

OPEN

# Artificial intelligence predicts the progression of diabetic kidney disease using big data machine learning

Masaki Makino<sup>1</sup>, Ryo Yoshimoto<sup>1</sup>, Masaki Ono<sup>2</sup>, Toshinari Itoko<sup>2</sup>, Takayuki Katsuki<sup>2</sup>, Akira Koseki<sup>2</sup>, Michiharu Kudo<sup>2</sup>, Kyoichi Haida<sup>3</sup>, Jun Kuroda<sup>4</sup>, Ryosuke Yanagiya<sup>5</sup>, Eiichi Saitoh<sup>6</sup>, Kiyotaka Hoshinaga<sup>7</sup>, Yukio Yuzawa<sup>8</sup> & Atsushi Suzuki<sup>1</sup> 

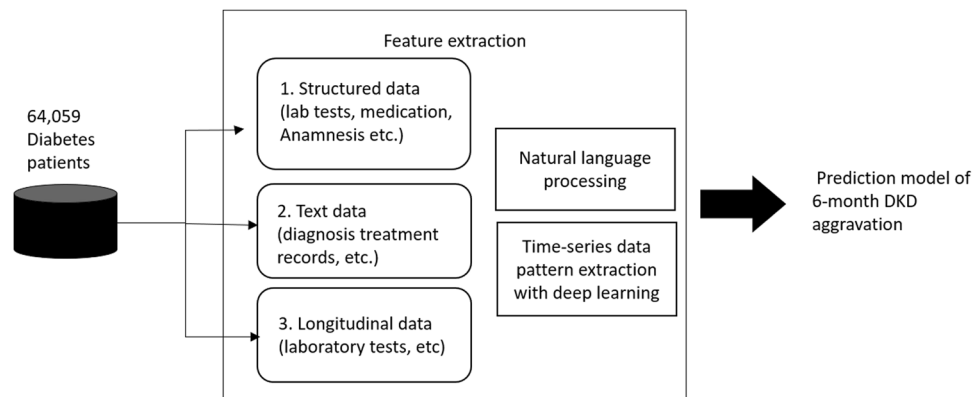
Artificial intelligence (AI) is expected to support clinical judgement in medicine. We constructed a new predictive model for diabetic kidney diseases (DKD) using AI, processing natural language and longitudinal data with big data machine learning, based on the electronic medical records (EMR) of 64,059 diabetes patients. AI extracted raw features from the previous 6 months as the reference period and selected 24 factors to find time series patterns relating to 6-month DKD aggravation, using a convolutional autoencoder. AI constructed the predictive model with 3,073 features, including time series data using logistic regression analysis. AI could predict DKD aggravation with 71% accuracy. Furthermore, the group with DKD aggravation had a significantly higher incidence of hemodialysis than the non-aggravation group, over 10 years ( $N = 2,900$ ). The new predictive model by AI could detect progression of DKD and may contribute to more effective and accurate intervention to reduce hemodialysis.

Today, type 2 diabetes mellitus (T2DM) is a worldwide burden afflicting developed and developing countries<sup>1</sup>. Chronic hyperglycemia and the subsequent accumulation of advanced glycation end-products result in multiple complications, including micro- and macrovascular diseases<sup>2</sup>. Among them, diabetic kidney disease (DKD), such as diabetic nephropathy, is the most frequent cause of hemodialysis (HD) and is associated with cardiovascular diseases<sup>3</sup>. Several clinical risk factors, such as hyperglycemia, dyslipidemia, hypertension and smoking, are related to the progression of DKD<sup>4</sup>. Microalbuminuria is known to be a good predictor of further progression of diabetic nephropathy and subsequent cardiovascular diseases<sup>5</sup> and early intervention for DKD, such as anti-hypertensive medicine, could induce remission of DKD with microalbuminuria<sup>6–9</sup>. However, a more precise predictive model is needed for the very early intervention in DKD to prevent its further progression in diabetes patients without apparent symptoms or signs.

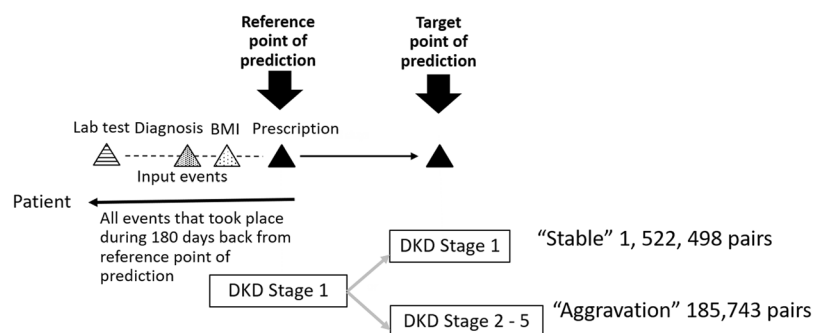
Artificial intelligence (AI) is changing our modern life and, in medicine, AI has two main branches, virtual and physical<sup>10</sup>. The physical branch includes robotics, which can assist surgery and rehabilitation. The virtual branch includes informatics, which is expected to assist physicians in their clinical diagnosis and treatment decisions. The recent progress of machine-learning, with big data analysis, is contributing greatly, especially in the field of clinical imaging<sup>11,12</sup>, pharmacokinetics<sup>13</sup>, genetics<sup>14</sup> and oncology<sup>15</sup>. However, there is so far little information about predictive models of prognosis and/or progression of complications in life-style related diseases, such as T2DM<sup>16–19</sup>.

