

Predicting Patients' Treatment Behavior by Medical Data Analysis Using Machine Learning Technique

*Hisashi Kurasawa, Akinori Fujino,
and Katsuyoshi Hayashi*

Abstract

The analysis of medical information including electronic health records and medical image data using artificial intelligence (AI) has been an active area of research in recent years. The objective is to develop a means of supporting decision-making about treatment and medicine prescription by medical doctors. NTT has also been designing an AI for supporting diabetes treatment in collaboration with the University of Tokyo Hospital. We introduce here our machine-learning based technique to predict missed scheduled clinical appointments, which are likely to trigger treatment discontinuation by patients with diabetes.

Keywords: diabetes, treatment discontinuation, machine learning

1. Treatment discontinuation in diabetes treatment

The number of diabetes patients has been gradually increasing throughout the world. There are now over 3 million diabetic patients in Japan and approximately 380 million worldwide. Diabetes progression causes complicating diseases including diabetes retinopathy and nephropathy and results in a reduction in the quality of life and an increase in medical costs. Thus, it is vital for diabetic patients to begin medical treatment early and keep their regular hospital appointments to control their blood glucose level. However, in Japan, about 10% of diabetic patients discontinue treatment and return to the hospital after diabetes progression. This is one of the major problems in diabetic care (**Fig. 1(a)**).

Researchers working to solve this problem have been studying the factors related to treatment discontinuation and have classified diabetic patients into two groups: those continuing and those discontinuing treatment. Multifactorial data were used to identify the differences between the two groups as risk factors

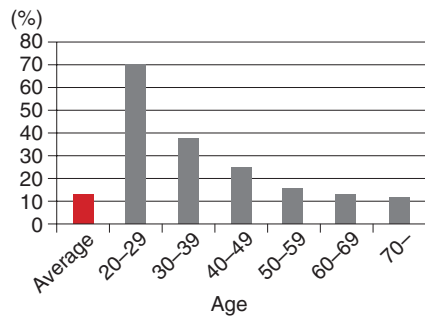
by testing their statistical significance. Factors such as being *male* or *holding a job* proved to be high risks for treatment discontinuation [1] (**Fig. 1(b)**).

There are various factors related to treatment discontinuation, and it is therefore quite difficult for diabetes specialists to identify the patients who need a higher level of support. Therefore, NTT has been collaborating with the University of Tokyo Hospital in designing a prediction model that can predict treatment discontinuation at an individual level (**Fig. 2**). We designed the model to have features generated by a machine-learning technique that NTT developed and by applying knowledge of medical data analysis and the clinical experience of staff at the University of Tokyo Hospital [2].

2. Design of prediction model of treatment discontinuation

We focused our efforts on designing a model to predict a missed clinical appointment (MA) that can be a trigger for treatment discontinuation in order to identify the diabetic patients who require stronger

(a) Treatment discontinuation ratio of diabetes patients by age bracket



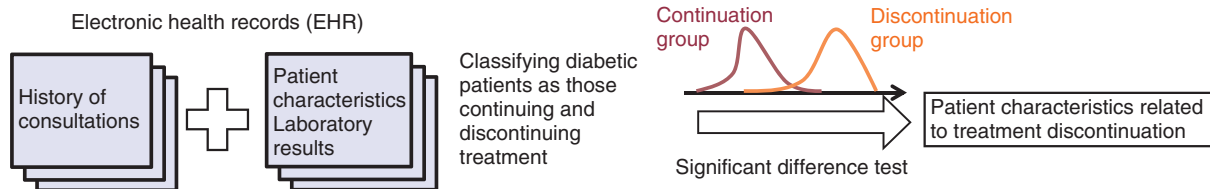
Source: Heisei 24-year National Health and Nutrition Survey (in Japanese).
<http://www.mhlw.go.jp/bunya/kenkou/eiyoudl/h24-houkoku-06.pdf>

(b) Examples of patient characteristics related to treatment discontinuation

- Males; people holding jobs
- Young people (under 50 years of age, especially those 20–30 years old)
- Patients whose blood glucose level is well-controlled/poorly controlled
- Patients who have discontinued treatment in the past

Fig. 1. Statistics related to treatment discontinuation by diabetes patients.

