

RNN을 이용한 제2형 당뇨병 예측모델 개발

장진수¹, 이민준², 이태노^{3*}

¹고려대학교 대학원 보건과학과 BK21플러스 인간생명-사회환경 상호작용융합사업단 박사과정,
²고려대학교 대학원 보건과학과 BK21플러스 인간생명-사회환경 상호작용융합사업단 석박사통합과정,
³고려대학교 대학원 보건과학과 BK21플러스 인간생명-사회환경 상호작용융합사업단 교수

Development of T2DM Prediction Model Using RNN

Jin-Su Jang¹, Min-Jun Lee², Tae-Ro Lee^{3*}

¹BK21PLUS Program in Embodiment: Health-Society Interaction, Department of Health Science, Graduate School, Korea University, Doctor's Course

²BK21PLUS Program in Embodiment: Health-Society Interaction, Department of Health Science, Graduate School, Korea University, Integrated PhD program

³BK21PLUS Program in Embodiment: Health-Society Interaction, Department of Health Science, Graduate School, Korea University, Professor

요약 제2형 당뇨병은 고혈당이 특징인 대사성 분비 장애로 여러 합병증을 야기하는 질병이며, 장기적인 치료가 필요하기 때문에 매년 많은 의료비를 지출한다. 이를 해결하기 위해 많은 연구들이 있어왔지만, 기존의 연구들은 한 시점에서의 데이터를 학습시켜 예측함으로써 정확도가 높지 않았다. 그래서 본 연구는 제2형 당뇨병 발생 예측에 대한 정확도를 높이기 위하여 RNN을 이용한 모델을 제안하였다. 본 모델을 개발하기 위해 한국인유전체역학조사 지역사회 코호트(안산-안성) 데이터를 이용하였으며, 시간의 흐름에 따른 데이터들을 모두 학습시켜 당뇨병 발생 예측모델을 만들었다. 예측 모델의 성능을 검증하기 위해 기존의 기계 학습 방법인 LR, k-NN, SVM과 정확도를 비교하였다. 비교한 결과 제안한 예측모델의 accuracy는 0.92, AUC는 0.92로 다른 기계 학습 방법보다 높은 정확도를 보였다. 따라서 본 연구에서 제안한 제2형 당뇨병 발생 예측 모델을 활용하여 발병을 조기 예측함으로써 생활습관 개선 및 혈당조절을 통해 당뇨병 발병을 예방하고 늦출 수 있을 것이다.

주제어 : 제2형 당뇨병, 질병 예측, 기계 학습, 딥러닝, RNN, 의료 인공지능

Abstract Type 2 diabetes mellitus(T2DM) is included in metabolic disorders characterized by hyperglycemia, which causes many complications, and requires long-term treatment resulting in massive medical expenses each year. There have been many studies to solve this problem, but the existing studies have not been accurate by learning and predicting the data at specific time point. Thus, this study proposed a model using RNN to increase the accuracy of prediction of T2DM. This work propose a T2DM prediction model based on Korean Genome and Epidemiology study(Ansan, Anseong Korea). We trained all of the data over time to create prediction model of diabetes. To verify the results of the prediction model, we compared the accuracy with the existing machine learning methods, LR, k-NN, and SVM. Proposed prediction model accuracy was 0.92 and the AUC was 0.92, which were higher than the other. Therefore predicting the onset of T2DM by using the proposed diabetes prediction model in this study, it could lead to healthier lifestyle and hyperglycemic control resulting in lower risk of diabetes by alerted diabetes occurrence.