In general, clinical studies are designed to elucidate specific clinical risk factors by arranging background data or conditions before recruitment. On the other hand, we performed clinical medicine under non-arranged

<sup>1</sup>Department of Endocrinology and Metabolism, Fujita Health University, Toyoake, Aichi, Japan. <sup>2</sup>IBM Research, Tokyo, Japan. <sup>3</sup>Business Process Planning Department, The Dai-ichi Life Insurance Company, Limited, Tokyo, Japan. <sup>4</sup>IT Business Process Planning Department, The Dai-ichi Life Insurance Company, Limited, Tokyo, Japan. <sup>5</sup>Division of Medical Information Systems, Fujita Health University, Toyoake, Aichi, Japan. <sup>6</sup>Department of Rehabilitation Medicine, Fujita Health University, Toyoake, Aichi, Japan. <sup>7</sup>Department of Urology, Fujita Health University, Toyoake, Aichi, Japan. <sup>8</sup>Department of Nephrology, Fujita Health University, Toyoake, Aichi, Japan. Correspondence and requests for materials should be addressed to A.S. (email: [aslapin@fujita-hu.ac.jp](mailto:aslapin@fujita-hu.ac.jp))



**Figure 1.** Feature extraction for deep learning. Clinical features for the predictive model of 6-months aggravation of diabetic kidney disease (DKD) were extracted using three different approaches: structural data, text data and longitudinal data from the electronic medical records (EMR) of 64,059 type 2 diabetes patients.



**Figure 2.** Extraction of “Stable” and “Aggravation” groups of diabetic kidney disease (DKD) for 6 months. We constructed 180 days-long event pairs between the reference point and target point of prediction in stage 1 DKD patients and obtained 1,708,241 pairs, including 1,522,498 in the stable group and 185,743 in the aggravation group. Then, we selected 15,422 in the stable group and 15,388 in the aggravation group by under-sampling.

conditions. Therefore, population-based analysis is used for the assessment, considered as the so-called, real-world setting. However, the analyses have some disadvantages, such as many confounding factors which may cause several biases affecting the main conclusion. We hypothesized here that AI could provide more useful analysis by big-data-based machine learning without preconception.

In this study, we constructed a new predictive model of DKD in diabetes patients by big data machine learning, based on electronic medical records (EMR).

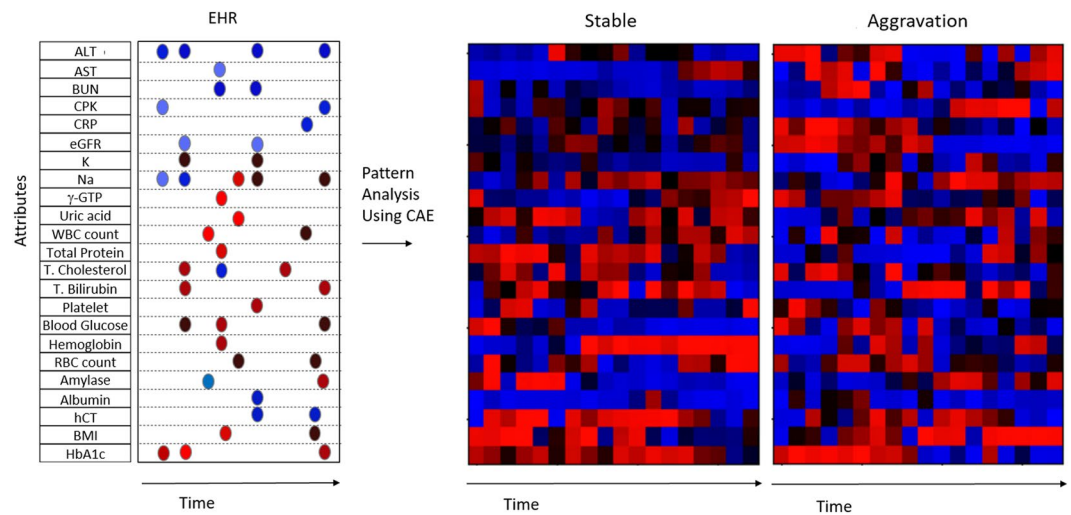
## Results

From 858,660 EMR, we extracted 451,584 cases with relevant clinical data. According to our criteria, 64,059 patients could be defined as T2DM. From these patients, we extracted the clinical features using three different approaches: structural data, text data and longitudinal data from EMR (Fig. 1).

During this process, AI extracted structural features such as laboratory tests, diagnosis, prescription and ICD 10 codes. AI picked up the past history, current diseases and prescriptions from the EMR text by natural language processing. Then, we constructed 180 days-long event pairs between the reference point and target point of prediction in stage 1 DKD patients and obtained 1,708,241 pairs, including 1,522,498 in the stable group and 185,743 in the aggravation group (Fig. 2). Finally, we selected 15,422 in the stable group and 15,388 in the aggravation group by under-sampling.

At first, we examined how much the information that longitudinal data of EMR records have affected the DKD prediction. To this end, for those “Stable” and “Aggravation” groups, AI extracted raw features during the 6 months prior to the reference point of prediction for selected 24 factors to reveal typical time series patterns relating to 6-month DKD aggravation, using a convolutional autoencoder<sup>20,21</sup> (Fig. 3). The 24 factors whose longitudinal information would affect the DKD were selected before the analysis. Figure 3 shows extracted typical time series patterns for “Aggravation” and “Stable” groups on the right, including some intriguing time series patterns that creatine phosphokinase (CPK) and body mass index (BMI) have a conspicuous increasing patterns for the aggravation group. The results of above autoencoder experiment showed the importance of taking account of longitudinal data.