Conventional: Effort to clarify patient characteristics related to treatment discontinuation



This study: Effort to predict missed clinical appointments

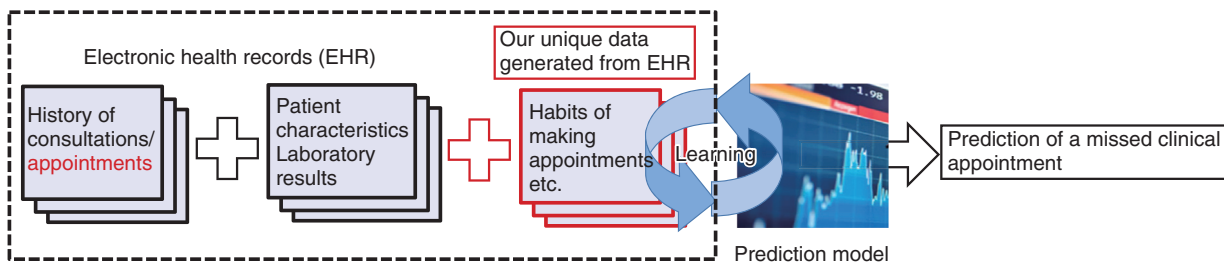


Fig. 2. Prediction model to identify the likelihood of treatment discontinuation by diabetes patients.

support. We designed a logistic regression model that predicts a class y from a feature vector x of a patient’s target appointment, with y representing a clinical appointment missed ($y = +1$) or kept ($y = -1$) and x generated from the patient’s electronic health record (EHR), representing the time from the initial visit to just before the target appointment. We modeled the probability of an appointment with x attributed to y with a logistic regression,

$$P(y | \mathbf{x}; \mathbf{w}) = \frac{1}{1 + \exp(-y\mathbf{w} \cdot \mathbf{x})}$$

We computed the w estimate by using training data $\{(x_1, y_1), \dots, (x_n, y_n), \dots, (x_N, y_N)\}$. In the feature design explained below, observed values of each quantitative variable, for example, blood test results, were linearly transformed (normalized) to make a variance of each variable = 1 and assigned to vector x .

When the training data size N is smaller than the dimension of feature vectors, or a training data sampling is biased, maximum-likelihood estimation often overfits a logistic regression model into the training

Table 1. Examples of X1 group.

Category	Features
Consultation	Past/recent consultation of medical departments
Disease	Past/recent diagnosed diseases; recovered from/under treatment
Medicine	Past/recent prescribed medicines, prescription term, number of doses taken per day

Table 2. Examples of X2 group.

Category	Features
Property	Gender, age Distance and time from home to hospital by public transport
Consultation	Frequency of clinic visits Probability of visit on a given day of the week
Appointment	Interval between the date on which a clinical appointment was made and the scheduled appointment date; Day of the week when an appointment was made; Whether the patient kept an appointment on the day it was made or missed the appointment without notice; Weather on the appointment day
Medicine	Length of the prescription term of the latest medicine Total amount of medication a day How many times a day medication is taken Maximum size of prescribed tablets
Laboratory results	HbA1c, HDL-C, LDL-C, TG, and T-Cho

data, leading the model to classify many appointments inaccurately. Thus, we used an L2-norm regularization method to mitigate overfitting and improve the model's generalizability.

3. Feature design

To obtain a prediction model with higher accuracy, we designed features using the knowledge of diabetes specialists, patients' opinions, and the results of our behavior analysis. We designed two types of features: one that was automatically generated from EHR (X1 group, $n = 29,025$), and another that could be related to treatment discontinuation.

Examples of the X1 group are listed in **Table 1**. These features concern the diagnosis and treatment department that each patient visited before the clinic appointment day, the name of their disease and the prescribed medication, the length of the prescription term (number of weeks, months, etc.), surgical procedures, and exams.