Key Words : T2DM, Disease Prediction, Machine Learning, Deep Learning, RNN, Medical AI

*Corresponding Author : Tae-Ro Lee(trlee@korea.ac.kr)

Received May 15, 2019
Accepted August 20, 2019

Revised June 19, 2019
Published August 28, 2019

1. 서론

당뇨병은 인슐린 분비 장애 및 인슐린 작용의 결함으로 인해 발생한 고혈당이 특징인 대사성 분비 장애이다. 그리고 당뇨병은 눈, 신장 및 신경에 영향을 미치고, 심혈관 질환에 대한 위험 증가 등 여러 합병증을 야기하는 질병으로 알려져 있다[1]. 게다가 당뇨병으로 인한 과혈당이 오랜 기간 몸 속 장기에 영향을 미치게 되면 결국에는 의식 불명 상태, 혼수상태, 심지어 죽음을 초래한다. 당뇨병은 크게 제1형 당뇨병, 제2형 당뇨병, 임신 당뇨병으로 나뉜다. 이 중에서 제2형 당뇨병(T2DM : Type 2 Diabetes Mellitus)은 성인기에 인슐린 저항성과 인슐린 분비기능 저하로 인해 발병하는 후천성 질병이다.

현재 전 세계적으로 T2DM의 발생률은 전례없이 빠른 속도로 증가하고 있다. 국제당뇨병연맹(IDF : International Diabetes Federation)은 18~99세 성인 11명 중 1명(약 4억 1,500만 명)이 2015년 전 세계적으로 당뇨병을 앓고 있다고 추정했다. 이 중에서 90% 이상이 T2DM 환자로 분류되고 있다[2]. 이와 같은 추정치는 2045년까지 6억 9천 3백만 명으로 증가할 것으로 예상하고 있다[3]. 국내의 경우 2015년 발표된 국민건강영양조사에 따르면 30세 이상의 한국인 당뇨 유병률은 남성이 11.1%, 여성이 8.0%로 보고되었으며[4], 2030년에 30세 이상의 당뇨 유병률은 남자 29.2%, 여자 19.7%으로 예측하고 있다[5].

당뇨병은 장기적인 치료가 필요하기 때문에 많은 의료비와 전문 의료인이 필요한 질병이지만 초기에 잘 관리하면 치료효과가 매우 큰 질병이다. 하지만 당뇨병은 초기 단계에 특별한 증상을 보이지 않아서 증상이 악화될 때까지 병원에 가지 않는 경우가 많다. 미국 당뇨병 협회(American Diabetes Association)에서 발표한 자료에 따르면, 당뇨병을 조기 진단받고 질병관리를 한다면 당뇨병과 관련된 합병증을 줄이는 데 효과적임을 알 수 있다[6]. 또한 당뇨병 환자의 조기 진단은 T2DM으로 인한 합병증을 예방하거나 지연시킬 수 있음이 밝혀졌고[7], 최적화된 예측모형을 통해 T2DM을 조기 예측함으로써 생활습관 개선 및 혈당조절이 가능하며, 고위험군의 T2DM 발생률을 낮출 수 있다고 알려져 있다[8]. 이는 당뇨병의 조기발견 및 예측이 당뇨병 치료에 매우 효과적이라는 것을 보여준다. 현재까지 당뇨병 위험을 예측하기 위한 전통적인 방법은 인구통계학적 데이터와 임상적 데이터를 사용하여 위험 수준을 파악하는 통계 모델을 구축하는 데 중점을 두었다[9-13]

최근에 의학 분야에서 기계 학습을 활용한 질병 발생

예측 분야가 주목을 받고 있다. 기계 학습이란, 컴퓨터가 인간과 같이 학습하고 추론하도록 유도하는 기술이다. 컴퓨터는 현실세계의 관측값을 바탕으로 현실세계 정보 간의 상호 작용을 자율적으로 학습하여 원하는 결과를 추론한다. 기계 학습은 임상 데이터를 적용하여 질병의 진행 상태 및 입원 기간 예측과 같은 작업에서 높은 예측률을 증명하였으며, 임상 의사결정 지원시스템(CDSS : Clinical Decision Support Machine)에 큰 기여를 할 것으로 예상되어 왔다. 특히 최근에는 시계열 데이터를 이용한 질병 발생 예측 연구가 점차 확대되고 있다.