Second, AI also constructed the predictive model with 3,073 features, including longitudinal data using logistic regression analysis (Table 1). We used longitudinal explanation variables by summarizing past 180-day EMR



**Figure 3.** Time-series data pattern extraction with deep learning. Artificial intelligence extracted raw features during the 6 months prior to the reference point of prediction for selected 24 factors to reveal time series patterns relating to 6-month DKD aggravation, using a convolutional autoencoder (CAE) and inverse analysis. Red and blue mean high and low values, respectively, and brightness of the color mean the magnitude of the values.

Category	Number of characteristics
<b>Structural data</b>	
Laboratory tests	168 (24 × 7)
Serum: Albumin, ALT, Amylase, AST, BG, BUN, CPK, CRP, eGFR, Creatinine, $\gamma$ -GTP, HbA1c, Hemoglobin, Hematocrit, K, Na, Platelet, RBC, WBC, Total bilirubin, Total cholesterol, Total protein, Uric acid	1. Mean of all data
	2. The difference of the highest value and the lowest value
	3. S.D. of all data
	4. The difference of the last and first monthly means
	5. Mean of monthly mean data
Urine: Albuminuria, Protein	6. S.D. of monthly mean data
Profile	6
Prescription (YJ code)	408
ICD 10 (top 3 numbers)	1,265
<b>Text data from electronic medical records</b>	
Past history of diseases	613 (names of diseases)
Current disease	613 (names of diseases)

**Table 1.** Extraction of time-series data and text data by natural language processing. ALT: alanine aminotransferase, AST: aspartate aminotransferase, BG: blood glucose, BUN: blood urea nitrogen, CPK: creatinine phosphokinase, CRP: C-reactive protein, eGFR: estimated glomerular filtration rate,  $\gamma$ -GTP:  $\gamma$ -glutamyl transpeptidase, K: potassium, Na: sodium, RBC: red blood cell, WBC: white blood cell.

records using average, standard deviation, and so on, taking the previous experiments into account. We then performed 5-fold cross validation and obtained a predictive evaluation result for each fold. The resultant average of the AUC was 0.743 and the average accuracy was 71%. Interestingly, 180-day statistical scores of laboratory tests before each reference time point seem to have good influence on prediction of DKD stage defined by urinary protein in 180 days. Therefore, the aggravation of urinary protein observation is strongly affected by its variance over the past 180 days. As actually shown in Table 2, as feature categories are added to the model, prediction performance improved. We observed that conspicuous improvement was shown when longitudinal features were added. Table 3 showed the resultant confusion matrix of our prediction.

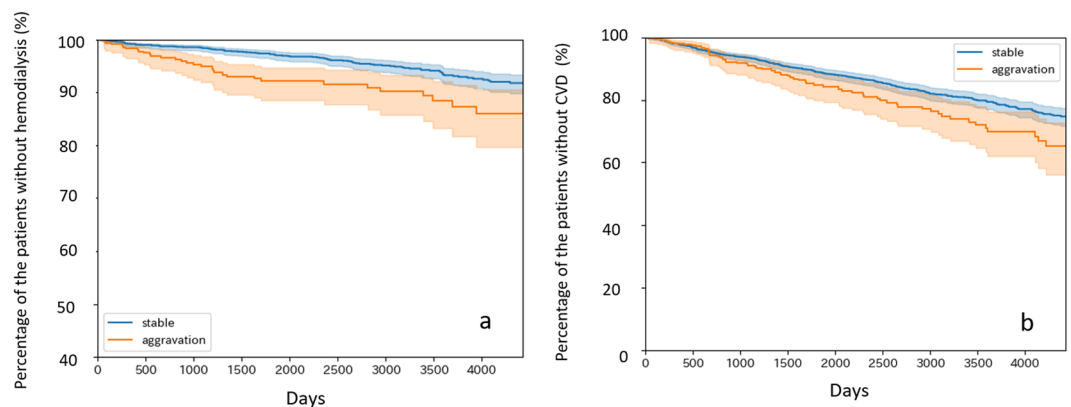
Third and last, we examined long-term relations with this 180-day prediction. When using the same “Stable” and “Aggregation” label for the patients, the DKD aggravation group had a significantly higher incidence of HD than the stable group over 10 years (Fig. 4a). Cardiovascular events were also more frequent in the DKD aggravation group than in the stable group (Fig. 4b).

Features	AUC	Accuracy
Profile	0.562	0.548
Profile + ICD10	0.562	0.557
Profile + ICD10 + YJCode	0.613	0.594
Profile + ICD10 + Blood Tests (latest)	0.644	0.606
Profile + ICD10 + YJCode + Blood Tests (latest and longitudinal)	0.656	0.610
Profile + ICD10 + YJCode + Blood Tests (latest and longitudinal) + Urinary Tests (latest and longitudinal)	0.729	0.691
Profile + ICD10 + YJCode + Blood Tests (latest and longitudinal) + Urinary Tests (latest and longitudinal) + Current Disease + Disease History	0.743	0.701

**Table 2.** Accuracy of prediction in each model. AUC: area under the curve; ICD10: 10<sup>th</sup> revision of the International Statistical Classification of Diseases and Related Health Problems, YJ code: national health insurance drug list code.

