DMNet: A Personalized Risk Assessment Framework for Elderly People With Type 2 Diabetes

Ziyue Yu, *Graduate Student Member, IEEE*, Wuman Luo, *Member, IEEE*, Rita Tse, and Giovanni Pau

***Abstract*—Type 2 diabetes is the most common chronic disease for the elderly people. This disease is difficult to be cured and causes continued medical expenses. The early and personalized risk assessment of type 2 diabetes is necessary. So far, various type 2 diabetes risk prediction methods have been proposed. However, these methods have three major issues: 1) not fully considering the importance of personal information and rating information of healthcare system, 2) not adopting the long-term temporal information, and 3) not comprehensively capturing the correlation between the diabetes risk factor categories. To address these issues, the personalized risk assessment framework for elderly people with type 2 diabetes is needed. However, it is very challenging due to two reasons, namely imbalanced label distribution and high-dimensional features. In this paper, we propose diabetes mellitus network framework (DMNet) for type 2 diabetes risk assessment of elderly people. Specifically, we propose tandem long short-term memory to extract the long-term temporal information of different diabetes risk categories. In addition, the tandem mechanism is used to capture the correlation between the diabetes risk factor categories. To balance the label distribution, we adopt the method of synthetic minority over-sampling technique with Tomek links. To form the better feature representations, we utilize entity embedding to solve the problem of high-dimensional features. To evaluate the performance of our proposed method, we conduct the experiments on a real-world dataset called Research on Early Life and Aging Trends and Effects. The experiment results show that DMNet outperforms the baseline methods in terms of six evaluation metrics (i.e., accuracy of 0.94, balanced accuracy of 0.94, precision of 0.95, F1-score of 0.95, recall of 0.95 and AUC of 0.94).**

***Index Terms*—DMNet, elderly people, personalized risk assessment, Type 2 diabetes.**

Manuscript received 3 June 2022; revised 30 September 2022 and 21 November 2022; accepted 24 December 2022. Date of publication 4 January 2023; date of current version 7 March 2023. This work was supported by the Macao Polytechnic University – Edge Sensing and Computing: Enabling Human-centric (Sustainable) Smart Cities under Grant RP/ESCA-01/2020. *(Corresponding author: Wuman Luo.)*

Ziyue Yu is with the Faculty of Applied Sciences, Macao Polytechnic University, Macao 999078, China (e-mail: ziyue.yu@mpu.edu.mo).

Wuman Luo and Rita Tse are with the Engineering Research Centre of Applied Technology on Machine Translation and Artificial Intelligence of Ministry of Education, Macao Polytechnic University, Macao 999078, China, and also with the Faculty of Applied Sciences, Macao Polytechnic University, Macao 999078, China (e-mail: luowuman@mpu.edu.mo; ritatse@mpu.edu.mo).

Giovanni Pau is with the Department of Computer Science and Engineering, University of Bologna, Bologna, Italy, and also with the Department of Computer Science, University of California, Los Angeles, CA 90095 USA (e-mail: giovanni.pau@unibo.it).

Digital Object Identifier 10.1109/JBHI.2022.3233622

I. Introduction

DIABETES mellitus (DM), commonly known as diabetes, is a global prevalent chronic illness. According to International Diabetes Federation, the number of worldwide diabetes patients is more than 537 million in 2021 [1]. Among the elderly people, type 2 diabetes (T2D) is the most common chronic disease, and the cure of it is difficult [2], [3]. The annual cost of treatment worldwide exceeds $1.3 trillion [4]. Moreover, T2D patients require constant medical care, which places a burden on the health insurance system. To make matters worse, nearly 1 in 2 elderly people with this disease go undiagnosed [5], [6]. Facts have proved that early risk assessment of T2D is the most effective way to alleviate the suffering of the patients and to avoid the high cost of medical treatment [7]. In addition, personalized assessment based on individual differences (e.g., environment, lifestyle, etc.) can lead to more accurate diagnosis, especially in the early stage of the disease [8], [9]. Therefore, in this paper, we focus on personalized early detection of T2D for elderly people.

So far, various T2D risk prediction methods based on behavioural logs and pathophysiological records have been proposed [10], [11], [12]. However, these works have three major issues. First, they failed to explore the personal information (e.g., early life condition) and the rating information of the healthcare systems thoroughly. It has been pointed out in [13] that these two kinds of information are substantial in the diagnosis of diabetes. Second, these methods did not consider the long-term impact of the patients' temporal information. Since diabetes is a chronic disease, capturing the long-term effects of different patients' features can increase the accuracy of diabetes diagnosis [14]. Third, the correlations between the diabetes risk factor categories were ignored. Risk factors are usually organized into different categories. For example, the factors like gender, age and wealth fall under the category of demographic information, while the factors like smoke and exercise fall under the category of lifestyle. Different demographic information has different impacts on adult lifestyle [15]. As such, the correlations between different risk factor categories will influence the chances of getting the diabetes.

To address these issues, a reliable personalized T2D risk assessment framework for elderly people is necessary. However, it is very challenging due to two reasons: 1) imbalanced label distribution is common for diabetes patients' data. It adversely affects the generalization ability of the models, and 2) features in healthcare domain usually have very high dimensionality. The reason is that most works use one-hot encoding to generate large feature representation matrix. This kind of matrices are usually high dimensional and very sparse.

In this paper, we propose a framework called diabetes mellitus network (DMNet) to predict whether the elderly people are at risk of T2D. Specifically, we propose tandem long short-term memory (T-LSTM) to capture long-term temporal information of diabetes risk factor categories. The tandem mechanism can also extract the correlations between different risk factor categories. Besides, we adopt synthetic minority oversampling technique with Tomek links (SMOTETomek) to balance the label distribution. We also utilize entity embedding to generate dense feature representation matrix with much lower dimensions. In addition, the multilayer perceptron (MLP) is applied to process both the personalized information from T-LSTM and the non-personalized rating information of the healthcare systems. We implement the DMNet using the dataset "Research on Early Life and Aging Trends and Effects (RELATE)" [16]. To evaluate the performance of the DMNet, we select accuracy, balanced accuracy, precision, recall, F1-score and Area Under Curve (AUC) as evaluation metrics. The experiment results show that the DMNet outperforms the baseline methods.

In summary, the main contributions of this paper are as follows:

1) We propose DMNet for personalized risk assessment of T2D for elderly individuals. To the best of our knowledge, we are the first to use long-term temporal information (12 years) to predict the risk of T2D for elderly people.
2) We apply entity embedding to form the better five feature representations of risk factor categories. And we design T-LSTM to capture the long-term temporal information and the correlations between the five feature representations.
3) We compare the performance of DMNet and baseline methods on the dataset called RELATE. The experiment results show that DMNet outperforms all baseline methods in terms of accuracy, balanced accuracy, precision, recall, F1-score, and AUC (i.e., accuracy of 0.94, balanced accuracy of 0.94, precision of 0.95, recall of 0.95, F1-score of 0.95, and AUC of 0.94).
4) Compared with other state-of-the-art risk assessment frameworks, our work focus on the elderly people. This work can be widely used in geriatric hospitals and nursing homes.

The remainder of this paper is organized as follows. In Section II, we review related works of diabetes risk assessment. DMNet will be given in Section III. Section IV describes the baseline methods, evaluation methods, experiment environment, performance analysis for DMNet and components experiment. We summarize this paper in Section V.