Lipton et al. [14]은 임상 데이터 세트에 LSTM을 적용한 모델을 제안했다. 저자들은 Children 's Intensive Care Unit(ICU) 데이터 세트에서 LSTM을 사용하여 13개의 병리적 테스트 결과를 사용하여 여러 가지 질병 발생(예 : 천식, 고혈압 및 빈혈)을 예측했다. 또한 Choi, Edward, et al.의 연구를 살펴보면[15], RNN 기술 중 GRU(Gated Recurrent Unit)로 일반 환자의 임상 기록 중 시간 순서로 측정된 환자 데이터를 이용하여 예측 모델을 만들었다. Lipton et al.의 연구와 마찬가지로 Choi, Edward, et al.의 연구는 주로 질병 발생을 예측하는 것이었다. [14,15]에서 연구한 질병 예측 모델은 다중 질병 발생 예측과 관련하여 유의미한 결과를 보여주었다.

이러한 기계 학습을 이용한 모델링 및 예측 기술은 다양한 종류의 건강 데이터를 컴파일하고 학습할 수 있는 최상의 방법으로 꼽히고 있다. 특히 딥러닝 기술 중에서도 RNN(Recurrent Neural Network)은 시계열 데이터를 바탕으로 다음에 일어날 미래를 예측하는 데 중점을 두었기 때문에 기존의 기계학습보다 더 정확한 질병 예측이 가능하다. 또한, 대부분의 임상 데이터는 시간의 순서에 따라 작성되기 때문에 RNN의 적용이 가능하다.

따라서 본 연구는 딥러닝 방법론 중 시계열 데이터를 바탕으로 미래를 예측하는 데 중점을 둔 RNN을 이용하여 일반인들의 당뇨병 발생을 예측할 수 있는 모델을 만들었다. 데이터는 국내 코호트 데이터인 한국인유전체역학조사(KoGES : Korean Genome and Epidemiology Study)의 지역사회 기반 코호트(안산, 안성) 자료를 사용하였다.

2. 연구방법

2.1 연구대상

본 연구에서 사용된 자료원은 질병관리본부에서 질병 연구 기반구축을 위해 대규모 추적 조사한 '한국인유전체역

학조사(KoGES) 중 안성(농촌), 안산(도시) 지역사회 코호트 자료이다. 자료원의 수집 기간은 2001년 ~ 2016년까지이고, 2년에 한 번씩 총 7번 추적 관찰한 자료이다. 대상 연령은 40 대 이상의 성인남녀이고, 관찰대상 수는 10,038명이다.

본 연구는 자료원에서 연속형 변수만을 이용하였다. 연속형 변수는 혈당, 당화혈색소(HbA1C), 혈액요소질소(BUN), 간 수치(AST, ALT), 콜레스테롤, 인슐린 등 혈액 검사를 통해 분석한 수치와 소변 산도, 요비중(Urine-SG) 등 소변 검사를 분석한 수치, 그리고 혈압, 키, 몸무게, BMI, 허리둘레 등 신체 계측치를 포함하여 총 67개 변수를 사용하였다.

2.2 데이터 전처리

본 연구에서는 당뇨병 환자군과 대조군으로 데이터를 나누었다. 먼저 1차 조사 이전에 당뇨를 겪은 환자를 제외하고, 다음과 같은 기준으로 당뇨병 환자군과 대조군을 나누었다. 첫째, HbA1C ≥ 6.5 이고, 둘째, 당뇨약의 복용 유무이다. 왜냐하면, 공복혈당 측정이나 경구당부하검사를 위해서는 반드시 금식이 필요하지만, 데이터 전처리 과정 중 금식하지 않은 상태에서 검사를 받은 관찰 대상자가 많아 보였기 때문이다. 그러므로 본 연구에서는 최근 3개월간 헤모글로빈의 평균 혈당 상태를 반영하는 것으로 알려진 당화혈색소만을 변수로 채택하였다. 국내 연구를 살펴보면, 당뇨병 진단 기준으로서 당화혈색소 측정이 유용하다는 연구 결과가 있다[16].

위와 같은 기준으로 당뇨병 환자군을 추적 기간별로 분류해 보면 Table 1과 같다. 당뇨병 환자군과 대조군은 똑같이 759명으로 맞춰 자료량의 치우침을 방지하였다. Table 1을 보면, 1기에 해당하는 당뇨병 환자군은 제외되었다. 그 이유는 본 연구가 시계열 자료를 이용한 질병발생 예측 모델을 개발하는 것이기에 때문에, 이전 자료가 없는 1기의 당뇨병 환자군의 데이터를 포함시키는 것은 적합하지 않다.