	Predicted: Stable	Predicted: Aggravation
Actual: Stable	12,822	2,600
Actual: Aggravation	6,261	9,127

**Table 3.** Confusion Matrix of Prediction Results.



**Figure 4.** Kaplan-Meier survival analysis for hemodialysis (a) and cardiovascular disease (CVD) (b) after the first visit in the stable and aggravation groups. (a) The blue line shows the percentage of patients without hemodialysis in the “Stable” group ( $n = 2,477$ ) at each time point, while the red line shows that of the “Aggravation” group ( $n = 423$ ). Log rank test result marked  $P = 0.00024$ . (b) The blue line shows the percentage of patients without a cardiovascular event in the “Stable” group ( $n = 2,367$ ) at each time point, while the red line shows that of the “Aggravation” group ( $n = 407$ ). Log rank test result marked  $P = 0.01434$ . In this study, cardiovascular events are defined as hospitalized heart failure, myocardial infarction, performance of coronary artery bypass grafting, percutaneous coronary intervention, and death due to heart disease. The shaded areas show 95% confidence interval.

## Discussion

In this study, we showed that AI could predict the progression of DKD using big data machine learning, according to the EMR of T2DM patients. Our study used three novel approaches to improve the predictive capacity of disease-specific complications. First, we constructed a new predictive model of diabetic complications before the patients showed clinical signs or symptoms such as microalbuminuria. Second, we used big EMR data for machine-learning by AI without any objective of clinical research; we included cases not defined clinically as T2DM in their text on EMR. Third, AI used time-series data from 6 months before the reference periods and predicted the progression of DKD for 6 months after the reference periods.

DKD is one of the most common diabetic complications and its progression results in hemodialysis for end-stage renal disease (ESRD)<sup>3,22</sup>. DKD is the major cause of hemodialysis in many countries<sup>3</sup>. Diabetes patients with normoalbuminuria have been reported to progress to microalbuminuria at 2.8%/year<sup>23,24</sup>. Because microalbuminuria is considered to be an early marker predicting diabetic nephropathy and subsequent ESRD, remission of microalbuminuria should mean less ESRD in the future<sup>25,26</sup>. There are several reports concerning

the remission of early stage DKD, such as diabetic nephropathy at stage 2, where patients have 30–300 mg urinary microalbumin/day<sup>6–9</sup>. Theoretically, the earlier the intervention for DKD, the better the outcome we can expect in terms of remission. However, early intervention in stage 1 DKD must be less cost-effective and has the risk of overdiagnosis and/or overtreatment. Other biomarkers in diabetes patients with normoalbuminuria, such as urinary L-type fatty acid binding protein and serum tumor necrosis factor- $\alpha$  and its receptors, could also be surrogate markers of diabetic nephropathy<sup>27</sup>, but none of these markers is perfect. Liao *et al.* recently reported that urinary proteomics analysis could be useful to detect early diabetic nephropathy and that the haptoglobin-to-creatinine ratio might provide a better predictive value for early renal functional decline in 4.2 years than the microalbumin-to-creatinine ratio<sup>28</sup>. In this study, we constructed a model of early stage DKD at stage 1 to 2 diabetic nephropathy. With this approach, we could define T2DM at an early stage of DKD but with a higher future risk of its progress. It may be beneficial to provide more intensive care, such as statins and anti-hypertensive medicine, for these patients.

Proteinuria is related to atherosclerosis, resulting in cardiovascular diseases, such as ischemic heart disease and apoplexy<sup>29</sup>. For diabetes patients, we used microalbuminuria as a surrogate marker to predict ESRD and other atherogenic cardiovascular events<sup>30</sup>. We showed here that progression of DKD in 6 months could result in a higher incidence of hemodialysis due to chronic renal failure in our patients over 10 years. Furthermore, the unstable proteinuria group had a higher incidence of cardiovascular events than the stable non-proteinuria group. These results suggest that very early intervention to reduce proteinuria could contribute to a better prognosis for both renal and cardiac diseases. Many countries have progressed to super-aging societies and elderly patients are liable to have several diseases at the same time. Therefore, clinical medicine in super-aging societies is more complicated and clinical trials to find effective treatments are more difficult. In previous works to predict diabetes complications, they calculated the risks with several clinical information such as current age, sex, ethnicity, smoking status, presence or absence of microalbuminuria or worse, and laboratory data for diabetes, hypertension and dyslipidemia<sup>31</sup>. With this approach, we could pick up known risk factors according to the previous report but could not include unknown risk factors. In addition, when we perform the clinical trials to prove the efficacy of the treatment, we need to register many untreated patients as control. In our approach to find an AI-supported predictive model for chronic diseases such as T2DM, we can elucidate the combination of clinical risk factors with less expense and less effort than current clinical trials. AI has the advantages of improving clinical medicine in the field with digital data such as imaging<sup>11,12</sup>, pharmacokinetics<sup>13</sup>, genetics<sup>14</sup> and oncology<sup>15</sup>. Recent studies with AI in the field of diabetes represent a diverse and complex set of innovative approaches that aim to transform diabetes care in four main areas: automated retinal screening, clinical decision support, predictive population risk stratification, and patient self-management tools<sup>32</sup>. AI could improve imaging techniques such as diabetic retinopathy screening<sup>33</sup>, because digital imaging is an aggregation of many pixels with the same processing condition. Recently, the US Food and Drug Administration (FDA) permitted the marketing of the first medical device to use AI to detect diabetic retinopathy. There are several innovative studies using a machine-learning approach to develop phenotyping frameworks to detect diabetes<sup>16</sup>, the progression of diabetes<sup>18</sup> and hypoglycemia<sup>17</sup>. However, an AI-oriented predictive model of diabetic complications has not yet been developed. Our current study suggests that AI could support our decision to reduce future clinical events at an early stage of complications in chronic diseases such as T2DM.