II. Related Works

Type 2 diabetes is becoming a major chronic disease for elderly people. This disease can cause a series of complications such as stroke, heart disease and impose financial burden to patients [17], [18]. Therefore, the early risk assessment of diabetes is necessary for elderly people. First, in the selection of dataset, most existing works used the dataset called Pima Indian Diabetes (PID) from University of California Irvine (UCI) [19], [20], [21] for personalized diabetes risk assessment. This dataset only includes the medical information such as the number of pregnancies, the BMI, insulin level, etc. However, many studies demonstrated that comprehensive personal information and the rating information of healthcare system can affect the prevalence of diabetes [22]. In addition, the patient will not develop diabetes in the short term [23]. Therefore, comprehensive personal and medical information of elderly people and long-term temporal information need to be included. Furthermore, different categories from personal and medical information exist the potential correlations [24]. Besides, existing works focused on overall population-based diabetes risk assessments, but not for elderly people. However, as the population ages, the proportion of elderly people with diabetes is increasing among all people with diabetes [25]. Therefore, research on personalized risk assessment framework for elderly people with diabetes is urgently needed.

Based on the public dataset above, major works adopted Naive Bayes (NB), random forest (RF), gradient boosting decision tree (GBDT), neural network (NN), LSTM to predict the likelihood of developing diabetes [21], [26], [27], [28]. More specifically, NB is based on Bayes' theorem and the assumption of independence of features. This method is insensitive to missing data and has robustness. However, since there is a correlation between the causes of diabetes [29], [30], the assumption that each feature in NB is independent of each other is incorrect. RF can handle high-dimensional data without dimension reduction [31]. However, RF causes overfitting problems during the training process. GBDT can prevent the problem of overfitting. However, the algorithm is not suitable for high-dimensional sparse features. NN can construct nonlinear models, but the simple network structure can lead to insufficient learning of the information from the features. LSTM includes input gate, output gate, forget gate and memory cell. These gate mechanisms can learn the correlation between hidden states of LSTM. However, using single LSTM cannot capture the correlations between the categories from personal and medical information. More recently, Alex et al. [32] proposed the framework called Deep Convolutional Neural Network (DCNN) to predict whether the people are at risk of T2D. However, convolutional neural network-based framework cannot extract the temporal information. Madan et al. [33] proposed the hybrid model named Convolutional Neural Network-Bidirectional Long Short-term Memory (CNN-BiLSTM) for diabetes risk assessment. However, CNN-BiLSTM failed to capture the correlations between the risk factor categories.

Imbalanced label distribution is a challenge in classification tasks. In the case of imbalanced label, models may predict

Fig. 1. An overview of our work. Briefly, DMNet consists of three parts, namely data preprocessing, model design and classification.

patients will not be at risk for diabetes, which would lead to poor classification accuracy [34]. Random oversampling is a well-known solution to imbalanced label distribution. Basso et al. and Kaltenecker et al. employed random oversampling, which randomly copied samples from the minority class and added them to the training dataset [35], [36]. However, this method can increase the likelihood of occurring overfitting and cause the sample selection bias [37]. Another method to balance the label distribution is Synthetic Minority Oversampling Technique for Nominal and Continuous (SMOTENC) [38]. This method randomly selects another sample from its nearest neighbor for each minority sample, and then randomly synthesizes a new sample within the space of these two selected sample. However, the synthetic data generated by SMOTENC increases the probability of overlapping between the label classes. On the other hand, the high dimensional fea tures are also challenging for risk assessment of diabetes. Existing works used one-hot to encode each feature in the diabetes data, which allows the model to handle categorical features [39]. However, the feature representation matrix formed by one-hot encoding method is sparse and huge, which means there are many "0" in the matrix. The sparse and large feature representation matrix can only store limited information. Therefore, it is not beneficial for the learning of models.

III. DMNet Framework

In this section, we first give the overview of our work in Section III-A. Then, the Section III-B introduces the selected dataset and data preprocessing. After that, we will present the model design of DMNet in Section III-C. Finally, Section III-D introduces the classification of DMNet.

A. Overview of Framework

Fig. 1 shows the overview of DMNet framework. It consists of three parts, namely data preprocessing module, model design module and classification module. First, we perform data cleaning and data sampling method (i.e., SMOTETomek) in the data preprocessing module. After that, model design module is used for learning the personalized feature representations from the elderly people. Specifically, we adopt the entity embedding to generate feature representations matrices. Then, the proposed tandem LSTM is used to capture the correlation between the risk factor categories and temporal information. Finally, the classification module uses the personalized feature representations and the non-personalized feature to predict whether the elderly people are at risk for type 2 diabetes.

TABLE I
Categories and Variables of Selected Dataset

Category	Variables
Demographics	Gender, Age, Marital status, Education, Quintile of income, Wealth indicator
Regime	Hscore, Demographic regime
Early Life Condition	Born in rural not
Adult Lifestyle	Smoke, Exercise, Drink, Number of contacts with doctors, Poor self-reported health, Composite measure of good health
Adult Health Outcome	NadlGT0, NadlGR0alt, NiadlGT0, NadlGT0harm, Bath, Dress, Toilet, Transfer, Bow, Hypertension, Respiratory, Heart disease, Stroke, Arthritis, Number of chronic conditions, Weight, Waist, HIP, BMI, Obesity

Note that Hscore represents the healthcare system rating information of countries/regions based on the evaluation results from World Health Organization. Higher values represent the worse rating for healthcare system. A good healthcare system can reduce the risk of diseases. NadlGT0 means the activities of daily living (ADLs). NadlGR0alt represents the disability scales. Instrumental activities of daily living (IADLs) is defined by NiadlGT0. And NadlGT0harm means the harmonized version based on ADLs, IADLs and disability scales.

B. DMNet: Data Preprocessing

We select the real-world dataset called RELATE supported by the United States Department of Health and Human Services, National Institutes of Health and National Institute on Aging [16]. This dataset is compiled from the major studies of elderly people around the world and covers a 12-year. In addition, the samples of RELATE represent the characteristics of the elderly people in major cities or provinces around the world. Specifically, the dataset includes harmonized data of elderly people from the following regions: 1) Africa (e.g., South Africa and Ghana), 2) Asia (e.g., China and India), 3) Latin America (e.g., Costa Rica), 4) United State (Puerto Rico and Wisconsin) and 5) European (e.g., Netherlands). There are 88273 elderly people's records in the dataset. And the number of variables is 180.

Fig. 2 shows that data preprocessing is the first part of DMNet. Initially, the selected dataset contains 180 variables and 88273 samples. First, we delete variables with missing values, which are caused by restrictions on personal privacy protection laws in different countries or regions. Then, we remove the elderly people's ID because it is the patients' identification code and does not contain useful personal and medical information. After that, there are 35 variables and 23845 elderly people's records in the dataset. And the P values of variables are less than 0.001.

The Table I shows the categories' names and variables. The 35 variables are included by five risk factor categories (i.e., demographics, regime, early life condition, adult lifestyle and