Table 1. Diabetic patients and control group in each time period

Routing period	Diabetes patients group	Control group
1 st follow-up period	exclusion	759
2 nd follow-up period	202	
3 rd follow-up period	145	
4 th follow-up period	108	
5 th follow-up period	115	
6 th follow-up period	106	
7 th follow-up period	83	
Total	759	759

모든 환자군과 대조군은 ID별로 조사한 시기에 따라 순서대로 데이터를 정렬하였다. 각 ID마다 시퀀스 별로 67개의 연속형 변수로 구성하였다. 시퀀스 간의 간격은 1~2년이다.

2.3 LSTM 네트워크

LSTM은 RNN 네트워크의 한 종류이고, 1997년 Hochreiter, Sepp, and Jürgen Schmidhuber에 의해 제안되었다[17]. LSTM은 cell state와 게이트를 활용하여 기존의 전통적인 RNN 네트워크의 문제점인 데이터와 데이터 사이의 시간 간격이 멀 경우 학습능력이 크게 저하되는 것을 해결하기 위해 고안되었다. LSTM의 기본 구조는 Fig. 1처럼 입력 게이트(input gate) i , 망각 게이트(forget gate) f , 제어 게이트(control gate) c 및 출력 게이트(output gate) o 의 4개의 게이트를 갖는 LSTM 셀의 구조를 나타낸다.

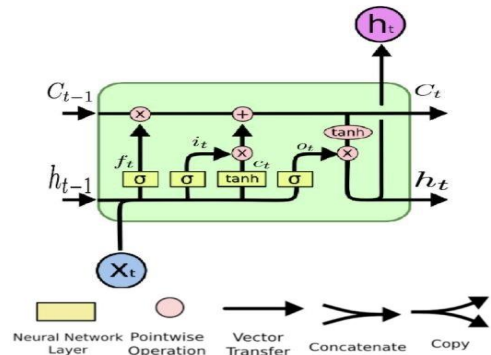


Fig. 1. Basic Structure Design of the LSTM cell

입력 게이트는 새로 들어오는 정보 중 어떤 것을 cell state에 저장할 것인지를 정하는 역할을 수행한다. 이 수식은 (1)과 같다.

$$i_t = \sigma(W_{xh_i}x_t + W_{hh_i}h_{t-1} + b_{h_i}) \quad (1)$$

망각 게이트는 cell state로부터 어떤 입력 정보를 버릴 것인지를 정하는 역할을 수행하며, 수식은 (2)와 같다.

$$f_t = \sigma(W_{xh_f}x_t + W_{hh_f}h_{t-1} + b_{h_f}) \quad (2)$$

제어 게이트는 이전 state인 C_{t-1} 를 업데이트해서 새로운 cell state인 C_t 를 만드는 역할을 수행한다. 수식은 (3), (4)과 같다.

$$g_t = \tanh(W_{xh_g}x_t + W_{hh_g}h_{t-1} + b_{h_g}) \quad (3)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \quad (4)$$

출력 게이트는 히든 벡터 h_{t-1} 에 출력과 업데이트를 수행하는 역할을 한다. 수식은 (5), (6)과 같다.

$$o_t = \sigma(W_{xh_o}x_t + W_{hh_o}h_{t-1} + b_{h_o}) \quad (5)$$

$$h_t = o_t \odot \tanh(c_t) \quad (6)$$

(1)에서 (6)의 계산 중 σ 는 시그모이드 활성화 함수이고, W 는 weight 계산 행렬이다. 그리고 \tanh 는 -1 에서 1 사이의 범위 값을 사용하는 함수이다.

2.4 예측모델 설계

본 연구에서는 T2DM 발생을 예측하기 위하여 LSTM 모델을 설계하였다. 모델 구축에 사용한 소프트웨어는 Python 3.7.1 환경에서 다차원 신경망 API(Application Programming Interface)인 Keras이다. 모델의 학습은 데이터의 80%, 확인(validation)은 데이터의 10%를 사용하였다. 나머지 10%는 테스트에 사용하였다. 모델은 0.001의 학습률을 가진 adam optimizer이며, 손실 함수(loss function)는 binary cross-entropy를 이용하였다.