Examples of the X2 group are listed in **Table 2**. We designed these features based on patients' behavior. Thus, the behavior characteristics and environmental factors of each patient weight the MA. For example, we included the day of the week and the weather on the day of a clinical appointment, and the pill size of

the prescribed medication. We also referred to interviews with specialists and patients. A diabetes specialist explained that with patients who often forget to take their medication, the specialist monitored compliance by comparing the intervals between scheduled clinical appointments, the length of the prescription term, and the amount of medication the patient still had, so we incorporated the calculation as a check on compliance.

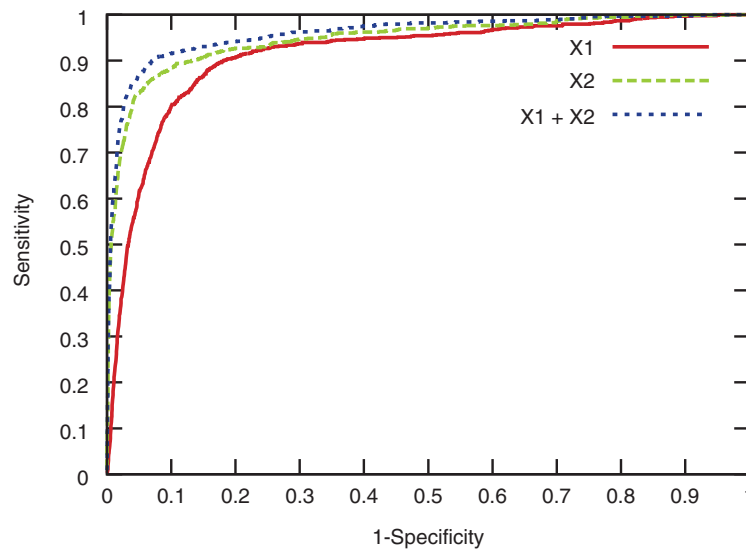
A diabetic patient told us that the accumulated experience with medical treatment in general stimulated his motivation to continue diabetes treatments. Thus, we used the history of clinic visits and the number of diagnosed diseases. We referred to databases maintained by the Japan Meteorological Agency and the Pharmaceuticals and Medical Devices Agency to respectively find weather information and the size of prescribed medication. In addition, we used gender, age, distance from home to hospital, and travel time, which previous research had identified as factors related to treatment discontinuation.

4. Evaluation of prediction performance

All prediction experiments were performed using records from the University of Tokyo Hospital, which included the history of 16,026 clinical appointments

Table 3. Actual results based on patients' behavior in making appointments.

Item description	Number (number of patients)
New appointment registration	32,962 (879)
Appointment schedule change	4,680 (619)
Medical consultation with an appointment (continuation group)	15,104 (869)
Medical consultation without an appointment	228 (116)
Missed appointment but patient had other appointment in the diabetes department on a later day	493 (232)



The larger the area under the ROC curve (AUC value) (i.e., the closer the curve is to the upper left), the higher the prediction performance (0.5 in a completely random case).

Fig. 3. ROC curve measuring prediction performance of logistic regression model with L2-norm regularization by using the X1, X2, and X1 + X2 groups.

scheduled by 879 patients whose initial clinical visit had been made after January 1, 2004, who had diagnostic codes indicative of diabetes, and whose HbA1c had been tested within three months after their initial visit. The records were dated between April 1, 2011 and June 30, 2014. We used the data to predict MAs in appointments. The actual number of missed appointments and kept appointments was 922 and 15,104, as indicated in **Table 3**.

We examined the difference in prediction performance between models by using the X1, X2, and X1 + X2 groups with a receiver operating curve (ROC) (**Fig. 3**). The model compared the area under the curve (AUC) values between them. An AUC value of 0.943 was obtained for the X2 group, leading to a prediction of MAs with higher accuracy than for the

Table 4. AUC value, precision, recall, and F-measure.

