ICIC Express Letters
Volume **18**, Number **6**, June **2024**

ICIC International ©2024 ISSN 1881-803X
pp. **615–624**

DEVELOPMENT OF ASSISTANCE SYSTEM TO DIAGNOSE CHANGES IN KIDNEY FUNCTION USING MACHINE LEARNING

Yuki Yamaguchi[1], Noritaka Shigei[1,*], Masanobu Miyazaki[2]
Yoichi Ishizuka[3], Shinichi Abe[4], Tomoya Nishino[4] and Hiromi Miyajima[1]

[1]Graduate School of Science and Engineering
Kagoshima University
1-21-40 Korimoto, Kagoshima City, Kagoshima 890-0065, Japan
{ k6072255; k2356323 }@kadai.jp; *Corresponding author: shigei@ibe.kagoshima-u.ac.jp

[2]Miyazaki Medical Clinic
3-12 Shiratorimachi, Nagasaki City, Nagasaki 852-8042, Japan
msnbmiya@gmail.com

[3]Graduate School of Engineering
Nagasaki University
1-14 Bunkyo, Nagasaki City, Nagasaki 852-8521, Japan
isy2@nagasaki-u.ac.jp

[4]Department of Nephrology
Nagasaki University Hospital
1-7-1 Sakamoto, Nagasaki City, Nagasaki 852-8501, Japan
{ s-abe; tnishino }@nagasaki-u.ac.jp

Received August 2023; accepted October 2023

Abstract. *To reduce the labor required for physicians to make diagnosis from time series data on kidney function from specific health checkups, this paper proposes 1) cooperative learning in which a sequence of creating a flowchart (FC) and training machine learning (ML) models is iteratively repeated and 2) a decision assistance system presenting suggestion based on the created FC and ML models. The cooperative learning aims to reduce the labor for creating FC and preparing supervised data. The decision assistance system aims to achieve both evidence-based judgments and accuracy. The effectiveness of the proposed framework is demonstrated in an example implementation using real data.*
Keywords: Kidney function, Time series data, Trend, Machine learning, Flowchart

1. **Introduction.** The number of chronic dialysis patients in Japan exceeds 340,000, and the number is increasing year by year [1]. One of the social issues is to reduce the number of dialysis patients due to chronic kidney disease (CKD) since this will lead to an increase in healthy life expectancy and a reduction in medical costs. Early detection of deteriorating kidney function (KF) and early initiation of treatment are important to prevent chronic kidney disease. Therefore, a major challenge is how to identify patients at high risk of KF deterioration and how to create a system to provide health guidance and referral to specialists.

Some municipal governments and medical associations hold case review meetings for specific health checkups and prepare comments for family doctors and examinees in order to encourage patients who need treatment to be examined by nephrologists. However, the judgment and preparation of comments are largely manual work by physicians and public health nurses and require much labor. In order to predict the deterioration of KF, it is important to grasp the trend of KF from time series data. However, it is not easy to diagnose the deterioration of renal function because the values of physical examinations may

DOI: 10.24507/icicel.18.06.615

fluctuate up and down depending on various factors, regardless of the trend. Although a flowchart-like criterion would help to reduce the burden, it is difficult to establish an appropriate criterion in the situation above. Furthermore, in such cases, the creation of the flowchart itself would be a difficult task.

There have been many studies on machine learning (ML) using CKD data. However, most of these studies, such as in [2, 3, 4], are conducted using data from the UCI machine learning repository [5]. This dataset is non-time series data consisting of 250 CKD patients and 150 non-CKD patients and is mainly used for binary classification. The research by Ventrella et al. in [6] uses time series data to examine the problem of 2- to 4-classification to predict the time to dialysis. However, it does not explicitly determine trends in KF, and what it uses is inpatient data and not data from health screening systems. Unlike previous studies, this study determines the trend of renal function in medical checkup data.

There are two issues in applying ML. One of the challenges in applying ML is to prepare sufficient supervised data. 1) Manual annotation of patient data by medical specialists requires a large amount of time and effort. In addition, ML may not always be based on medical evidence. 2) The output process of ML and deep learning is a black box. It cannot be logically explained to physicians and patients. Therefore, this study proposes the use of flowcharts as a solution to both of these problems.