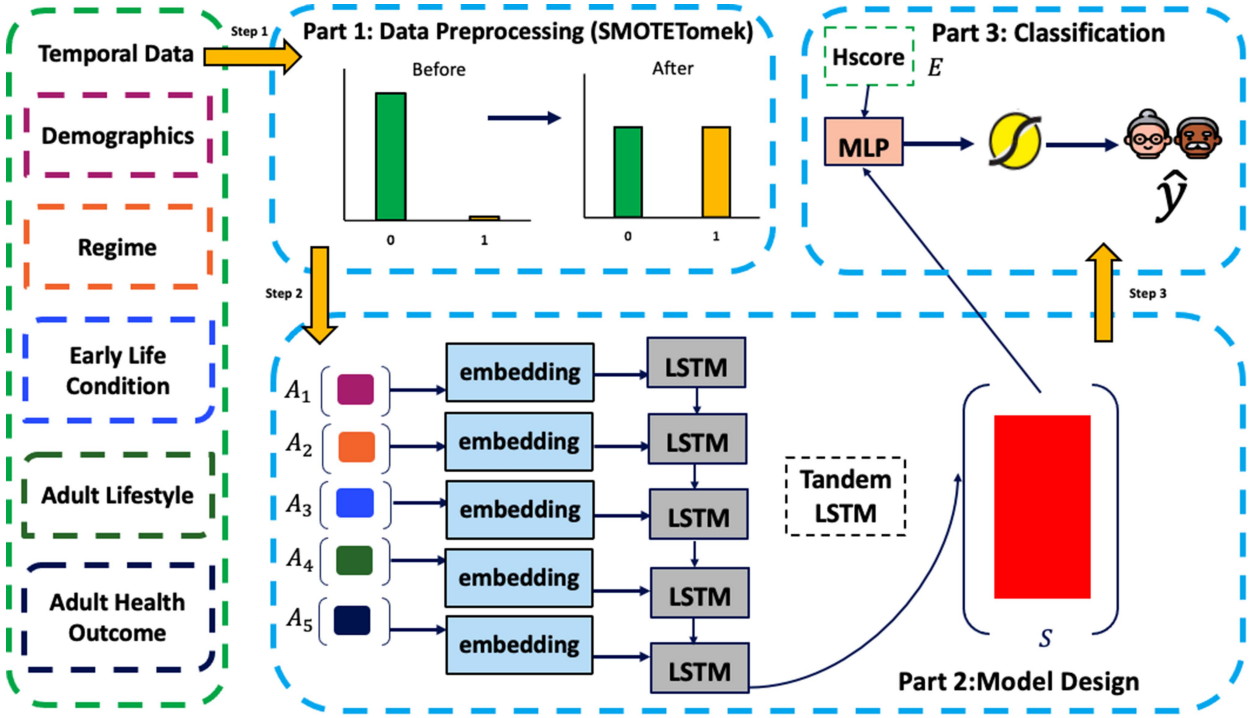

Fig. 2. The structure of DMNet. Part 1 use the elderly people records and apply SMOTETomek to overcome the challenge of imbalanced label distribution. The main components of part 2 include entity embedding and tandem LSTM. In part 3, MLP receives the output from part 2 and use non-personalized variable (i.e., Hscore) to finish the classification. The activation function is sigmoid.

adult health outcome). These categories contain rich and diverse personal and medical risk factors information. For risk factor categories of T2D, we used canonical correlation analysis (CCA) to demonstrate the correlations between the five categories. CCA can be used to calculate the relatedness between two matrices (vectors) with different dimensions [40]. The Fig. 3 shows that there is the positive correlation between the five risk categories. In risk factors, Hscore is numerical variable, while others are categorical variables. The age range of elderly people is 50 to 106 years, and the average of age is 70 years. We use these 35 variables as features. Specifically, Hscore can be regarded as the non-personalized feature. This is because the Hscore represents the evaluation score of a regional/national health system instead of a person. Therefore, Hscore can be used for all residents of the same region/country [41]. The rest of 34 features are the personal profiles (e.g., smoke, obesity) of patients.

The next step is data preprocessing. In the diabetes dataset, only 10.7% (i.e., 2551/23845) are elderly people with T2D and 89.3% (i.e., 21294/23845) are non-diabetic patients. The imbalanced label of dataset can cause the model classifying all elderly people as the non-diabetic patients, thus causing the poor classification performance. In this paper, we adopt SMOTETomek to relieve the negative impact of imbalanced label distribution. Specifically, the SMOTETomek uses Tomek link to remove the overlap between the label classes caused by synthetic data. Tomek link is defined as the pair of samples of opposite label classes that are their own nearest neighbors. The Tomek algorithm is used to find such pairs and cut most samples of the pair.

C. DMNet: Model Design

The second part of DMNet is model design. This part is used to learn the personalized information of the elderly people. The personalized information includes 34 features except Hscore. As a popular method, one-hot encoding was widely used for categorical features. This method converts categorical features into binary vectors where only one value is 1. However, previous studies concluded that the one-hot encoding method led to the explosion of dimension and only stored limited information [42]. Therefore, to reduce the dimensions and form better representation of features, we adopt entity embedding in our work. Entity embedding can transform discrete features into continuous features, which can be applied by various machine learning and deep learning applications [43], [44].

In Fig. 2, A_1, A_2, A_3, A_4, A_5 represent five feature matrices of risk factor categories, namely demographics, regime, early life condition, adult lifestyle and adult health outcome. Each feature matrix is formed by concatenating the feature vectors of each risk factor. Then, the five feature matrices are inputted to five embeddings, respectively. The entity embedding is formalized as:

$$D_i = W_{di} A_i, \tag{1}$$

where W_{di} is mapping matrix. $i = 1, 2, .., 5$ because we have five feature matrices.

After entity embedding, the D_i is used as input for LSTM. In this paper, we concatenate five LSTM called Tandem LSTM (T-LSTM). The reason for proposing T-LSTM is that the risk

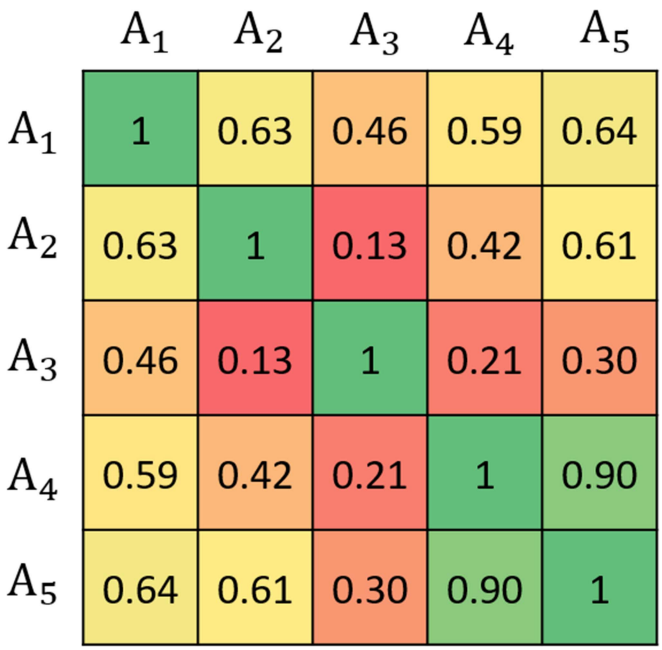

Fig. 3. The correlation for T2D risk factor categories. A positive value denotes a positive correlation. The larger the value, the stronger the correlation. The five diabetes risk factor category matrices are all positively correlated with each other.

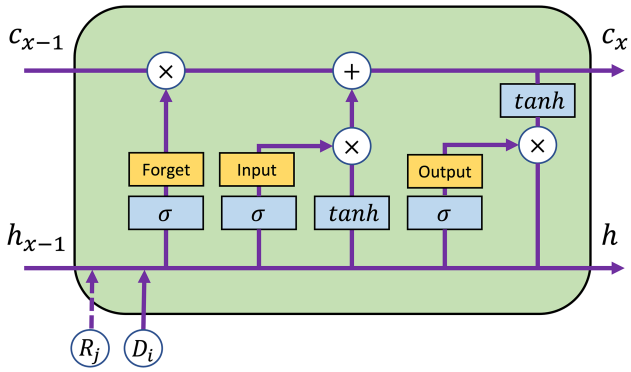

Fig. 4. The structure of T-LSTM. The yellow rectangles represent forget gate, input gate and output gate, respectively. The blue rectangles mean the activation functions. D_i $(1 \leq i \leq 5)$ is the output of each entity embedding, and is used as input for the i-th LSTM. R_j is the output of the j-th LSTM $(1 \leq j \leq 4)$, and is used as input for the $(j+1)$-th LSTM.