There are some limitations to this study. First, the information obtained from each EMR, especially from the medical doctors' records, varies considerably and we could not unify the data extraction from each patient. Second, the duration between each laboratory test was not uniform and depended on the individual patient. Third, this study was carried out in a single center and has not been reevaluated using EMR from other institutions. Fourth, we could not find any relationship between progression of DKD for 6 months and medication. Because the patients without DKD are likely to be treated less intensively, we suggest that the medication itself may not have affected the progression of DKD in this study. Therefore, we still need prospective study to prove very early intervention to DKD could prevent macrovascular events including ESRD and CVD in T2DM patients.

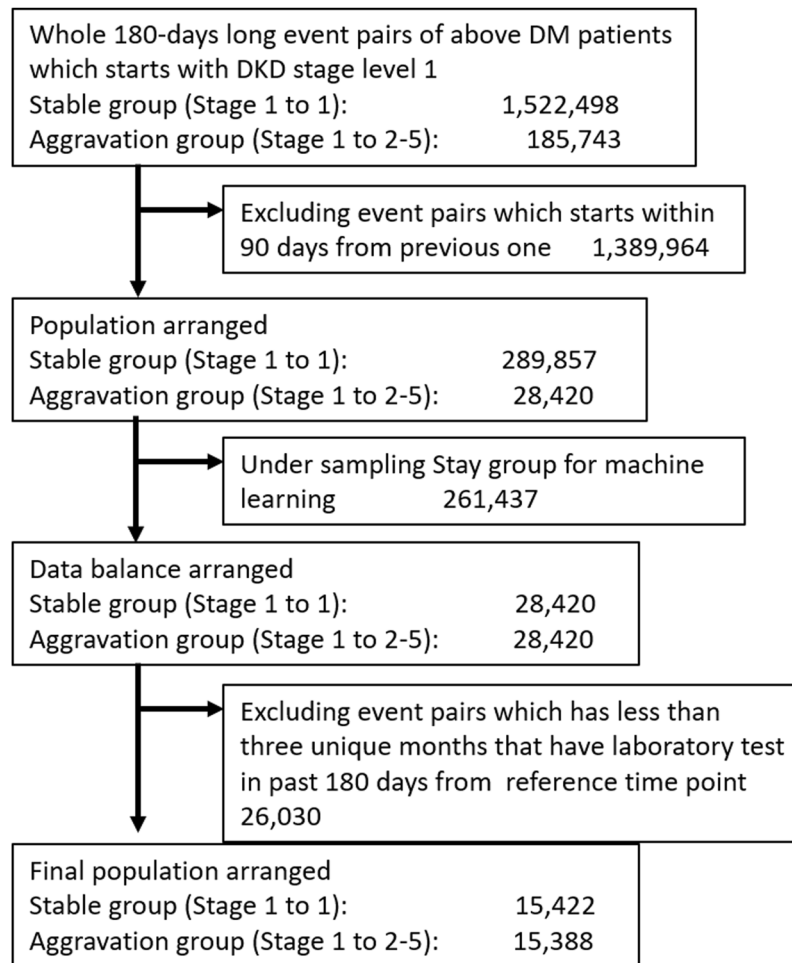
In conclusion, the new predictive model using AI could detect the progression of DKD, which may contribute to more effective and accurate intervention to reduce hemodialysis and cardiovascular events.

## Methods

**Definition of T2DM and the EMR used for this study.** We started to use the EMR system in our hospital in 2005 and had 858,660 EMR in 2016. There are 407,076 EMR without any clinical data, so we could extract 451,584 EMR in total. Among them, we found 64,059 patients with a diagnosis of T2DM according to our criteria as follows: (1) T2DM was recorded in the medical billing, (2) the HbA1c level was equal to or above 6.5% (NGSP), (3) the fasting plasma glucose level was equal to or above 126 mg/dL, except for in an emergency room, (4) the postprandial plasma glucose level was equal to or above 200 mg/dL, except for in an emergency room, (5) anti-diabetic medicine (137 drugs) was prescribed. This study was approved by the Fujita Health University Ethical Committee (HM17-159). Informed consent from each patient was not available; therefore, the opportunity to opt out of this research was announced on our homepage ([http://www.fujita-hu.ac.jp/~endylabo/research/watson\\_01/index.html](http://www.fujita-hu.ac.jp/~endylabo/research/watson_01/index.html)). All private personal information was protected and removed during the process of analysis and publication. All data associated with this study are present in the paper or in the Supplementary Materials.

**Staging categorization of DKD in this study.** Diabetic nephropathy is defined by albuminuria and a decrease of the estimated glomerular filtration rate (eGFR). In typical diabetic nephropathy, the microalbuminuria precedes the decrease of the eGFR because hyperglycemia induces glomerular damage resulting in a dysfunction of the barrier system in glomerulus. However, patients with low eGFR without apparent proteinuria have been increasing in number recently because of the treatment and the aging population. Therefore, we have recently started to use DKD instead of diabetic nephropathy<sup>3</sup>. In this study, we defined the DKD stage by





**Figure 5.** Study population. After identifying labeled pairs, we manually under-sampled the pairs to balance the numbers. In this test, the stable group was much larger than the aggravation group, so pairs in the stable group were under-sampled to match the numbers better. We used these labeled pairs for supervised learning; the final numbers are 15,422 for the stable group and 15,388 for the aggravation group. The characteristics of the study population are shown in Table S2.

proteinuria, the albumin to creatinine ratio or the eGFR at stages 1 to 4 (Table S1). Stage 5 is defined as maintenance hemodialysis or continuous ambulatory peritoneal hemodialysis.