The purpose of this study is to reduce the amount of labor involved in selecting high-risk groups and preparing commentary. We propose 1) cooperative learning in which a sequence of creating a flowchart (FC) and training an ML model is iteratively repeated and 2) a decision assistance system presenting suggestion based on the created FC and ML models. The cooperative learning aims to reduce the labor for creating FC and preparing supervised data. The decision assistance system aims to achieve both evidence-based judgments and accuracy. These proposals solve the above described two issues in applying ML. The effectiveness of the proposed cooperative learning is demonstrated by showing that supervised data can be prepared with less annotation work and that the accuracies of FC and ML improve. Furthermore, the effectiveness of the proposed judgment method is demonstrated by showing that ML can provide better results for some cases than FC. The rest of this paper is organized as follows. Section 2 describes the diagnosis of trends in KF that this study addresses. Section 3 describes the methods used in this study to determine trends in KF using ML. Sections 4 and 5 propose a framework for cooperative learning and a judgment suggestion system, respectively. Section 6 demonstrates the effectiveness of our proposal. Finally, Section 7 is the conclusion of this paper.

2. CKD Prevention Based on Specific Health Checkups.

2.1. Specific health checkups and specific health guidance.

Some municipalities in Japan provide health guidance based on specific health checkups. The specific health checkup focuses on metabolic syndrome, which can be taken once a year by people aged 40-74 years to prevent lifestyle-related diseases. In addition, for those who are at high risk of developing lifestyle-related diseases and for those who are expected to be highly effective in preventing lifestyle-related diseases by improving their lifestyle, public health nurses and dietitians provide specific health guidance to review their lifestyle habits.

2.2. CKD prevention efforts based on specific health checkups.

CKD is characterized by persistent kidney damage and KF decline that progresses to end-stage renal failure requiring dialysis therapy or kidney transplantation. The number of patients on chronic dialysis continues to increase, posing a significant healthcare economic problem. CKD also increases the risk of cardiovascular diseases such as myocardial infarction, stroke, heart failure, and death [7]. The onset of CKD is closely related to lifestyle diseases and age-related decline in KF, making the family physician the center of care for

CKD patients. However, CKD includes renal diseases such as IgA nephropathy that require treatment by nephrologists or specialized medical institutions in the community. Therefore, family physicians, nephrologists, and specialized medical institutions need to cooperate in treating CKD patients [7].

One of co-authors, a nephrologist, participates in a case study group using health checkup data, in which nephrologists and diabetes specialists, the government, and medical associations collaborate [8]. In this study group, health checkup data are used to identify high-risk groups for CKD and to develop recommendations for health guidance and referrals to specialists. The health checkup data are time-series data covering 1 to 5 years. The advice consists of the status of KF, urinary protein, blood glucose, blood pressure, etc., and whether or not a referral to a specialist is required.

Glomerular filtration rate (GFR) is used to diagnose and classify the severity of CKD. It represents the amount of wastes, such as creatinine and inulin, that the kidneys eliminate as urine. GFR can be measured by clearance testing, which is not easily performed in an outpatient setting. Therefore, the estimated glomerular filtration rate (eGFR) calculated from gender, serum creatinine Cr, and age α is generally used. For Japanese, $eGFR$ (ml/min/1.73m^2) is calculated as follows [9]: For male, $eGFR = 194 \times Cr^{-1.094} \times \alpha^{-0.287}$, and for female, $eGFR = 194 \times Cr^{-1.094} \times \alpha^{-0.287} \times 0.739$.

2.3. **Diagnosis of kidney function trends.** In preparing the advice described in Section 2.2, it is necessary to appropriately capture trends in KF. In our initial study, the time series data of eGFR for each examinee were classified into the seven classes of trends as shown in Figure 1. R6 is assigned when the number of checkups is two and KF is decreasing; R7 is assigned when the number of checkups is one and KF is low.