Features	AUC	Precision	Recall	F-measure
X1	0.912	0.466	0.438	0.449
X2	0.943	0.750	0.577	0.652
X1 + X2	0.958	0.757	0.659	0.704

X1 group (AUC = 0.912). The AUC for both the X1 and X2 groups (AUC = 0.958) yielded even greater accuracy, as indicated in **Table 4**. An AUC value of more than 0.9 generally means excellent discrimination in such performance evaluations. We can therefore say that our model can predict MAs with high accuracy and help to identify the patients needing

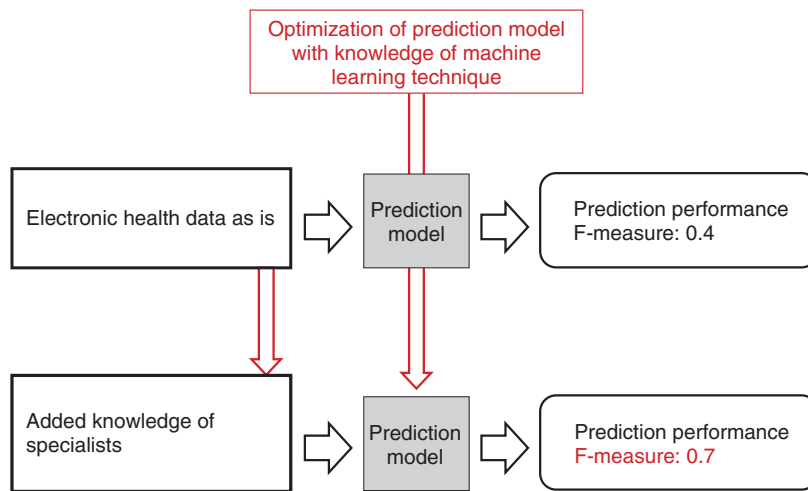


Fig. 4. Relationship between input data and prediction performance.

Table 5. Features contributing to MA prediction.

Feature	Weight
Appointment was made on a Sunday	2.22
Appointment was scheduled for a Friday	1.91
Interval between when a clinical appointment was made and the date for which it was scheduled	-0.912
Graves' disease was diagnosed and recovered from	-1.36
Probability the previous appointment was made on a Monday	0.897
Rilmazafone hydrochloride hydrate recently prescribed	1.29

stronger support to maintain treatment; it also helps to determine when doctors should support them. The precision, recall, and F-measure of the model when using both X1 and X2 groups were 0.757, 0.659, and 0.704, respectively, although the F-measure was 0.449 when using only the X1 group, as indicated in Table 4. We can also say that we succeeded in obtaining a model with high prediction performance by designing features related to treatment behaviors. The relationship between input data and prediction performance is shown in Fig. 4.

5. Features contributing to MA prediction

We compared the weights of features used in this study to investigate what the contributive features were for MA prediction. The contributive features are listed in Table 5. These features were given the largest absolute weight in our trained model and thus contributed strongly to MA prediction. We found that

features related to when and how appointments were made, rather than to patients' clinical condition, influenced the accuracy in predicting MAs. Some contributive features were, for example, "Appointment was made on Sunday," "Appointment was scheduled on Friday," and "Interval between when a clinical appointment was made and the date for which it was scheduled." Of these, the last characteristic amount is a negative value, which means that the shorter the interval between when a clinical appointment was made and the date for which it was scheduled, the more likely a missed clinical appointment will occur. These features found by our model were valued by a diabetes specialist because it is generally difficult for medical doctors to find them.

We believe that our model will be a powerful tool to improve patients' medical condition since it makes it possible to identify patients who need more support, to assess when they should be strongly supported, and to control the strength of intervention with each

patient.

6. Future development

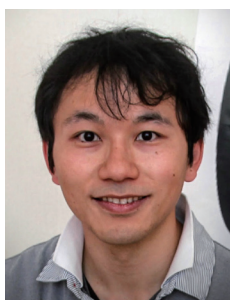
We designed a model that can predict with high accuracy when a diabetic patient is likely to miss a scheduled hospital appointment, possibly leading to discontinuation of treatment. This is the first study utilizing machine learning to design this kind of model.