factor categories have the potential correlation. In the T-LSTM, the first LSTM only uses the entity embedding result of the first category (i.e., demographics) as input, and the other four LSTM receive the output of the previous LSTM as additional input. The Fig. 4 shows that the main mechanism of T-LSTM includes input gate, output gate, forget gate and candidate memory cell. Mathematically, The first LSTM of T-LSTM is formalized as:

$$I_x = \sigma(W_{ID}D_{i,x} + W_{Ih}h_{x-1} + b_I) \tag{2}$$

$$F_x = \sigma(W_{FD}D_{i,x} + W_{Fh}h_{x-1} + b_F) \tag{3}$$

$$c_x = F_x \odot c_{x-1} + I_x \odot tanh(W_{cD}D_{i,x} + W_{ch}h_{x-1} + b_c) \tag{4}$$

$$O_x = \sigma(W_{OD}D_{i,x} + W_{Oh}h_{x-1} + b_O) \tag{5}$$

$$h_x = O_x \odot tanh(c_x) \tag{6}$$

$$\sigma(l) = \frac{1}{1+e^{-l}} \tag{7}$$

$$tanh(l) = \frac{e^l - e^{-l}}{e^l + e^{-l}} = \frac{1-e^{-2l}}{1+e^{-2l}}, \tag{8}$$

where I_x, F_x, c_x, O_x and h_x represent input gate, forget gate, memory cell, output gate and hidden state, respectively. $D_{i,x}$ is the output of entity embedding. In the first LSTM, the $i = 1$. W denotes the weight matrices and b is bias. $\odot$ means the element wise product. Sigmoid function is represented by σ.

The second to the fifth LSTM of T-LSTM is formulated as follows:

$$I_x = \sigma(W_{ID}D_{i,x} + W_{Ih}h_{x-1} + W_{IR}R_j + b_I) \tag{9}$$

$$F_x = \sigma(W_{FD}D_{i,x} + W_{Fh}h_{x-1} + W_{FR}Rj + b_F) \tag{10}$$

$$O_x = \sigma(W_{OD}D_{i,x} + W_{Oh}h_{x-1} + W_{OR}R_j + b_O), \tag{11}$$

where R_j $(1 \leq j \leq 4)$ represents the output from $(j+1)$-th LSTM. In the second to the fifth LSTM, i is equal to 2, 3, 4, 5, respectively. The calculation of memory cell and hidden state for the second to the fifth LSTM are the same as (4) and (6), respectively. The mechanism of T-LSTM considers the potential influence among the five categories (i.e., demographics, regime, early life condition, adult lifestyle and adult health outcome) to improve the performance of model.

D. DMNet: Classification

The final part of DMNet is classification. The main idea of this part is that different rating information of healthcare system (i.e., Hscore) have different effects on T2D risk assessment. In other words, for the same elderly person, the poorer healthcare system causes the higher risk of developing T2D compared with the better healthcare system. In Fig. 2, the new feature representation matrix S from fifth LSTM is inputted into MLP. In addition, non-personalized rating information of healthcare system is also used in MLP. The classification result is generated using the following equations:

$$\hat{y} = \sigma(W_y S + W_{Hs}E + b_y), \tag{12}$$

where $\hat{y}$ is used to predict whether the elderly people are at risk of developing T2D, and E denotes Hscore. W_y and W_{Hs} are the learnable weight. b_y is the bias of MLP.

In this paper, our loss function is binary cross entropy (BCE). The reason for using the BCE is that it alleviates the negative impact of gradient disappearance when using sigmoid as the activation function [45]. The BCE loss function is formalized as:

$$\text{loss} = -(y \cdot log(\hat{y}) + (1-y) \cdot log(1-\hat{y}), \tag{13}$$

where y is the label of test set, and $\hat{y}$ is the predicted label.

We use adaptive moment estimation (Adam) as optimizer in the training process of model. Adam is an algorithm that performs first-order gradient optimization on a stochastic objective function and can converge faster.

IV. Experiments and Results

In this section, we empirically study the performance of the proposed DMNet. First, we describe the baseline methods for T2D risk assessment. Second, we describe the evaluation metrics of experiments. Then, we describe the experiment environment. Finally, we discuss the experiments analysis (i.e., comparison of DMNet and baseline methods, the performance of components analysis, and explanation of personalized patient representation in T-LSTM).

A. Baseline Methods

To compare with our proposed DMNet, we adopt the naive bayes (NB), random forest (RF), gradient boosting decision tree (GBDT), neural network(NN), and state-of-the-art models DCNN [32], CNN-BiLSTM [33] as baseline methods. Since the baseline methods were not used for temporal diabetes data, we concatenate the temporal information in each feature to form the single feature vector.

1) Naive Bayes: In the diabetes risk assessment task, we can predict the elderly people whether are at risk of developing diabetes by bayes formula.

2) Random Forest: RF is an ensemble learning algorithm. This algorithm samples the training set of diabetes data N times to obtain N training subsets. Each training subset is trained with one decision tree to form the random forest.

3) Gradient Boosting Decision Tree: GBDT is an iterative decision tree algorithm. The algorithm constructs a set of weak learning trees and accumulates the results of multiple decision trees as the final prediction output.

4) Neural Network: NN can be applied to classification tasks. The structure of NN includes one input layer, one hidden layer and one output layer. Nodes in the hidden layer compute the nonlinear transformation of the inputs.

5) Deep Convolutional Neural Network: Alex et al. proposed the DCNN for diabetes classification [32]. In DCNN, the input x first enter the two 1-dimension convolutional layers for feature extraction. Then, one max pooling layer is used for reducing the dimensions of feature representation. After that, the output from the max pooling layer enter two 1-dimension convolutional layers. Finally, one fully connected with sigmoid function is used to evaluate the risk of T2D.

6) Convolutional Neural Network-Bidirectional Long Short-Term Memory: Madan et al. proposed the CNN-BiLSTM for diabetes risk assessment task [33]. The CNN module contains two parts, each of which consists of a convolutional layer and a max pooling layer. The output of CNN module is inputted into Bi-LSTM. Finally, fully connected layer receives the result of Bi-LSTM for diabetes classification.

B. Evaluation Metrics

In this paper, we adopt six evaluation metrics to illustrate the performance of DMNet and baseline methods, i.e., accuracy, balanced accuracy (BA), precision, F1-score, recall and AUC. The formulas and explanations of evaluation metrics are shown in the following.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{14}$$

$$\text{Balanced Accuracy} = \frac{1}{2}\left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP}\right) \tag{15}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{16}$$

$$F1 - \text{score} = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{17}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{18}$$

where true positive (TP) denotes the positive label is correctly predicted as positive. True negative (TN) represents negative label is correctly predicted as negative. False positive (FP) means negative label is incorrectly classified as positive. False negative (FN) represents positive label is incorrectly classified as negative. AUC is used to evaluate the performance of the classifier and it takes the value from 0 to 1 (larger values indicate better classifiers).

C. Experiment Environment

We implement the baseline methods and proposed DMNet on RELATE dataset. The reason we do not perform our method on the PID dataset is that this dataset does not have the risk factor categories. We use Python 3.8 and PyTorch framework for all experiments. The experiments are conducted using a service with a 16 GB memory Intel Core i5-2.0 GHz processor and a 12 GB NVIDIA Tesla K80 graphical processor. For all experiments, the 80% of the data as the training set and the rest as the test set. In training set, we adopt data sampling and 10-fold cross validation. We adopt the 10 random seeds and use the average and standard deviation to evaluate the models' performance. We present the optimal parameters of baseline methods and DMNet as follow. The number of estimators and max depth for random forest and gradient boosting decision tree are 50 and 5, respectively. For NN, the dimension of hidden layer is 10. For DCNN and CNN-BiLSTM, the number of filters is 32. The kernel size is 5. For DMNet, the dimensions of D_1 to D_5 are 15, 5, 5, 10, 30, respectively. The dimensions of hidden layers from five LSTM are 5, 5, 10, 30, 50, respectively. The learning rate of 0.001 and the batch size of 128 are used for DMNet, NN, DCNN and CNN-BiLSTM.