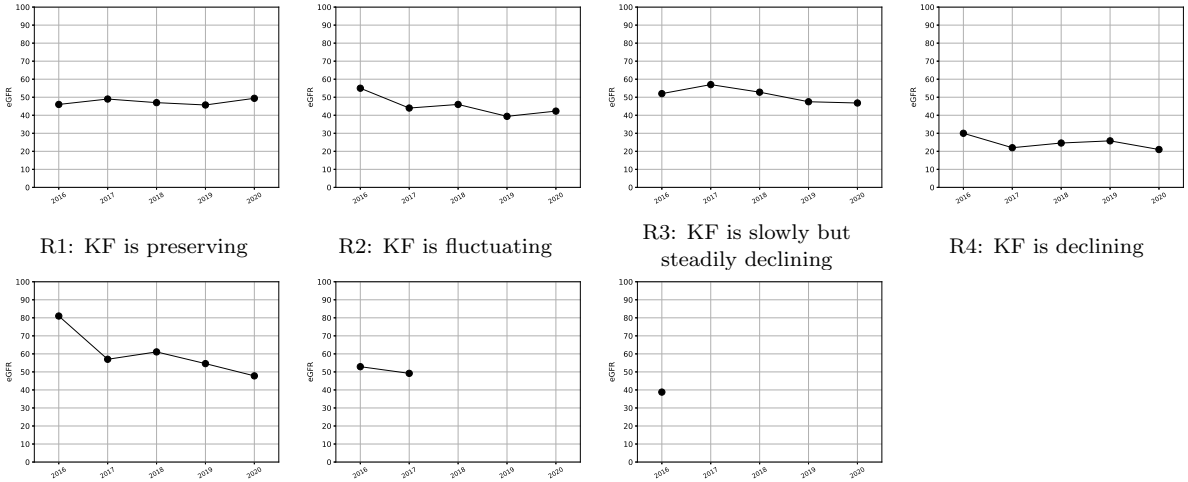

FIGURE 1. Examples of eGFR time-series data for seven classes of trends in kidney function

In progress of our study, we reviewed the classification classes and considered two cases, six-class and four-class as follows. In the case of six-class, the data of R7, which are not time series data, were deleted. In the case of four-class, in addition to that, R2 was merged into R3 and R6 into R4, resulting in four classes of R1, R3, R4, and R5.

In the health checkup data, the number of checkups and the duration of checkups vary from examinee to examinee, and there is a problem of missing values. In addition, it is difficult to make appropriate judgments based only on simple conditions such as slope in time-series data, where the series length varies and fluctuates.

3. **Estimation of Kidney Function Trends Using Machine Learning.** This section describes the machine learning-based trend estimation method proposed in [10] for the problem described in Section 2.3. The method is used in this study. This previous study proposed data preprocessing to deal with different numbers of examinations and different periods of examinations and three ensemble learning-based methods using the Gradient Boosting Decision Tree (GBDT) as models for ML. This study uses LightGBM for the final ML model. The data used to determine the trend of KF are the health checkup data of eGFR for the maximum examination period of Y years: x_1, x_2, ..., and x_Y. Here, x_y is the value of eGFR of $Y - y$ years ago, and $x_y = \mathrm{NA}$ if $Y - y$ years ago is not examined. Let $S_\mathrm{Y} = \{y | x_y \neq \mathrm{NA}\}$ be the set of years of examination. In addition to the eGFR values x_1, x_2, ..., and x_Y, the explanatory variables are the *eGFR*'s difference $\Delta = x_Y - x_{\min_{y \in S_\mathrm{Y}} y}$, the regression line's slope a given by Equation (1), the number of examinations $N = |S_\mathrm{Y}|$, and the examination period $T = Y + 1 - \min_{y \in S_\mathrm{Y}} y$.

$$a = \frac{N \sum_{y \in S_\mathrm{Y}} y x_y - \sum_{y \in S_\mathrm{Y}} y \sum_{y \in S_\mathrm{Y}} x_y}{N \sum_{y \in S_\mathrm{Y}} y^2 - \left(\sum_{y \in S_\mathrm{Y}} y \right)^2} \tag{1}$$

If there are years in the examination period T in which no examination is taken, the explanatory variable x_y for those years is linearly interpolated using the values before and after the examination. That is, for all $y > \min_{y' \in S_\mathrm{Y}} y'$ such that $x_y = \mathrm{NA}$, the value of the medical examination is given by the following equations. $x_y = x^- + (x^+ - x^-)(y - y^-)/(y^+ - y^-)$, where y^- and x^- are the year and the value of the medical checkup immediately before year $y \in S_\mathrm{Y}$ and y^+ and x^+ are the year and the value of the medical checkup immediately after year $y \in S_\mathrm{Y}$.