Various techniques that can provide information appropriate to humans have been proposed as recommendation techniques in the machine learning field. We plan to continue studying machine learning in order to develop solutions to prevent treatment discontinuation. One of our collaborators who has been studying telemedicine demonstrated that patients' behavior related to their meals can be improved if they are given feedback on their lifestyle habits based on the data [3]. Thus, we believe that giving patients

feedback on the confirmation of appointments will be one of the solutions to motivate patients who may potentially discontinue treatment. The standards for EHR and SS-MIX2 (Standardized Structured Medical Information Exchange) are now being widely implemented, and an environment for handling massive amounts of clinical data is being prepared, so we will focus our efforts on improving our prediction model while utilizing these advances.

References

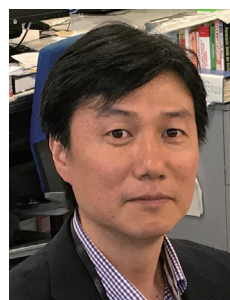
- [1] M. Noda, et al., "Diabetes Consultation Discontinuation Countermeasure Comprehensive Guide," 2014 (in Japanese).
- [2] H. Kurasawa, K. Hayashi, A. Fujino, K. Takasugi, T. Haga, K. Waki, T. Noguchi, and K. Ohe, "Machine-learning-based Prediction of a Missed Scheduled Clinical Appointment by Patients with Diabetes," *J. Diabetes Sci. Technol.*, Vol. 10, No. 3, pp. 730–736, 2016.
- [3] K. Waki, H. Fujita, Y. Uchimura, K. Omae, E. Aramaki, S. Kato, H. Lee, H. Kobayashi, T. Kadowaki, and K. Ohe, "DialBetics: A Novel Smartphone-based Self-management Support System for Type 2 Diabetes Patients," *J. Diabetes Sci. Technol.*, Vol. 8, No. 2, pp. 209–215, 2014.



Hisashi Kurasawa

Research Engineer, Ubiquitous Service Systems Laboratory, NTT Network Innovation Laboratories.

He received a B.E., M.E., and Ph.D. in information science and technology from the University of Tokyo in 2006, 2008, and 2011. He joined NTT Network Innovation Laboratories in 2011, where he studied sensor data mining. He joined a medical AI research project in 2013 and is studying prediction models of clinical condition, including treatment discontinuation prediction. He is also currently studying biomedical signal analysis at NTT Basic Research Laboratories. He is a member of the Institute of Electronics, Information and Communication Engineers, the Information Processing Society of Japan, the Japan Association for Medical Informatics, and the Database Society of Japan.



Katsuyoshi Hayashi

Senior Research Engineer, Social Device Technology Laboratory, NTT Device Technology Laboratories.

He received a B.E., M.E., and Ph.D. in applied chemistry from Waseda University, Tokyo, in 1996, 1998, and 2005. He joined NTT Basic Research Laboratories in 1998 and began studying electrochemical biosensing devices to detect biomolecules. During 2007–2008, he was a visiting scientist in the Biomedical Engineering Department, University of Wisconsin-Madison, USA, where he studied cell biology with microfluidic devices. He launched a medical AI research project with Drs. Kurasawa and Fujino in 2013, and is studying prediction models of clinical conditions, including treatment discontinuation prediction. He is also currently studying surface plasmon resonance-based microfluidic sensing devices to detect biochemical reaction and biomarkers for point-of-care testing at NTT Device Technology Laboratories. He is a member of the Chemical Society of Japan, the Electrochemical Society of Japan, and the Japan Association of Chemical Sensors.



Akinori Fujino

Senior Research Scientist, Learning and Intelligent Systems Research Group, NTT Communication Science Laboratories.

He received a B.E. and M.E. in precision engineering from Kyoto University in 1995 and 1997, and a Ph.D. in informatics from Kyoto University in 2009. He joined NTT in 1997. His current research interests include machine learning and knowledge discovery from complex data.