또한 5 개의 LSTM 레이어의 출력공간(units)은 50, 20, 20, 20, 20개로 설정하였으며, LSTM 레이어의 활성화 함수는 sigmoid를 사용했고, 각 레이어의 drop out 은 0.2로 설정하였다. 1개의 완전 연결 레이어(fully connected layer)의 활성화 함수(activation function)는 RELU를 이용하였다. 출력은 최종 예측된 당뇨 발생 환자를 출력하는 데 사용하였다. Fig. 2는 각 레이어의 입력 및 출력 크기를 나타낸 것이다. 이 모델은 10,000 반복(epoch)으로 고정하였다.

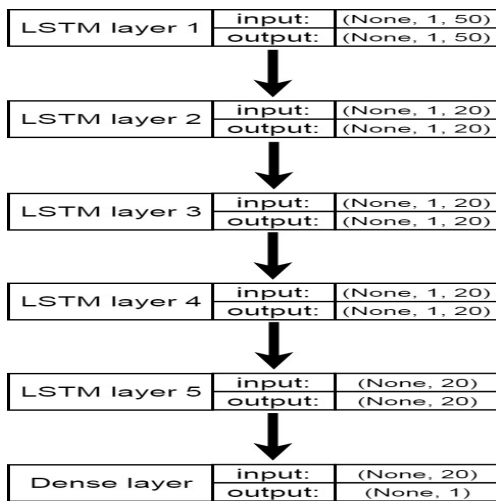


Fig. 2. Input and output dimensions of each layer

또한 본 연구의 구현 환경은 H/W는 AWS(Amazon Web Service) EC2의 GPU 기반 컴퓨팅의 장점을 활용하기 위해 구축된 g3.4xlarge 인스턴스를 사용하였고, S/W는 Pycharm을 사용하였다.

3. 연구결과

3.1 성능 비교를 위한 기저모델

RNN을 이용한 당뇨병 예측 모델의 성능 평가는 기존의 기계 학습에서 널리 사용되고 있는 모델에 동일한 데이터 세트로 학습시키고 테스트하여 비교 분석하였다. 본 연구에서 비교대상 모델은 LR(Logistic Regression), k-NN(K-Nearest Neighbor), SVM(Support Vector Machine)이다.

3.2 모델의 정확도 평가

구축한 모델을 평가하기 위해 대표적인 판단 기준인 accuracy와 AUC(Area under the curve)로 모델의 정확도를 보고자 했다. Base line 모델들은 데이터에서 시간적 동역학(temporal dynamics)을 모델링할 수 없으므로 각 ID 별로 한 시퀀스 전의 데이터를 바탕으로 다음 시퀀스에 발생하는지 예측하는 모델을 만들었다.

모든 모델 간의 교차 검증한 accuracy는 Fig 3.과 같다. 모델 간의 accuracy는 LR, k-NN, SVM, RNN이 각각 0.75, 0.81, 0.76, 0.92로 RNN 방법으로 구축한 모델의 accuracy가 가장 높았다.

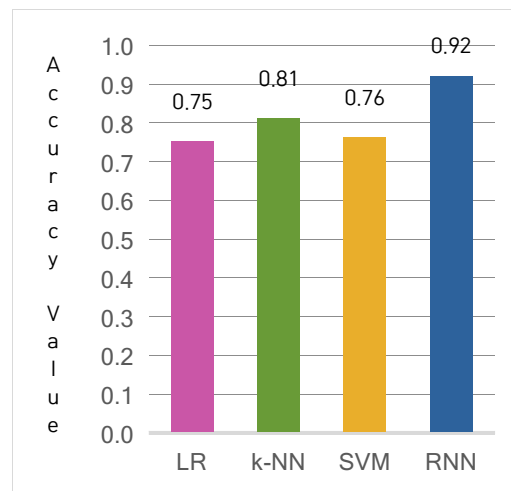


Fig. 3. Prediction accuracy of each machine learning

모델 간의 교차 검증한 AUC는 Fig 4.와 같다. 모델 간의 AUC는 LR, k-NN, SVM, RNN이 각각 0.86, 0.91, 0.87, 0.92로 나타났다. 이와 같이 모델의 AUC는 RNN 방법을 사용한 모델이 다른 모델들에 비해 모든 분류의 임계값에서 가장 정확한 결과를 보였다. 따라서 모델의 accuracy와 AUC 결과를 살펴보면, RNN 모델이 다른 기계학습 모델에 비해 정확도 측면에서 모두 우수한 것으로 나타났다.