D. Performance Analysis Under Different Data Sampling Methods

The performances of DMNet and baseline methods based on SMOTETomek, SMOTENC and random oversampling in terms of accuracy, balanced accuracy, precision, F1-score, recall and AUC are shown in Tables II, III, and IV, respectively. The average and standard deviation (in parenthesis) of evaluation metrics are reported.

TABLE II
THE PERFORMANCE OF DMNET AND BASELINE METHODS (SMOTETOMEK)

Method	Accuracy	Balanced Accuracy	Precision	F1-score	Recall	AUC
NB	0.82 (0.000)	0.81 (0.002)	0.80 (0.004)	0.82 (0.003)	0.84 (0.007)	0.82 (0.002)
RF	0.79 (0.003)	0.78 (0.009)	0.76 (0.006)	0.79 (0.009)	0.84 (0.016)	0.79 (0.008)
GBDT	0.86 (0.005)	0.86 (0.002)	0.88 (0.018)	0.86 (0.002)	0.83 (0.006)	0.86 (0.002)
NN	0.78 (0.009)	0.78 (0.008)	0.85 (0.004)	0.77 (0.008)	0.79 (0.008)	0.78 (0.008)
DCNN	0.82 (0.000)	0.82 (0.004)	0.86 (0.002)	0.81 (0.002)	0.82 (0.002)	0.81 (0.005)
CNN-BiLSTM	0.83 (0.005)	0.83 (0.004)	0.87 (0.002)	0.82 (0.005)	0.83 (0.005)	0.83 (0.004)
DMNet	**0.94 (0.004)**	**0.94 (0.004)**	**0.95 (0.005)**	**0.95 (0.004)**	**0.95 (0.004)**	**0.94 (0.004)**

TABLE III
THE PERFORMANCE OF DMNET AND BASELINE METHODS (SMOTENC)

Method	Accuracy	Balanced Accuracy	Precision	F1-score	Recall	AUC
NB	0.81 (0.003)	0.80 (0.003)	0.79 (0.006)	0.81 (0.003)	0.83 (0.009)	0.81 (0.003)
RF	0.76 (0.005)	0.75 (0.006)	0.73 (0.003)	0.75 (0.005)	0.83 (0.014)	0.76 (0.006)
GBDT	0.82 (0.001)	0.85 (0.009)	0.86 (0.009)	0.83 (0.002)	0.78 (0.003)	0.82 (0.009)
NN	0.76 (0.005)	0.76 (0.007)	0.84 (0.005)	0.75 (0.008)	0.77 (0.007)	0.76 (0.007)
DCNN	0.80 (0.005)	0.79 (0.004)	0.85 (0.002)	0.79 (0.005)	0.80 (0.005)	0.80 (0.004)
CNN-BiLSTM	0.81 (0.008)	0.81 (0.007)	0.86 (0.002)	0.80 (0.005)	0.81 (0.008)	0.81 (0.008)
DMNet	**0.93 (0.005)**	**0.90 (0.004)**	**0.94 (0.004)**	**0.93 (0.005)**	**0.94 (0.005)**	**0.93 (0.004)**

TABLE IV
THE PERFORMANCE OF DMNET AND BASELINE METHODS (RANDOM OVERSAMPLING)

Method	Accuracy	Balanced Accuracy	Precision	F1-score	Recall	AUC
NB	0.73 (0.002)	0.72 (0.002)	0.73 (0.004)	0.73 (0.003)	0.69 (0.001)	0.72 (0.002)
RF	0.74 (0.002)	0.73 (0.002)	0.72 (0.006)	0.74 (0.002)	0.77 (0.009)	0.74 (0.002)
GBDT	0.75 (0.001)	0.76 (0.006)	0.79 (0.017)	0.75 (0.001)	0.66 (0.012)	0.75 (0.006)
NN	0.75 (0.005)	0.75 (0.004)	0.83 (0.007)	0.74 (0.002)	0.75 (0.004)	0.75 (0.004)
DCNN	0.76 (0.005)	0.77 (0.003)	0.83 (0.002)	0.75 (0.004)	0.76 (0.004)	0.76 (0.003)
CNN-BiLSTM	0.79 (0.008)	0.79 (0.009)	0.85 (0.005)	0.78 (0.012)	0.80 (0.008)	0.79 (0.009)
DMNet	**0.89 (0.008)**	**0.87 (0.009)**	**0.90 (0.006)**	**0.89 (0.008)**	**0.89 (0.010)**	**0.88 (0.009)**

From Tables II, III, and IV, we can find that SMOTETomek-based methods have the better performance than SMOTENC-based and random oversampling-based methods. Therefore, we can conclude that the SMOTETomek can effectively improve the classification performance.

In particular, the experiment results show that the SMOTETomek-based DMNet achieves the best performance, i.e., accuracy of 0.94, balanced accuracy of 0.94, precision of 0.95, F1-score of 0.95, recall of 0.95, AUC of 0.94. The reason why DMNet has better performance is that DMNet can extract long-term temporal information. In addition, the tandem mechanism of DMNet can capture the correlation between different T2D risk factor categories. Therefore, the proposed DMNet can effectively improve the classification performance of T2D risk assessment for elderly people. Besides, we change the input order of five risk categories and the results show that the input order does not affect the model performance.

E. Performance Analysis for Components of DMNet

To illustrate the effectiveness of the proposed framework, we perform the component analysis on DMNet. This analysis aims to explore the impact of entity embedding, non-personalized feature (i.e., Hscore), tandem mechanism of DMNet on T2D risk assessment. Specifically, we design four experiments: 1) single LSTM and DMNet, 2) one-hot encoding-based single LSTM and entity embedding-based single LSTM, 3) one-hot encoding-based DMNet and entity embedding-based DMNet, 4) DMNet without Hscore and DMNet. All experiments are implemented on three data sampling methods (i.e., SMOTETomek, SMOTENC and random oversampling).

1) Single LSTM and DMNet: The tandem mechanism (i.e., T-LSTM) is not present in the experiment. This means that all features are received by single LSTM. The goal is to explore whether capturing the correlation between different risk factor categories through DMNet would perform better than the single LSTM that received all features.

2) One-Hot Encoding-Based Single LSTM and Entity Embedding-Based Single LSTM: In this experiment, we discuss the performance of entity embedding-based single LSTM and one-hot encoding-based single LSTM. The goal is to investigate whether entity embedding outperforms one-hot encoding when both are single LSTM.