GBDT has shown good accuracy in various applications, and LightGBM is a framework for implementing GBDT. This previous study uses LightGBM as a baseline model for ML. Decision trees classify data from the root to leaves based on the given features. GBDT is an ensemble learning method that creates decision tree models sequentially and combines these models. The created decision tree models improve accuracy by learning the residuals between the terget values and the sum of the predictions of all models up to the previous decision tree model as the loss function. LightGBM has a shorter computation time than other frameworks, because it incorporates gradient-based one-side sampling (GOSS), which excludes from training data with small contribution to accuracy improvement, and exclusive feature bundling, which effectively reduces the number of features.

4. **Cooperative Learning between Machine Learning and Flowchart Creation.** It is difficult to create a flowchart for a classification task for which clear decision criteria are unknown. On the other hand, in ML, the preparation of teacher data required for training models is labor-intensive in the annotation process. We propose a framework for cooperative learning between ML and flowchart creation as a method to solve both of these problems.

4.1. **General algorithm of the proposed framework.** The outline of the proposed framework is shown in Figure 2. In the proposed framework, Step 1 begins with the annotation of a tiny amount of data by experts of the target domain. This work may utilize existing data produced by the experts. In Step 2, the worker observes this data, extracts the main conditions, and creates a flowchart. In Step 3, annotation work on the increased data is performed by assigning the judgment results of the flowchart as labels for the data. In Step 4, an ML model is trained by using the data in Step 3. In Step 5, experts review only the data for which the judgment result of the ML model differs from that of the flowchart, correct the labels and modify the flowchart. After that, Steps 3, 4 and 5 are repeated a certain number of times to update the ML model and modify the flowchart.

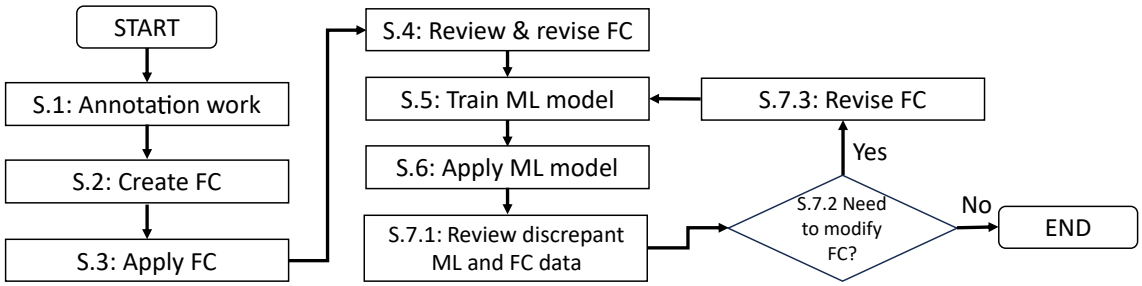

FIGURE 2. The outline of the proposed framework of cooperative learning

This flow reduces the expert's effort for the following reasons. The expert only needs to process a tiny number of data in Step 1 for the annotation work, and in Step 5, the expert only needs to review a relatively small number of data. In addition, the expert's effort in creating the flowchart is small because it is mainly created by the worker based on the data. On the other hand, the annotation work for ML can be semi-automated for a large amount of data since the flowchart is used to perform the annotation work.

The general algorithm for this framework is as follows.

Step 1: Experts of the target domain perform the annotation work on a small dataset.
Step 2: Workers create a flowchart (FC) by the annotation data.
Step 3: Apply the FC to a larger dataset and obtain the judgment result.
Step 4: Experts review and revise the judgment result in Step 3. Workers revise the FC according to the revised judgment result.
Step 5: Apply the FC to a larger and/or the same dataset and obtain the judgment results of the FC as labeled datasets.
Step 6: Train an ML model (or ML models with cross-validation) by using revised labeled dataset or dataset labeled by FC, and apply the ML model (or ML models) to the largest dataset so far to obtain judgment result by ML models.
Step 7: Experts review only the data for which the judgment result of the ML model or models differs from that of the FC and correct the labels. If any revisions are made, workers modify the FC and go to Step 5. Otherwise, terminate the algorithm.