**Predictive model of DKD.** Focusing on DKD, one of the diabetic complications that will incur heavy treatment, including dialysis, if aggravated, a model was developed to predict its progression. Based on laboratory test results, this model uses the stage of DKD (one to five, Table S1) as labels and predicts whether the diabetes of DKD stage 1 patients will progress, in terms of DKD stage, after 180 days. A predictive model was created using a variety of features, including patient profiles, name of the disease and treatment (details of treatment and medication), extracted from EMR. These records are in the clinical data base system of Fujita Health University Hospital.

**Label definition.** Labels for prediction are created from the DKD stage change in 180 days. We first find any pair of events of the “reference point” and “target point”, which are measured around 180 days apart. Closely measured pairs were filtered to avoid using similar data. Then, we labeled the pairs as the stable and aggravation groups. The stable group ( $n = 1,522,498$ ) were in stage 1 at the reference and target points (Fig. 2). The aggravation group ( $n = 185,743$ ) are those whose DKD stage at the reference point was 1 but who progressed to stage 2 or more at the target point. We excluded 1,389,964 event pairs whose reference and target points were less than 90 days apart. After this exclusion, there were 289,857 patients in the stable group and 28,420 in the aggravation group. In order to adjust the numbers of each group for machine learning, we performed under-sampling of the stable group. Finally, we excluded 26,030 event pairs which had less than three unique months that had had laboratory tests in past 180 days from reference point. The final population for this study was selected as 15,422 in the stable group and 15,388 in the aggravation group (Fig. 5).

**Structured features.** We extracted a variety of features from the EMR. They include laboratory tests, profiles, medication, disease history and so on. We used the values which were collected at most 180 days prior to

$$Y = \frac{1}{1 + e^{-\sum_{i=0}^n \alpha_i x_i}}$$

**Figure 6.** Formula to predict the probability of the DKD stage a half-year later. The probability of the DKD stage a half-year later as  $Y$ , where  $x_i$  are the input variables (the number is  $n$ ) and  $\alpha_i$  are parameters, which were calculated using logistic regression.

the reference point. Table 1 shows the categories of structured features and their numbers. Among these feature values, we created processed features by statistical aggregation of feature sequences, because we have a time-series of measured values for each feature. We calculated 180-day S.D., mean, and others described in Table 1 and then used these as the features.

**Unstructured features and text processing.** Other than structured features, we processed texts from medical examinations and nutrition consultations. The texts are recorded in the free text format for each consultation. From those we extracted disease names of current and historical diseases as keywords by traditional natural language processing. We extracted those keywords using disease name dictionaries for name aggregation. Other than keywords extraction, we also used topic information as features. As using the medical text records as a corpus, we conducted topic analysis using Latent Dirichlet Allocation (LDA)<sup>34</sup>, which are commonly used for topic extraction. Such extracted topics are also used for our features.

**Time series pattern analysis.** To selected 24 factors whose longitudinal information would affect the DKD, we conducted an analysis to find time series patterns relating to 6-month DKD aggravation using a convolutional autoencoder<sup>20,21</sup>. In the convolutional autoencoder for the input vector  $X$  of time series data of 24 factors, the encoder has 5 hidden layers consisting of: 1) one-dimensional convolutional layer with  $64 \times 24 \times 3$  filters for  $X$ ; 2) a max-pooling layer with  $1 \times 2$  filter; 3) a one-dimensional convolutional layer with  $64 \times 1 \times 3$  filters per map; 4) a max-pooling layer of  $1 \times 2$  filter; 5) a fully connected layer which extracts a hidden representation as 128 neurons as function  $f(X)$ , while the decoder is the transposed convolution<sup>35</sup>, which constructs a reverse function  $g(\cdot)$  where  $X = g(f(X))$ . We then minimize the reconstruction error of  $X - g(f(X))$  to learn the autoencoder. Next we learn the classification model to predict “Aggravation” or “Stable” using the extracted hidden representation. From those results, we generated the typical input time series patterns by inverse analysis, which finds maximum input time series values so that the hidden representation corresponding to “Aggravation” or “Stable” is activated. Details of the network architecture and mathematical definitions are described in our previous paper<sup>21</sup>.

**Prediction model.** We applied logistic regression using the Python code with scikit-learn library (<https://scikit-learn.org/>) for model solving. Among many machine learning packages including R (<https://www.r-project.org/>), SPSS (<https://www.ibm.com/analytics/spss-statistics-software>), Matlab (<https://www.mathworks.com/products/matlab.html>), SAS (<https://www.sas.com/home.html>), Weka (<https://www.cs.waikato.ac.nz/ml/weka/>) and other, we chose scikit-learn because our feature extraction processes are written in Python. Due to the large number of explanation variables, we used L2-regularization to avoid overfitting. With regularization, we also used a stepwise method to choose explanation variables. To speed up the computation, we adopted a concurrent algorithm to choose those variables. Using the feature set once selected, we used a 5-fold cross validation method to evaluate the prediction performance. Figure 6 shows the formula used to predict the probability of the DKD stage a half-year later.