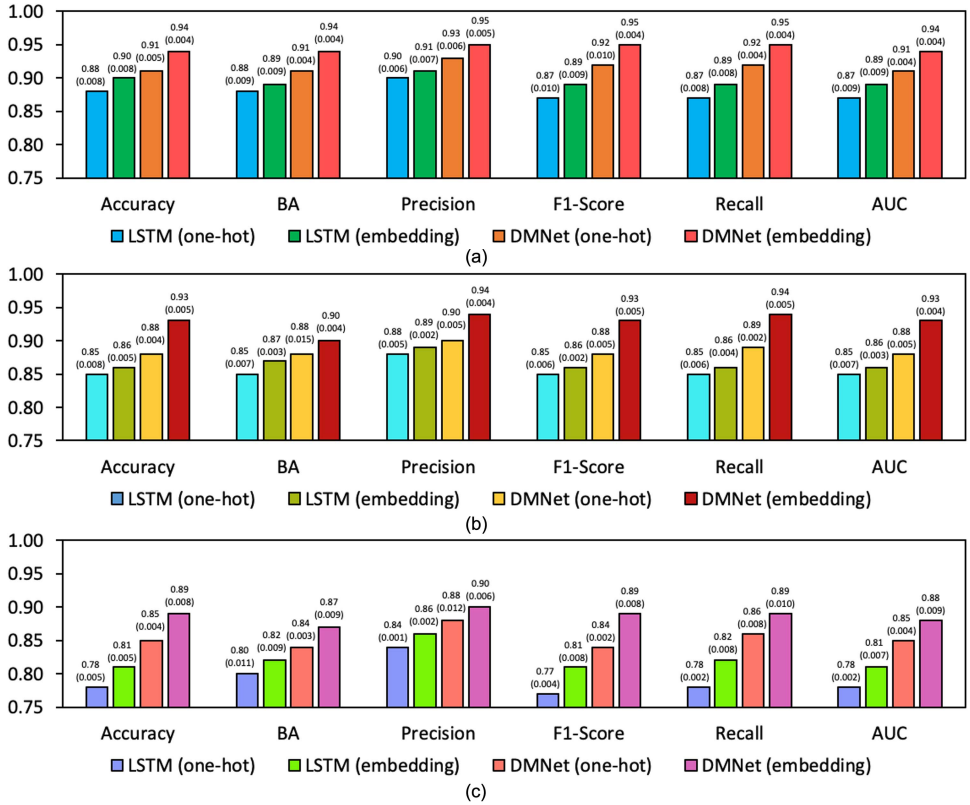

Fig. 5. Component analysis results of single LSTM and DMNet, one-hot encoding-based single LSTM and entity embedding-based single LSTM, one-hot encoding-based DMNet and entity embedding-based DMNet. (a), (b) and (c) are the experimental results based on SMOTETomek, SMOTENC and random oversampling, respectively.

3) One-Hot Encoding-Based DMNet and Entity Embedding-Based DMNet: We adopt one-hot and entity embedding to encode the features from diabetes factor categories on DMNet, respectively. The goal is to explore whether DMNet with entity embedding has better performance than DMNet with one-hot encoding.

4) DMNet Without Hscore and DMNet: In this experiment, we remove the non-personalized feature (i.e., Hscore) for performance analysis. The goal is to explore whether the rating information of healthcare system can effectively increase the classification performance.

Figs. 5 and 6 present the performance of component analysis. The experimental results show that DMNet (entity embedding) achieves the best performance in terms of six evaluation metrics in SMOTETomek, SMOTENC and random oversampling.

Based on Fig. 5, we can see that the tandem mechanism of DMNet has better performance than the single LSTM. In SMOTETomek method, compare with one-hot encoding-based single LSTM, one-hot encoding-based DMNet improves 3.3% by accuracy, 3.3% by balanced accuracy, 3.2% by precision, 5.4% by F1-score, 5.4% by recall, and 4.4% by AUC. On the other hand, compare with entity embedding-based single LSTM, entity embedding-based DMNet improves 4.3% by accuracy, 5.3% by balanced accuracy, 4.2% by precision, 6.3% by F1-score, 6.3% by recall, and 5.3% by AUC. In SMOTENC methods, compare with one-hot encoding-based single LSTM, one-hot encoding-based DMNet improves 3.4% by accuracy, 3.4% by balanced accuracy, 2.2% by precision, 3.4% by F1-score, 4.5% by recall, and 3.4% by AUC. In addition, compare with entity embedding-based single LSTM, entity embedding-based DMNet improves 7.5% by accuracy, 3.3% by balanced accuracy, 5.3% by precision, 7.5% by F1-score, 8.5% by recall, and 7.5% by AUC. In random oversampling methods, compare with one-hot encoding-based single LSTM, one-hot encoding-based DMNet improves 8.2% by accuracy, 4.8% by balanced accuracy, 4.5% by precision, 8.3% by F1-score, 9.3% by recall, and 8.2% by AUC. Besides, compare with entity embedding-based single LSTM, entity embedding-based DMNet improves 9.0% by accuracy, 5.7% by balanced accuracy, 4.4% by precision, 9.0% by F1-score, 7.9% by recall, and 8.0% by AUC. This experiment demonstrates capturing the potential correlation between different risk factor categories by tandem mechanism in DMNet can effectively improve the classification performance of the model.

Fig. 5 also shows the result of one-hot-based single LSTM and entity embedding-based single LSTM. The experimental results show that the entity embedding-based single LSTM outperforms the single LSTM based on one-hot encoding. In SMOTETomek method, compare with one-hot encoding-based single LSTM, entity embedding-based single LSTM improves 2.2% by accuracy, 1.1% by balanced accuracy, 1.1% by precision, 2.2% by F1-score, 2.2% by recall, and 2.2% by AUC.

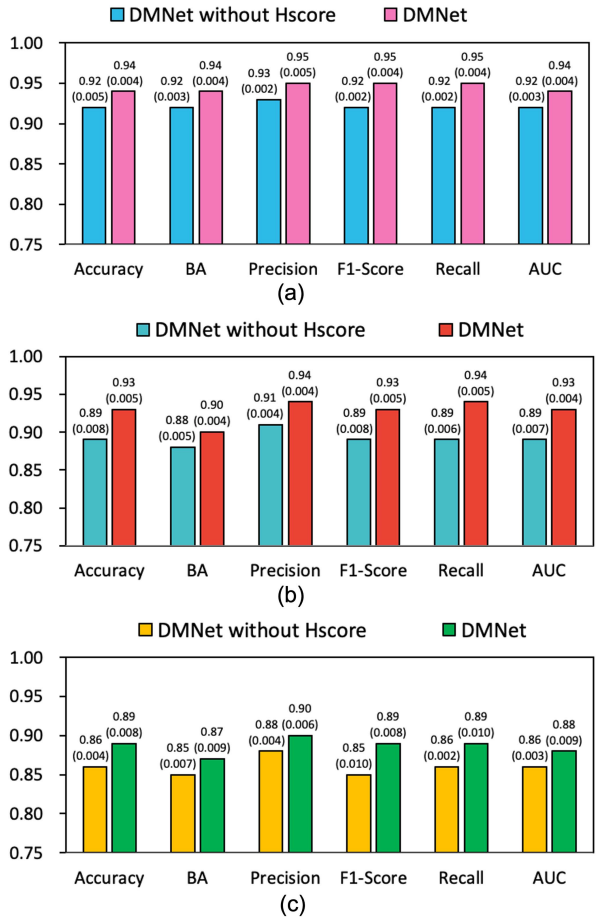

Fig. 6. Results for DMNet without Hscore and DMNet. (a) is SMOTETomek-based experiment result. (b) is SMOTENC-based experiment result. (c) is random oversampling-based experiment result.

In SMOTENC method, compare with one-hot encoding-based single LSTM, entity embedding-based single LSTM improves 1.2% by accuracy, F1-score, recall, AUC, respectively, 2.3% by balanced accuracy and 1.1% by precision. In random oversampling method, compare with one-hot encoding-based single LSTM, entity embedding-based single LSTM improves 3.7% by accuracy, 2.4% by balanced accuracy, 2.3% by precision, 4.9% by F1-score, 4.9% by recall, and 3.7% by AUC. For the performance of one-hot-based DMNet and entity embedding-based DMNet. The results show that DMNet with entity embedding achieves the better performance in terms of six evaluation metrics. In SMOTETomek method, compare with one-hot encoding-based DMNet, entity embedding-based DMNet improves 3.2% by accuracy, balanced accuracy, F1-score, recall, AUC, respectively, and 2.1% by precision. In SMOTENC method, compare with one-hot encoding-based DMNet, entity embedding-based DMNet improves 5.4% by accuracy, 2.2% by balanced accuracy, 4.3% by precision, 5.4% by F1-score, 5.6% by recall, and 5.4% by AUC. In random oversampling method, compare with one-hot encoding-based DMNet, entity embedding-based DMNet improves 4.5% by accuracy, 3.4% by balanced accuracy, 2.2% by precision, 5.6% by F1-score, 3.4% by recall, and 3.4% by AUC. The results of these two component analyses illustrate