4.2. **Implemented flow of the proposed framework.** The proposed framework described in Section 4.1 was implemented using three different datasets: a small dataset D_a consisting of 28 cases, a medium dataset D_b consisting of 310 cases, and a large dataset D_c consisting of 3157 cases. The specific flow is shown below.

Step 1: A nephrologist assigns one of the seven classes from R1 to R7 to each case in the small dataset D_a consisting of 28 cases based on comments made by physicians and public health nurses. The obtained labeled dataset of D_a is referred to as JD_a.
Step 2: Workers create an FC F_1 according to JD_a.
Step 3: For each case in the medium dataset D_b consisting of 405 cases, by applying F_1, one of the seven classes is assigned as a label. The labeled dataset of D_b by F_1 is referred to as JF_{b1}.
Step 4: The nephrologist reviews and revises the labels of JF_{b1}, where the classes of labels are reduced from the seven classes to the six classes of R1∼R6. The revised labeled dataset of D_b is referred to as JD_{b1}. According to JD_{b1}, the worker revises F_1 as F_2.
Step 5: For each case in D_b, by applying F_2, one of the six classes R1∼R6 is assigned as a label. As a result, a labeled dataset JF_{b2} is obtained for D_b.
Step 6: All cases of D_b are judged by ML models provided from leave-one-out cross-validation (LOOCV) with JD_{b1} of the training data set. The obtained ML models and the judgment result of D_b by M_{b1} are referred to as M_{b1} and JM_{b1}, respectively.
Step 7: The nephrologist reviews only cases where JD_{b1}, JF_{b2}, and JM_{b1} did not match, and revises JD_{b1} as JD_{b2}. According to JD_{b2}, the worker revises F_2 as F_3.

Step 5 (2nd): For each case in D_b and D_c consisting of 3157 cases, by applying F_3, one of the six classes R1∼R6 is assigned as a label. The obtained labeled datasets of D_b and D_c by F_3 are referred to as JF_{b3} and JF_{c3}, respectively.
Step 6 (2nd): An ML model M_{b2} is trained by using JD_{b2} as the training dataset. All cases of D_c are judged by the ML model M_{b2}, and the judgment result JM_{c2} on D_c by M_{b2} is obtained.
Step 7 (2nd): The nephrologist reviews only cases where JF_{c3} and JM_{c2} did not match, and creates the labeled dataset JD_{c3} for D_c, where the classes of labels are reduced to the four classes of R1, R3, R4, and R5. According to JD_{c3}, the worker revises F_3 as F_4.
Step 5 (3rd): For each case in D_c, by applying F_4, one of the four classes is assigned as a label. The obtained labeled datasets of D_c by F_4 are referred to as JF_{c4}.
Step 6 (3rd): All cases of D_c are judged by ML models provided from 5-fold cross validation with JF_{c4} of the training data set. The obtained ML models and the judgment result of D_c are referred to as M_{c4} and JM_{c4}, respectively.
Step 7 (3rd): The flow completes with the final FC F_4, FC's judgment JF_{c4}, ML models M_{c4} and ML's judgment JM_{c4}.

5. **Judgment Suggestion Using Machine Learning and Flowchart.** In the judgment of trends in KF, although the flowchart has clear criteria, there are cases where the physician's conclusion differs from the flowchart's. On the other hand, although machine learning can make flexible judgments that cannot be made with the flowchart, the basis for such conclusions is unclear. To address these issues, we propose a system that assists physicians in making decisions by combining flowchart judgments and machine learning judgments.

Using the final flowchart F_4 and machine learning model M_4 constructed in the flow presented in Section 4.2, we develop a system that presents information to help the physician make a decision. The LightGBM is used as the machine learning model, and its decision result can output the probability for each class in addition to the decision class. In presenting information, these probabilities are also utilized. The proposed system shows the following information related to judgment: 1) The judgment result of FC F_4 and 2) M_4's predicted probabilities of classes R1, R3, R4, and R5.