**Long-term effect of DKD aggravation.** In predicting 180-day DKD stage aggravation, it is interesting to know whether such short-term stage changes are related to changes in long-term patient status; the period is sufficiently long to observe even severe events, such as dialysis and other critical end points. To elucidate this short-term to long-term relationship, we conducted survival analysis using Kaplan-Meier methods, using two groups of 180-day DKD stage changes, namely the stable and aggravation groups. We estimated two Kaplan-Meier curves for the diabetic hard end points of hemodialysis and cardiovascular disease events. The cardiovascular disease events include hospitalized heart failure, myocardial infarction, coronary artery bypass grafting, percutaneous coronary intervention, and death from heart disease.

Observing the 180-day DKD stage change and to analyze each patient once, we collected 180-day DKD stage change data from the first visits of the patients to the hospital. More precisely, we collected the first DKD stage using laboratory test results from the patient’s first visit to the hospital and the second DKD stage using laboratory test results, which were taken between 180 and 240 days after the first visit. We then separated the data for the stable and aggravation groups. The definition of stable and aggravation is as we described for the predictive model. Finally, we estimated Kaplan-Meier curves from 240 days after the first visit to the last recorded date from the patient’s EMR. Note that, for the aggravation group, we excluded patients who already suffered from the hard end-points. This setting provides a fair and conservative investigation of the short-term to long-term relationship.

For survival analysis for hemodialysis, the number of patients in the stable group was 2,477; they remain at DKD stage 1. The number of patients in the aggravation group was 423, all of whom survived and at the DKD stage of 2 to 5 about 180 days after the first visit. For survival analysis for cardiovascular disease, the respective numbers are 2,367 and 407.

**Statistical analysis.** To determine the relationship between half-year prediction and long-term tendency, we also conducted a survival analysis. We estimated Kaplan–Meier functions and the curves of occurrence of hemodialysis and cardiovascular diseases and carried out a Log-rank test. We prepared two groups of “Stable” and “Aggravation”, as described above, and applied Kaplan–Meier estimations using LIFELINES (<https://lifelines.readthedocs.io/en/latest/>), a python library of survival analyses.  $P < 0.05$  was considered as statistically significant.