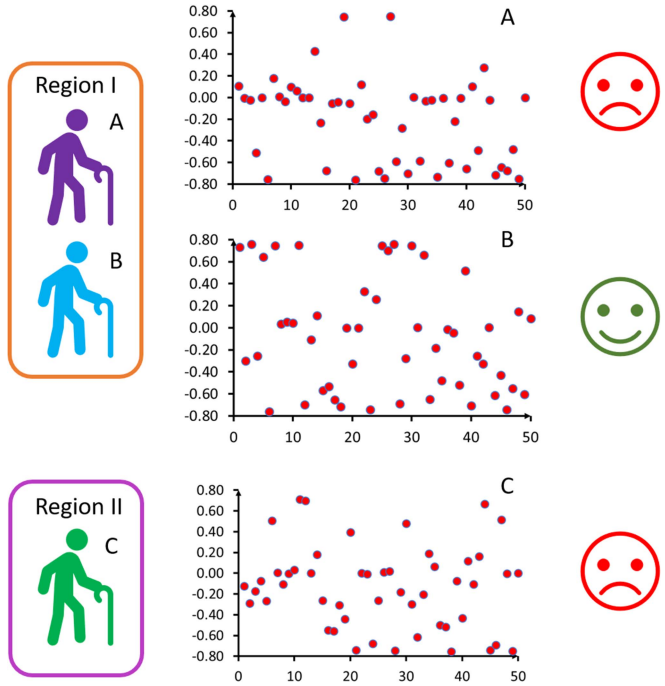

Fig. 7. Examples of personalized feature representation for elderly people. Region I has a poor health system and Region II has a good health system. A smiley face indicates not having diabetes. A sad face means suffering from diabetes.

that entity embedding enables different risk factor categories in dataset to form the better feature representation than one-hot encoding. Therefore, entity embedding can improve the classification performance.

Fig. 6 presents the performance comparison for DMNet without Hscore and DMNet. We can observe that DMNet achieves the better performance in terms of six evaluation metrics. In SMOTETomek method, compare with DMNet without Hscore, DMNet improves 2.1% by accuracy, balanced accuracy, precision, AUC respectively, and 3.2% by F1-score, recall, respectively. In SMOTENC method, DMNet improves 4.3% by accuracy, 2.2% by balanced accuracy, 3.2% by precision, 4.3% by F1-score, 5.3% by recall, and 4.3% by AUC. In random oversampling method, DMNet improves 3.4% by accuracy, 2.3% by balanced accuracy, 2.2% by precision, 3.4% by F1-score, 3.4% by recall, and 2.3% by AUC. This component analysis illustrates that the rating information of healthcare system is critical for T2D risk assessment.

F. Explanation of Personalized Patient Representation in T-LSTM

In this paper, the T-LSTM is proposed for forming the personalized feature representations by learning the 12-years' personal history profiles from five risk factor categories. To better explain the personalized feature representation of elderly people, we visualize the output of T-LSTM. As shown in Fig. 7, the healthcare system of region I has the lowest rating score (i.e.,175), while the healthcare system of Region II has the highest rating score (i.e., 33). Thus, the elderly people living in region I have the

highest chances of having diabetes, while those in region II have the lowest chances of having diabetes. Although these regional information are very useful for the assessment of T2D, we need more personal information to predict whether a specific person having diabetes or not. To illustrate this, we select three persons from these two regions. Specifically, person A and person B are from region I, and person C is from region II. The scatter plots in Fig. 7 show the personalized feature representations (i.e., the outputs of the T-LSTM) of these three persons. The vertical coordinate represents the value in the vector. The horizontal coordinate is the vector dimension of the feature representation. We find that for elderly people with diabetes, there were fewer points with values between 0.6 and 0.8 in the scatter plot. For elderly people without diabetes, there are many points in the range of 0.6 to 0.8. Thus, person B is healthy although he/she comes from a dangerous region, and person C has T2D although he/she comes from a safe region. By utilizing both personalized feature representations and the rating scores of the healthcare system, our model provides correct predictions of all these three persons.

V. Conclusion

In this paper, we propose the personalized T2D risk assessment framework called DMNet for elderly people. This framework includes data preprocessing, model design and classification. We use the real-world data called RELATE as the dataset, which contains five risk factor categories (i.e., demographics, regime, early life condition, adult lifestyle and adult health outcome) with 12 years.

Specifically, the SMOTETomek approach is applied to balance the label distribution in the selected dataset. We adopt entity embedding to construct the better feature representations matrices of different diabetes risk factors categories. The T-LSTM from DMNet is used to capture long-term temporal information and potential correlations between the different risk factor categories. We use the MLP for classifying the risk of having T2D in the elderly people. Besides, non-personalized feature is used as additional input in the MLP. We used NB, RF, GBDT, NN, DCNN and CNN-BiLSTM as the baseline methods, and these six methods have significant contributions to the risk assessment of T2D. In addition, we design the component analysis experiments to demonstrate the components effectiveness of DMNet.

Compared with previous works, this paper has the following advantages. First, our work is conducted on the worldwide population dataset and the characteristics of elderly people in the dataset are representative. Therefore, this dataset can increase the generalization ability of DMNet, and we can use it as the universal tool for risk assessment of elderly people. Second, we compare the performance of state-of-the-art methods and DMNet based on three data sampling methods. The results show that DMNet achieves the better performance. Besides, SMOTETomek can increase the classification performance for all methods. Third, the components analysis proves that entity embedding, rating information of healthcare system and tandem mechanism from DMNet increase the performance of classification. Fourth, the features used in DMNet are available for hospitals and nursing homes. Therefore, DMNet can be used in hospital information systems or nursing systems.

The exploration of the model generality based on the tandem mechanism will be our future work. For example, LSTM can be replaced by other models. This is because the proposed tandem mechanism can be regarded as the general structure, which can be used for the datatset including risk factor categories.

References

[1] H. Sun et al., "IDF diabetes atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045," *Diabetes Res. Clin. Pract.*, vol. 183, 2022, Art. no. 109119.

[2] R. Taylor and A. Barnes, "Can type 2 diabetes be reversed and how can this best be achieved? James lind alliance research priority number one," *Diabetic Med.*, vol. 36, no. 3, pp. 308–315, 2019.

[3] G. S. Meneilly, A. Knip, D. B. Miller, D. Sherifali, D. Tessier, and A. Zahedi, "Diabetes in older people," *Can. J. Diabetes*, vol. 42, pp. S283–S295, 2018.

[4] C. Bommer et al., "Global economic burden of diabetes in adults: Projections from 2015 to 2030," *Diabetes Care*, vol. 41, no. 5, pp. 963–970, 2018.

[5] F. Chentli, S. Azzoug, and S. Mahgoun, "Diabetes mellitus in elderly," *Indian J. Endocrinol. Metab.*, vol. 19, no. 6, 2015, Art. no. 744.

[6] K. Ogurtsova et al., "IDF diabetes atlas: Global estimates of undiagnosed diabetes in adults for 2021," *Diabetes Res. Clin. Pract.*, vol. 183, 2022, Art. no. 109118.

[7] M. Chen et al., "Task-wise split gradient boosting trees for multi-center diabetes prediction," in *Proc. 27th ACM SIGKDD Conf. Knowl. Discov. Data Mining*, 2021, pp. 2663–2673.

[8] A. Shaban-Nejad, M. Michalowski, and D. L. Buckeridge, "Health intelligence: How artificial intelligence transforms population and personalized health," *Nature*, vol. 1, 2018, Art. no. 53.

[9] C. Ni Ki, A. Hosseinian-Far, A. Daneshkhah, and N. Salari, "Topic modelling in precision medicine with its applications in personalized diabetes management," *Expert Syst.*, vol. 39, no. 4, 2022, Art. no. e12774.

[10] D. Dua and C. Graff, "UCI machine learning repository," 2017. [Online]. Available: http://archive.ics.uci.edu/ml

[11] R. S. Lindsay et al., "Adiponectin and development of type 2 diabetes in the Pima Indian population," *Lancet*, vol. 360, no. 9326, pp. 57–58, 2002.