Figure 3 shows an example of information presented by the proposed system. In the table of "Suggestion by FC and ML", the four classes of R1, R3, R4, and R5 are displayed in descending order of M_4's predicted probabilities. In column "FC", the judgment of F_4 is marked with "*". If the FC's decision agrees with the ML's one having the highest predicted probability, then the physician may simply adopt that decision. Otherwise, the physician may make a selection based on his or her own thinking from those that have higher prediction probabilities. This procedure of decision making would allow the physician to make a prompt decision.

eGFR by year

Year	2016	2017	2018	2019	2020
eGFR	81.0	57.0	61.1	54.6	47.8

Suggestion by FC and ML

FC	Probability	Trend
*	54.6 %	R5. Rapidly decreasing
	38.2 %	R4. Decreasing
	26.8 %	R3. Slow but steadily decreasing
	4.5 %	R1. Preserved

FIGURE 3. An example of information presented by the suggestion system

TABLE 1. Evaluation results on flowcharts F_1, F_2, F_3, and F_4

(a) For F_1 when ground truth is JD_{b1}

	Precision	Recall	F1-score	Support
R1	0.8294	0.9959	0.9050	244
R2	0.0000	0.0000	0.0000	9
R3	0.0000	0.0000	0.0000	27
R4	0.7778	0.4667	0.5833	15
R5	0.1667	0.5000	0.2500	2
R6	0.0000	0.0000	0.0000	13
Accuracy	0.8097			310

(b) For F_2 when ground truth is JD_{b2}

	Precision	Recall	F1-score	Support
R1	0.9750	0.9590	0.9669	244
R2	0.8333	0.5556	0.6667	9
R3	0.6053	0.8519	0.7077	27
R4	0.8000	0.5333	0.6400	15
R5	0.1667	0.5000	0.2500	2
R6	0.5000	0.3846	0.4348	13
Accuracy	0.8903			310

(c) For F_3 when ground truth is JD_{b2}

	Precision	Recall	F1-score	Support
R1	0.9556	0.8811	0.9168	244
R2	0.4167	0.5556	0.4762	9
R3	0.4545	0.9259	0.6098	27
R4	0.8000	0.5333	0.6400	15
R5	0.1667	0.5000	0.2500	2
R6	0.5000	0.0769	0.1333	13
Accuracy	0.8226			310

(d) For F_3 (4 classes) when ground truth is JD_{b2}

	Precision	Recall	F1-score	Support
R1	0.9556	0.8811	0.9168	244
R3	0.4627	0.8611	0.6019	36
R4	0.8333	0.3571	0.5000	28
R5	0.1667	0.5000	0.2500	2
Accuracy	0.8290			310

(e) For F_4 when ground truth is JD_{b2}

	Precision	Recall	F1-score	Support
R1	0.9569	0.9098	0.9328	244
R3	0.5200	0.7222	0.6047	36
R4	0.5556	0.5357	0.5455	28
R5	0.0000	0.0000	0.0000	2
Accuracy	0.8484			310

6. Evaluation and Consideration.

6.1. **Accuracy of flowcharts.** In this section, we evaluate the accuracy of flowcharts and ML models at each step of the flow presented in Section 4.2.

The FCs F_1, F_2, F_3, and F_4 use the number of examinations N, the last year's eGFR x_Y, the eGFR change per year $(x_Y - x_1)/(Y-1)$, the regression line's slope a given by Equation (1), and the number of up/down movements in eGFR as explanatory variables.

The evaluation results on FCs F_1, F_2, F_3, and F_4 are shown in Table 1. In the evaluation, since the cases of $N = 1$ are excluded, there is no data with the label R7. From the results, it can be observed that the accuracy of F_2 is improved from F_1. In particular, F1 has scores of 0 for R2, R3, and R6, which is due to inappropriate conditions. This defect has been fixed in F_2. The accuracy of F_3 is lower than F_2. This is due to the fact that the conditions for F_3 are more concise, giving priority to interpretability. At this point, we considered changing from 6 classes to 4 classes and confirmed that this would improve accuracy. F_4 is more accurate than F_3.