## References

1. Cho, N. H. *et al.* IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. *Diabetes Res. Clin. Pract.* **138**, 271–281 (2018).
2. Sarwar, N. *et al.* Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: A collaborative meta-analysis of 102 prospective studies. *Lancet* **375**, 2215–2222 (2010).
3. Thomas, M. C. *et al.* Diabetic kidney disease. *Nat. Rev. Disease Prim.* **1**, 1–20 (2015).
4. Gæde, P. *et al.* Years of life gained by multifactorial intervention in patients with type 2 diabetes mellitus and microalbuminuria: 21 years follow-up on the Steno-2 randomised trial. *Diabetologia* **59**, 2298–2307 (2016).
5. Ninomiya, T. *et al.* Albuminuria and Kidney Function Independently Predict Cardiovascular and Renal Outcomes in Diabetes. *J. Am. Soc. Nephrol.* **20**, 1813–1821 (2009).
6. Nishimura, M. *et al.* Effect of Home Blood Pressure on Inducing Remission/Regression of Microalbuminuria in Patients With Type 2 Diabetes Mellitus. *Am. J. Hypertens.* **30**, 830–839 (2017).
7. Roscioni, S. S., Heerspink, H. J. L. & De Zeeuw, D. The effect of RAAS blockade on the progression of diabetic nephropathy. *Nat. Rev. Nephrol.* **10**, 77–87 (2014).
8. Kawanami, D. *et al.* SGLT2 inhibitors as a therapeutic option for diabetic nephropathy. *Int. J. Mol. Sci.* **18** (2017).
9. Penno, G., Garofolo, M. & Del Prato, S. Dipeptidyl peptidase-4 inhibition in chronic kidney disease and potential for protection against diabetes-related renal injury. *Nutr. Metab. Cardiovasc. Dis.* **26**, 361–373 (2016).
10. Hamet, P. & Tremblay, J. Artificial intelligence in medicine. *Metabolism* **69**, S36–S40 (2017).
11. Urban, G. *et al.* Deep Learning Localizes and Identifies Polyps in Real Time With 96% Accuracy in Screening Colonoscopy. *Gastroenterology* **155**, 1069–1078.e8 (2018).
12. Narula, S., Shameer, K., Salem Omar, A. M., Dudley, J. T. & Sengupta, P. P. Machine-Learning Algorithms to Automate Morphological and Functional Assessments in 2D Echocardiography. *J. Am. Coll. Cardiol.* **68**, 2287–2295 (2016).
13. Deshpande, D. *et al.* Levofloxacin Pharmacokinetics/Pharmacodynamics, Dosing, Susceptibility Breakpoints, and Artificial Intelligence in the Treatment of Multidrug-resistant Tuberculosis. *Clin. Infect. Dis.* **67**, S293–S302 (2018).
14. Bouazziz, J. *et al.* How Artificial Intelligence Can Improve Our Understanding of the Genes Associated with Endometriosis: Natural Language Processing of the PubMed Database. *Biomed Res. Int.* **2018**, 6217812 (2018).
15. Boon, I., Au Yong, T. & Boon, C. Assessing the Role of Artificial Intelligence (AI) in Clinical Oncology: Utility of Machine Learning in Radiotherapy Target Volume Delineation. *Medicines* **5**, 131 (2018).
16. Kagawa, R. *et al.* Development of Type 2 Diabetes Mellitus Phenotyping Framework Using Expert Knowledge and Machine Learning Approach. *J. Diabetes Sci. Technol.* **11**, 791–799 (2017).
17. Sudharsan, B., Peeples, M. & Shomali, M. Hypoglycemia prediction using machine learning models for patients with type 2 diabetes. *J. Diabetes Sci. Technol.* **9**, 86–90 (2015).
18. Anderson, J. P. *et al.* Reverse Engineering and Evaluation of Prediction Models for Progression to Type 2 Diabetes: An Application of Machine Learning Using Electronic Health Records. *J. Diabetes Sci. Technol.* **10**, 6–18 (2016).
19. Ye, C. *et al.* Prediction of Incident Hypertension Within the Next Year: Prospective Study Using Statewide Electronic Health Records and Machine Learning. *J. Med. Internet Res.* **20**, e22 (2018).
20. Nishio, M. *et al.* Convolutional auto-encoders for image denoising of ultra-low-dose CT. *Heliyon* **3**, e00393 (2017).
21. Katsuki, T. *et al.* Risk prediction of diabetic nephropathy via interpretable feature extraction from EHR using convolutional autoencoder. In *Studies in Health Technology and Informatics* **247**, 106–110 (2018).
22. Alicic, R. Z., Rooney, M. T. & Tuttle, K. R. Diabetic kidney disease: Challenges, progress, and possibilities. *Clin. J. Am. Soc. Nephrol.* **12**, 2032–2045 (2017).
23. Katayama, S. *et al.* Low transition rate from normo- and low microalbuminuria to proteinuria in Japanese type 2 diabetic individuals: The Japan diabetes complications study (JDACS). *Diabetologia* **54**, 1025–1031 (2011).
24. Hanai, K. *et al.* Asymmetric dimethylarginine is closely associated with the development and progression of nephropathy in patients with type 2 diabetes. *Nephrol. Dial. Transplant.* **24**, 1884–1888 (2009).
25. Araki, S. Ichi Comprehensive risk management of diabetic kidney disease in patients with type 2 diabetes mellitus. *Diabetol. Int.* **9**, 100–107 (2018).
26. Tu, S.-T. *et al.* Prevention of Diabetic Nephropathy by Tight Target Control in an Asian Population With Type 2 Diabetes Mellitus. *Arch. Intern. Med.* **170**, 155–161 (2010).
27. Gohda, T. *et al.* Clinical predictive biomarkers for normoalbuminuric diabetic kidney disease. *Diabetes Res. Clin. Pract.* **141**, 62–68 (2018).
28. Liao, W.-L. *et al.* Urinary Proteomics for the Early Diagnosis of Diabetic Nephropathy in Taiwanese Patients. *J. Clin. Med.* **7**, 483 (2018).
29. Brown, W. W. & Keane, W. F. Proteinuria and cardiovascular disease. *Am. J. Kidney Dis.* **38**, S8–S13 (2001).
30. Kidney Disease: Improving Global Outcomes (KDIGO) CKD-MBD Update. KDIGO 2017 Clinical Practice Guideline Update for the Diagnosis, Evaluation, Prevention, and Treatment of Chronic Kidney Disease–Mineral and Bone Disorder (CKD-MBD). *Kidney Int. Int. Suppl.* 1–59 (2017).
31. Palmer, A. J. Computer modeling of diabetes and its complications: A report on the fifth Mount Hood challenge meeting. *Value Heal.* **16**, 670–685 (2013).
32. Dankwa-Mullan, I. *et al.* Transforming Diabetes Care Through Artificial Intelligence: The Future Is Here. *Popul. Health Manag.* **00**, pop.2018.0129 (2018).
33. Fenner, B. J., Wong, R. L. M., Lam, W.-C., Tan, G. S. W. & Cheung, G. C. M. Advances in Retinal Imaging and Applications in Diabetic Retinopathy Screening: A Review. *Ophthalmol. Ther.* **7**, 333–346 (2018).
34. Blei, D., Jordan, M. & Ng, A. Y. Latent Dirichlet allocation. *J. Mach. Learn. Res.* **3**, 993–1022 (2003).
35. Shelhamer, E., Long, J. & Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**, 640–651 (2017).

## Acknowledgements

The authors are indebted to the all the physicians and medical staffs in Fujita Health University Hospital to collect the medical information from EHR. They also thank the system engineers in the Division of Medical Information Systems of Fujita Health University and Kohtaroh Miyamoto and Toru Aihara in IBM Research. This study was supported by the contribution from The Dai-ichi Life Insurance Company, Limited. The funding sources had no role in the study design, data collection, management, analysis, interpretation of the data, or manuscript preparation.



### Author Contributions

M.M., R.Y., M.O., T.I., T.K., A.K., M.K., R.Y. and A.S. carried out the initial analysis and interpretation of data and drafted the manuscript. K.H. and J.K. finally approved the contents of the manuscript. E.S., K.H. and Y.Y. coordinated and supervised the study and critically reviewed the manuscript.

### Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-019-48263-5>.

**Competing Interests:** Dr. Suzuki's work has been funded by research support from MSD, Ono Pharmaceuticals, Chugai Pharmaceuticals, and Takeda Pharmaceuticals. Mr. Ono, Mr. Itoko., Mr. Katsuki, Mr. Koseki., Mr. Kudo are employee of IBM Co. Mr. Haida. and Mr. Kuroda. are employee of The Dai-ichi Life Insurance Company, Limited. Dr. Makino, Dr. Yoshimoto, Dr. Saitoh, Dr. Hoshinaga and Dr. Yuzawa declare no potential conflict of interest.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019