were given by doctors based on the patient's current condition and were highly correlated with the patient's readmission. In addition, dynamic physiological measures of patients have a significant impact on readmission prediction, because these dynamic physiological measures are recorded during the patient's ICU stay and are closely related to changes in the patient's condition, making them valuable information for readmission prediction.

4.3.5. Discussion

There are three major limitations in this study that could be addressed in future research. First, since all dates in the MIMIC-III data set are shifted to protect patient confidentiality, it is not possible to ascertain which patients are admitted after 2001 and have at least 12 months of prior data, possibly leading to some incorrect values for the number of ICU admissions in the year preceding discharge. Second, information from clinical notes[48] is not included and the simplifying assumption is made that various diagnose and procedure-related codes are available immediately at the time of discharge. Third, the week interpretability due to the continuous processing of patients' dynamic data by neural ODEs is difficult to analyze the influence of patient characteristics on the prediction effect.

5. Conclusions

In this paper, we propose a multi-task learning with Pearson and RNN-based neural Ordinary differential equations Model (MP-ROM) to predict ICU readmission. Neural ODE processing irregular interval sequence information is added into recursive neural network as the bottom model.

In addition, the model uses Pearson correlation coefficient to calculate task correlation score to enhance task sharing. The results showed that MP-ROM had better characterization and was validated in a large open source MIMIC-III dataset, effectively improving the predictive performance of ICU readmission.

Acknowledgments: This work was supported in part by the Beijing Information Science and Technology University Qin Xin Talents Cultivation Program under grant no. QXTCPC202112, Research Level Improvement Project under grant no. 2020KYNH214, and Beijing Educational Science Planning Project of China under grant no. CHCA2020102.

References

1. A, R.A.L.; Watts Charles, F.B. Patients Readmitted to ICUs: A Systematic Review of Risk Factors and Outcomes - ScienceDirect. *Chest* **2000**.
2. Xiao, C.; Edward, C.; Sun, J. Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review. *Journal of the American Medical Informatics Association* **2018**.
3. Ab, A.; Vc, B.; Hh, A.; Rkc, C.; Mr, C. FSS-2019-nCov: A deep learning architecture for semi-supervised few-shot segmentation of COVID-19 infection. *Knowledge-Based Systems* **2020**.
4. Jiménez-Cruz, R.; Velázquez-Rodríguez, J.L.; López-Yáez, I.; Villuendas-Rey, Y.; Yáez-Márquez, C. Supervised Classification of Diseases Based on an Improved Associative Algorithm **2021**.
5. Zhang, J.; Kowsari, K.; Harrison, J.H.; Lobo, J.M.; Barnes, L.E. Patient2Vec: A Personalized Interpretable Deep Representation of the Longitudinal Electronic Health Record. *IEEE Access* **2018**.
6. Junqueira, A.; Mirza, F.; Baig, M.M. A machine learning model for predicting ICU readmissions and key risk factors: analysis from a longitudinal health records. *Health and Technology* **2019**.
7. Wang, B.; Ding, S.; Liu, X.; Li, X.; Li, G. Predictive classification of ICU readmission using weight decay random forest. *Future Generation Computer Systems* **2021**.
8. Joel.; Sumner.; Adel.; Alaeddini. Analysis of Feature Extraction Methods for Prediction of 30-Day Hospital Readmissions. *Methods of information in medicine* **2019**.
9. N..; K..; Francis.; A..; Luther.; E..; Salib.; L..; Allanby.; D.. The use of artificial neural networks to predict delayed discharge and readmission in enhanced recovery following laparoscopic colorectal cancer surgery. *Techniques in Coloproctology* **2015**.
10. Huang, J.S.; Chen, Y.F.; Hsu, J.C. Design of a Clinical Decision Support Model for Predicting Pneumonia Readmission. International Symposium on Computer **2014**.
11. Lin, X. A Predictive Random Forest Model on Hospital 30-Day Readmissionusing Electronic Health Records. *Emory University* **2012**.
12. Madrid-García, A.; Font-Urgelles, J.; Vega-Barbas, M.; León-Mateos, L.; Rodriguez-Rodriguez, L. Outpatient Readmission in Rheumatology: A Machine Learning Predictive Model of Patient's Return to the Clinic. *Journal of Clinical Medicine* **2019**.
13. Matheny, M.E.; Ricket, I.; Goodrich, C.A.; Shah, R.U.; Brown, J.R. Development of Electronic Health Record–Based Prediction Models for 30-Day Readmission Risk Among Patients Hospitalized for Acute Myocardial Infarction. *JAMA Network Open* **2021**.
14. Kuris, E.O.; Veeramani, A.; Mcdonald, C.L.; Disilvestro, K.J.; Daniels, A.H. Predicting Readmission After Anterior, Posterior, and Posterior Interbody Lumbar Spinal Fusion: A Neural Network Machine Learning Approach. *World Neurosurgery* **2021**.
15. Ashfaq, A.; Sant'Anna, A.; Lingman, M.; Nowaczyk, S. Readmission prediction using deep learning on electronic health records. *Journal of Biomedical Informatics* **2019**.
16. Nguyen, P.; Tran, T.; Wickramasinghe, N.; Venkatesh, S. Deepr: A Convolutional Net for Medical Records. *IEEE Journal of Biomedical and Health Informatics* **2017**.
17. Liu, X.; Chen, Y.; Bae, J.; Li, H.; Sanger, T. Predicting Heart Failure Readmission from Clinical Notes Using Deep Learning. 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) **2019**.
18. Xie, J.; Zhang, B.; Ma, J.; Zeng, D.D.; Ciganic, J.L. Readmission Prediction for Patients with Heterogeneous Medical History: A Trajectory-Based Deep Learning Approach. *Social Science Electronic Publishing*.
19. Rajkomar, A.; Oren, E.; Chen, K.; Dai, A.M.; Hajaj, N.; Hardt, M.; Liu, P.J.; Liu, X.; Marcus, J.; Sun, M. Scalable and accurate deep learning for electronic health records. *npj Digital Medicine* **2018**.
20. Allam, A.; Nagy, M.; Thoma, G.; Krauthammer, M. Neural networks versus Logistic regression for 30 days all-cause readmission prediction. *Scientific Reports* **2019**.
21. Barbieri, S.; Kemp, J.; Perez-Concha, O.; Kotwal, S.; Jorm, L. Benchmarking Deep Learning Architectures for Predicting Readmission to the ICU and Describing Patients-at-Risk. *Scientific Reports* **2020**.
22. Dorajoo, S.R.; See, V.; Chan, C.T.; Tan, J.Z.; Tan, D.; Razak, S.; Ong, T.T.; Koomanan, N.; Yap, C.W.; Chan, A. Identifying Potentially Avoidable Readmissions: A Medication-Based 15-Day Readmission Risk Stratification Algorithm. *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy* **2017**.
23. Ortiz-Barrios, M.; Altamar-Maldonado, Z.; Martínez-Solano, C.; Petrillo, A.; Medina-Buelvas, A.M. Predicting 15-day unplanned readmissions in hospitalization departments: an application of logistic regression. *Ingeniare* **2021**.
24. Zhang, K.; Zhuang, X. Recognition and standardization of cardiac MRI orientation via multi-tasking learning and deep neural networks **2020**.
25. Deep multi-task prediction of lung cancer and cancer-free progression from censored heterogenous clinical imaging. Image Processing **2020**.
26. Chen, Z.; Badrinarayanan, V.; Lee, C.Y.; Rabinovich, A. GradNorm: Gradient Normalization for Adaptive Loss Balancing in Deep Multitask Networks **2017**.
27. Vandenhende, S.; Georgoulis, S.; Gansbeke, W.V.; Proesmans, M.; Dai, D.; Gool, L.V. Multi-Task Learning for Dense Prediction Tasks: A Survey **2020**.
28. Ma, J.; Zhao, Z.; Yi, X.; Chen, J.; Hong, L.; Chi, E.H. Modeling Task Relationships in Multi-task Learning with Multi-gate Mixture-of-Experts. *ACM* **2018**.