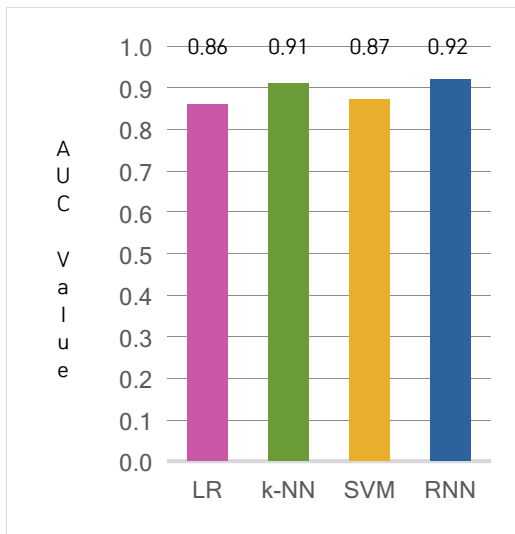


Fig. 4. Prediction AUC of each machine learning

4. 논의

본 연구의 한계점은 KoGES에서 조사된 자료원에는 당뇨 발생 환자의 데이터의 양이 적은 것이다. 그러나 적은 양의 데이터임에도 불구하고 다른 기계학습보다 높은 정확도를 보였다는 것은 질병을 예측하는 데 있어 기존의 기계학습 방법보다 RNN 방법이 미래를 예측하는 데 더 적합하다는 것을 의미한다.

기계학습을 이용한 당뇨병에 관한 연구는 대부분 빠른 진단을 목적으로 했다. Iyer et al.은 의사결정나무(J48 알고리즘)와 Naive Bayes을 이용하여 당뇨병을 예측하는 작업을 했고[18], Kumari와 Chitra은 SVM(Support Vector Machine)을 이용하여 T2DM을 예측할 수 있다고 제안하는 연구가 있었다[19]. Sarwar와 Sharma는 Naive Bayes에 대한 연구를 통해 T2DM을 진단 예측할 수 있다고 제안했다[20]. 또한 당뇨병 진단 예측 연구의

경우 시계열 데이터를 이용하진 않았지만 신경망 네트워크 모델이 적용되어 왔다[21-23]. 또한, Motka et al. [24]과 Polat et al. [25]은 인공 신경 퍼지 추론 시스템(ANFIS : Adaptive Neuro Fuzzy Inference System)을 사용했다. 그러나 기존의 기계학습을 이용한 당뇨병 예측모델 연구들은 질병의 발생을 예측하는 것이 아닌 정상인과 환자를 분류(classification)하는데 초점을 맞춘 연구이다. 또한 시간 정보가 없는 횡단적 데이터 세트를 사용했다는 점이 한계로 꼽힌다.

지금까지 유일하게 시계열 데이터를 바탕으로 당뇨병 발생 예측 모델을 만든 연구는 Alhassan, Zakhriya, et al. [26]의 연구가 유일했다. Alhassan, Zakhriya, et al.의 연구에서는 불규칙적인 자료원을 사용하였기 때문에 불규칙 잡음(random noise)이 있어, 학습에 나쁜 영향을 미쳤다. 따라서 데이터의 불균형을 없앤 후 학습을 진행할 필요가 있다.

또한 현재 기계 학습을 이용하여 국내 자료원을 사용한 당뇨병 발생 예측모델 연구는 KoGES 지역사회 기반 코호트(안산, 안성)를 이용한 Lee, Juyoung, et al.[8]의 연구가 유일하고, 시계열 데이터를 활용할 수 있는 RNN을 이용한 연구는 현재까지 없었다.