[12] T. Zhu, K. Li, P. Herrero, and P. Georgiou, "Deep learning for diabetes: A systematic review," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 7, pp. 2744–2757, Jul. 2021.

[13] J. A. Seiglie, D. Nambiar, D. Beran, and J. J. Miranda, "To tackle diabetes, science and health systems must take into account social context," *Nature Med.*, vol. 27, no. 2, pp. 193–195, 2021.

[14] I. Contreras et al., "Artificial intelligence for diabetes management and decision support: Literature review," *J. Med. Internet Res.*, vol. 20, no. 5, 2018, Art. no. e10775.

[15] X. Liu et al., "Prevalence and influencing factors of overweight and obesity in a Chinese rural population: The Henan rural cohort study," *Sci. Rep.*, vol. 8, no. 1, pp. 1–11, 2018.

[16] M. McEniry, *Research on Early Life and Aging Trends and Effects (RELATE): A Cross-National Study*, 2013. [Online]. Available: https://hrsdata.isr.umich.edu/data-products/research-early-life-and-aging-trends-and-effects-relate-files

[17] J. B. Cole and J. C. Florez, "Genetics of diabetes mellitus and diabetes complications," *Nature Rev. Nephrol.*, vol. 16, no. 7, pp. 377–390, 2020.

[18] K. Zarkogianni, M. Athanasiou, A. C. Thanopoulou, and K. S. Nikita, "Comparison of machine learning approaches toward assessing the risk of developing cardiovascular disease as a long-term diabetes complication," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 5, pp. 1637–1647, Sep. 2018.

[19] K. Kannadasan, D. R. Edla, and V. Kuppili, "Type 2 diabetes data classification using stacked autoencoders in deep neural networks," *Clin. Epidemiol. Glob. Health*, vol. 7, no. 4, pp. 530–535, 2019.

[20] D. Kaur et al., "Application of Bayesian networks to generate synthetic health data," *J. Amer. Med. Informat. Assoc.*, vol. 28, no. 4, pp. 801–811, 2021.

[21] S. Larabi-Marie-Sainte, L. Aburahmah, R. Almohaini, and T. Saba, "Current techniques for diabetes prediction: Review and case study," *Appl. Sci.*, vol. 9, no. 21, 2019, Art. no. 4604.
[22] V. Borhaninejad, A. Iranpour, M. Shati, A. N. Tahami, G. Yousefzadeh, and R. Fadayevatan, "Predictors of self-care among the elderly with diabetes type 2: Using social cognitive theory," *Diabetes Metabolic Syndrome: Clin. Res. Rev.*, vol. 11, no. 3, pp. 163–166, 2017.
[23] N. Musacchio et al., "Artificial intelligence and Big Data in diabetes care: A position statement of the Italian association of medical diabetologists," *J. Med. Internet Res.*, vol. 22, no. 6, 2020, Art. no. e16922.
[24] E. Agardh, P. Allebeck, J. Hallqvist, T. Moradi, and A. Sidorchuk, "Type 2 diabetes incidence and socio-economic position: A systematic review and meta-analysis," *Int. J. Epidemiol.*, vol. 40, no. 3, pp. 804–818, 2011.
[25] A. Sinclair, P. Saeedi, A. Kaundal, S. Karuranga, B. Malanda, and R. Williams, "Diabetes and global ageing among 65–99-year-old adults: Findings from the international diabetes federation diabetes atlas," *Diabetes Res. Clin. Pract.*, vol. 162, 2020, Art. no. 108078.
[26] H. Lai, H. Huang, K. Keshavjee, A. Guergachi, and X. Gao, "Predictive models for diabetes mellitus using machine learning techniques," *BMC Endocr. Disord.*, vol. 19, no. 1, pp. 1–9, 2019.
[27] L. Zhang, Y. Wang, M. Niu, C. Wang, and Z. Wang, "Machine learning for characterizing risk of type 2 diabetes mellitus in a rural Chinese population: The Henan rural cohort study," *Sci. Rep.*, vol. 10, no. 1, pp. 1–10, 2020.
[28] G. Swapna, R. Vinayakumar, and K. Soman, "Diabetes detection using deep learning algorithms," *ICT Exp.*, vol. 4, no. 4, pp. 243–246, 2018.
[29] T. Dendup, X. Feng, S. Clingan, and T. Astell-Burt, "Environmental risk factors for developing type 2 diabetes mellitus: A systematic review," *Int. J. Environ. Res. Public Health*, vol. 15, no. 1, 2018, Art. no. 78.
[30] N. Den Braver, J. Lakerveld, F. Rutters, L. Schoonmade, J. Brug, and J. Beulens, "Built environmental characteristics and diabetes: A systematic review and meta-analysis," *BMC Med.*, vol. 16, no. 1, pp. 1–26, 2018.
[31] Y. Qi, "Random forest for bioinformatics," in *Ensemble Machine Learning*. Berlin, Germany: Springer, 2012, pp. 307–323.
[32] S. A. Alex, J. Nayahi, H. Shine, and V. Gopirekha, "Deep convolutional neural network for diabetes mellitus prediction," *Neural Comput. Appl.*, vol. 34, no. 2, pp. 1319–1327, 2022.
[33] P. Madan et al., "An optimization-based diabetes prediction model using CNN and bi-directional LSTM in real-time environment," *Appl. Sci.*, vol. 12, no. 8, 2022, Art. no. 3989.
[34] J. Zhang, X. Wu, and V. S. Sheng, "Imbalanced multiple noisy labeling," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 2, pp. 489–503, Feb. 2015.
[35] S. Basso, A. Ceselli, and A. Tettamanzi, "Random sampling and machine learning to understand good decompositions," *Ann. Operations Res.*, vol. 284, no. 2, pp. 501–526, 2020.
[36] C. Kaltenecker, A. Grebhahn, N. Siegmund, and S. Apel, "The interplay of sampling and machine learning for software performance prediction," *IEEE Softw.*, vol. 37, no. 4, pp. 58–66, Jul./Aug. 2020.
[37] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, "A survey on bias and fairness in machine learning," *ACM Comput. Surv.*, vol. 54, no. 6, pp. 1–35, 2021.
[38] Z. S. H. Abad, D. M. Maslove, and J. Lee, "Predicting discharge destination of critically ill patients using machine learning," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 3, pp. 827–837, Mar. 2021.
[39] H. F. Ahmad, H. Mukhtar, H. Alaqail, M. Seliaman, and A. Alhumam, "Investigating health-related features and their impact on the prediction of diabetes using machine learning," *Appl. Sci.*, vol. 11, no. 3, 2021, Art. no. 1173.
[40] X. Yang, W. Liu, W. Liu, and D. Tao, "A survey on canonical correlation analysis," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 6, pp. 2349–2368, Jun. 2021.
[41] W. H. Organization, *The World Health Report 2000: Health Systems: Improving Performance*. Geneva, Switzerland: World Health Organization, 2000.
[42] H. Wang, Z. Cui, Y. Chen, M. Avidan, A. B. Abdallah, and A. Kronzer, "Predicting hospital readmission via cost-sensitive deep learning," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 15, no. 6, pp. 1968–1978, Nov./Dec. 2018.
[43] S. Chen, J. Wang, F. Jiang, and C.-Y. Lin, "Improving entity linking by modeling latent entity type information," in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 7529–7537.
[44] Y. Li et al., "Self-attention enhanced selective gate with entity-aware embedding for distantly supervised relation extraction," in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 8269–8276.
[45] A. S. Bosman, A. Engelbrecht, and M. Helbig, "Visualising basins of attraction for the cross-entropy and the squared error neural network loss functions," *Neurocomputing*, vol. 400, pp. 113–136, 2020.