29. Johnson, A.; Pollard, T.J.; Shen, L.; Lehman, L.W.H.; Feng, M.; Ghassemi, M.; Moody, B.; Szolovits, P.; Celi, L.A.; Mark, R.G. MIMIC-III, a freely accessible critical care database. *Scientific Data*.
30. Cho, K.; Merrienboer, B.V.; Gulcehre, C.; BaHdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. *Computer Science* **2014**.
31. Nguyen, P.; Tran, T.; Venkatesh, S. Resset: A Recurrent Model for Sequence of Sets with Applications to Electronic Medical Records. *IEEE* **2018**.
32. Chen, R.T.Q.; Rubanova, Y.; Bettencourt, J.; Duvenaud, D. Neural Ordinary Differential Equations **2018**.
33. Dupont, E.; Doucet, A.; Teh, Y.W. Augmented neural odes. *arXiv preprint arXiv:1904.01681* **2019**.
34. Maurer, A. Bounds for linear multi-task learning. *The Journal of Machine Learning Research* **2006**.
35. Zhang, H.R.; Yang, F.; Wu, S.; Su, W.J.; Ré, C. Sharp Bias-variance Tradeoffs of Hard Parameter Sharing in High-dimensional Linear Regression **2020**.
36. Meyerson, E.; Miikkulainen, R. Beyond Shared Hierarchies: Deep Multitask Learning through Soft Layer Ordering **2017**.
37. Ruder, S.; Bingel, J.; Augenstein, I.; Sgaard, A. Latent Multi-task Architecture Learning **2017**.
38. Progressive Layered Extraction (PLE): A Novel Multi-Task Learning (MTL) Model for Personalized Recommendations. RecSys '20: Fourteenth ACM Conference on Recommender Systems **2020**.
39. Field, A. *Discovering Statistics Using SPSS (2nd Ed* Discovering statistics using R / **2005**.
40. Mu, Y.; Liu, X.; Wang, L. A Pearson's correlation coefficient based decision tree and its parallel implementation. *Information Sciences* **2017**.
41. Lu, Y.; Kumar, A.; Zhai, S.; Yu, C.; Feris, R. Fully-adaptive Feature Sharing in Multi-Task Networks with Applications in Person Attribute Classification. *IEEE Computer Society* **2016**.
42. Puth, M.T.; Neuhaeuser, M.; Ruxton, G.D. Effective use of Spearman's and Kendall's correlation coefficients for association between two measured traits. *Animal Behaviour* **2015**.
43. Nahler.; Gerhard. Pearson correlation coefficient. *Springer Vienna* **2009**.
44. Youden, W.J. Index for rating diagnostic tests. *Cancer* **1950**.
45. Takaya, S.; Marc, R.; Guy, B. The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLoS ONE* **2015**.
46. Kingma, D.; Ba, J. Adam: A Method for Stochastic Optimization. *Computer Science* **2014**.
47. Weiss, G.M.; Mccarthy, K.; Zabar, B. Cost-Sensitive Learning vs. Sampling: Which is Best for Handling Unbalanced Classes with Unequal Error Costs? International Conference on Data Mining. **2007**.
48. Liu, J.; Zhang, Z.; Razavian, N. Deep EHR: Chronic Disease Prediction Using Medical Notes **2018**.