본 연구는 지도 학습(Supervised Learning)의 형태를 취하기 때문에, 예측하고자 하는 결과를 바탕으로 모델이 만들어졌다. 따라서 제안한 모델은 동일한 데이터 형태를 대상으로 했을 경우에만 예측이 가능하다는 점과 당뇨병만을 예측하는 모델이라는 한계점이 있다.

5. 결론

본 연구에서는 T2DM의 발생을 예측하기 위해 RNN을 이용한 질병발생 조기 예측 모델을 개발하였다. 모델은 KoGES 데이터 세트를 사용하여 다양한 입력 크기로 학습하고 테스트하였다. 당뇨병 발생 예측 결과는 동일한 데이터 세트를 사용하는 다른 기계학습(LR, k-NN 및 SVM)과 정확도를 비교한 결과 RNN 모델이 다른 기계학습보다 더 높은 성능을 보였다.

앞으로는 T2DM 예측 모델로 T2DM의 발생을 예측함으로써 환자에게 질병발생 전 경각심을 느끼게 해줄 수 있을 뿐만 아니라 당노가 발생하기 전에 생활습관 개선 및 혈당조절을 통해 당뇨병 발병을 예방하고 늦출 수 있을 것이다.

REFERENCES

- [1] Z. Punthakee, R. Goldenberg & P. Katz. (2018). Definition, classification and diagnosis of diabetes, prediabetes and metabolic syndrome. *Canadian journal of diabetes*, 42, S10–S15.
- [2] D. Atlas. (2015). International diabetes federation. IDF Diabetes Atlas, 7th edn. Brussels, Belgium: *International Diabetes Federation*.
- [3] N. Cho, J. E. Shaw, S. Karuranga, Y. Huang, J. D. da Rocha Fernandes, A. W. Ohlrogge & B. Malanda. (2018). IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. *Diabetes research and clinical practice*, 138, 271–281.
- [4] Korea Centers for Disease Control. (2015). *Korean national health and nutrition survey data*. <https://knkanes.cds.go.kr/>
- [5] I. Baik. (2019). Projection of Diabetes Prevalence in Korean Adults for the Year 2030 Using Risk Factors Identified from National Data. *Diabetes & metabolism journal*, 43(1), 90–96.
- [6] American Diabetes Association. (2004). Screening for type 2 diabetes. *Diabetes care*, 27(suppl 1), s11–s14.
- [7] J. L. Gross, M. J. De Azevedo, S. P. Silveiro, L. H. Canani, M. L. Caramori & T. Zelmanovitz. (2005). Diabetic nephropathy: diagnosis, prevention, and treatment. *Diabetes care*, 28(1), 164–176.
- [8] J. Y. Lee et al. (2011). Development of a predictive model for type 2 diabetes mellitus using genetic and clinical data. *Osong public health and research perspectives*, 2(2), 75–82.
- [9] A. M. Kanaya et al. (2005). Predicting the development of diabetes in older adults: the derivation and validation of a prediction rule. *Diabetes Care*, 28(2), 404–408.
- [10] J. Choi & Y. Suh. (2018). Deriving rules for identifying diabetic among individuals with metabolic syndrome. *Journal of digital convergence*, 16(11), 363–372.
- [11] Y. M. Kim & S. H. Kang. (2015). Changes and determinants affecting on geographic variations in health behavior, prevalence of hypertension and diabetes in Korean. *Journal of digital convergence*, 13(11), 241–254.
- [12] H. Y. Kim & H. S. Kim. (2018). Factors Affecting the Control of HbA1c in Type 2 Diabetic Patients. *Journal of Convergence for Information Technology*, 8(6), 75–84.
- [13] M. J. Lee, H. K. Kang & B. J. Seo. (2018). Correlation between Outpatient's Medical Adherence and National Insurance Types in the Type 2 Diabetes Mellitus. *Journal of Convergence for Information Technology*, 8(4), 9–14.
- [14] Z. C. Lipton, D. C. Kale, C. Elkan & R. Wetzel. (2015). Learning to diagnose with LSTM recurrent neural networks. *arXiv preprint arXiv:1511.03677*.
- [15] E. Choi, M. T. Bahadori, A. Schuetz, W. F. Stewart & J. Sun. (2016, December). Doctor ai: Predicting clinical events via recurrent neural networks. In *Machine Learning for Healthcare Conference* (pp. 301–318).
- [16] Y. S. Lee & S. S. Moon. (2011). The use of HbA1c for diagnosis of type 2 diabetes in Korea. *The Korean Journal of Medicine*, 80(3), 291–297.
- [17] S. Hochreiter & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735–1780.
- [18] A. Iyer, S. Jeyalatha & R. Sumbaly. (2015) Diagnosis of Diabetes Using Classification Mining Techniques. *International Journal of Data Mining & Knowledge Management Process (IJDKP)*, 5, 1–14. <https://doi.org/10.5121/ijdkp.2015.5101>
- [19] V. A. Kumari & R. Chitra. (2013). Classification of diabetes disease using support vector machine. *International Journal of Engineering Research and Applications*, 3(2), 1797–1801.
- [20] A. Sarwar & V. Sharma. (2012). Intelligent Naïve Bayes approach to diagnose diabetes Type-2. *International Journal of Computer Applications, IJCA Special Edition Nov*, 14–16.
- [21] P. Venkatesan & S. Anitha. (2006). Application of a radial basis function neural network for diagnosis of diabetes mellitus. *Current Science (00113891)*, 91(9), 1195–1199.
- [22] X. H. Meng, Y. X. Huang, D. P. Rao, Q. Zhang & Q. Liu. (2013). Comparison of three data mining models for predicting diabetes or prediabetes by risk factors. *The Kaohsiung journal of medical sciences*, 29(2), 93–99.
- [23] H. Temurtas, N. Yumusak & F. Temurtas. (2009). A comparative study on diabetes disease diagnosis using neural networks. *Expert Systems with applications*, 36(4), 8610–8615.
- [24] R. Motka, V. Parmarl, B. Kumar & A. R. Verma. (2013, September). Diabetes mellitus forecast using different data mining techniques. In *2013 4th International Conference on Computer and Communication Technology (ICCT)* (pp. 99–103). IEEE.
- [25] K. Polat & S. Güneş. (2007). An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease. *Digital Signal Processing*, 17(4), 702–710.
- [26] Z. Alhassan, A. S. McGough, R. Alshammari, T. Daghtani, D. Budgen & N. Al Moubayed. (2018). Type-2 diabetes mellitus diagnosis from time series clinical data using deep learning models. In *International Conference on Artificial Neural Networks*. p468–478.

장 진 수 (Jin-Su Jang)

[장학원]



- 2017년 2월 : 고려대학교 보건과학과 (보건학석사)
- 2017년 3월 ~ 현재 : 고려대학교 보건과학과 박사과정
- 관심분야 : Health Information Technology, Disease Early Prediction, Deep Learning

· E-Mail : runmc@korea.ac.kr

이 민 준 (Min-Jun Lee)

[장학원]



- 2019년 2월 : 연세대학교 의공학부 응용전자공학 전공(공학사)
- 2019년 2월 ~ 현재 : 고려대학교 일반대학원 보건과학과 석박통합과정
- 관심분야 : u-healthcare, Bio-medical Engineering, blockchain, Social Sentiment Analysis

· E-Mail : dlalswns444@korea.ac.kr

이 태 노 (Tae-Ro Lee)

[장학원]



- 1984년 2월 : 광운대학교 전산학과(이학사)
- 1989년 8월 : 경희대학교 교육대학원 전산학 전공(교육학 석사)
- 2001년 2월 : 경희대학교 대학원 컴퓨터공학과 (공학 박사)
- 2005년 9월 ~ 2006년 8월 : Visiting

Professor of Griffith University

- 1996년 9월 ~ 현재 : 고려대학교 보건과학대학 보건정책관리학부 교수
- 관심분야 : Hospital Information System, u-Healthcare, MIS, Signal Processing
- E-Mail : trlee@korea.ac.kr