6.2. **Accuracy of machine learning models.** Table 2 shows the accuracy of ML models assuming JF_{c4} is ground truth (GT). The ML models for JM_{c2} and JM_{c4} are logistic regression and LightGBM, respectively. JM_{c2} has much worse accuracy for R2, R4, R5, and R6. In contrast, in the final model JM_{c4}, which is for four-class classification, the accuracy of all classes has much improved and achieved accuracy of 98%. On the other hand, in [6], using in-hospital data, ML is applied to predicting the time to dialysis as a four-class classification problem, achieving 89% accuracy for extremely randomized trees.

Table 2. Confusion matrices for ML models

(a) For JM_{c2} when ground truth is JF_{c3}

ML\GT	R1	R2	R3	R4	R5	R6	Prec.
R1	1995	58	203	20	6	1	0.874
R2	63	3	7	5	3	3	0.036
R3	30	13	133	24	6	0	0.646
R4	24	7	37	46	16	2	0.348
R5	0	0	0	4	30	0	0.882
R6	23	1	34	11	32	4	0.038
Rec.	0.934	0.037	0.321	0.418	0.323	0.400	Acc.
F1	0.903	0.036	0.429	0.380	0.472	0.070	0.777

(b) For JM_{c4} when ground truth is JF_{c4}

ML\GT	R1	R3	R4	R5	Prec.
R1	2441	7	6	0	0.995
R3	15	366	7	0	0.943
R4	12	7	261	4	0.919
R5	0	0	4	27	0.871
Rec.	0.989	0.963	0.939	0.871	Acc.
F1	0.992	0.953	0.929	0.871	0.980

Table 3 shows the confusion matrices of the final ML model M_{c4} assuming JD_{c4} is GT, where JD_{c4} is nephrologist's judgment. Table 3(b) is the result when the first and second rankings are considered as the correct answer. It can be observed that most of the correct answers are contained in the top two predictions of the final ML model M_{c4}.

TABLE 3. Confusion matrices of ML model M_{c4} when ground truth is JD_{c4}

(a) When only the first ranking is considered as the correct answer

ML\GT	R1	R3	R4	R5
R1	2449	2	3	0
R3	14	371	3	0
R4	12	4	268	0
R5	0	1	3	27

(b) When the first and second rankings are considered as the correct answer

ML\GT	R1	R3	R4	R5
R1	2472	0	0	0
R3	0	376	0	0
R4	3	1	277	0
R5	0	1	0	27

TABLE 4. Confusion matrices for data where FC and ML did not match

(a) Confusion matrix of ML and FC

ML\FC	R1	R3	R4	R5
R1	0	7	4	0
R3	15	0	7	0
R4	11	7	0	4
R5	0	0	4	0

(b) Confusion matrix of ML and JD_{c4}

ML\GT	R1	R3	R4	R5
R1	8	2	1	0
R3	14	5	3	0
R4	11	4	7	0
R5	0	1	3	0

(c) Confusion matrix of ML's 1st and 2nd ranking and JD_{c4}

ML\GT	R1	R3	R4	R5
R1	30	0	0	0
R3	0	10	0	0
R4	3	1	14	0
R5	0	1	0	0

Table 4 shows the confusion matrices for data where FC and ML did not match. According to Table 4(a), in 59 out of 3157 cases, F_4 and M_{c4} decisions were in disagreement. Table 4(b) shows that M_{c4} gave the correct answer in 20 out of the 59 cases. Furthermore, Table 4(c) shows that in 54 out of the 59 cases, the correct decision is among the top two predictions of M_{c4}.

7. **Conclusion.** In this paper, we proposed cooperative learning and a decision assistance system to reduce the labor required for physicians to make diagnosis of kidney function. The evaluation results demonstrated that 1) collaborative learning improves the accuracy of the flowcharts and ML models and 2) the combined system of flowchart and ML models created by cooperative learning presents useful information to assist physicians in making decisions. In future works, we will apply the proposed system for actual review meetings to evaluate and improve the system.

Acknowledgments. We thank Ms. Y. Anami of Health Support Labo for her efforts in anonymizing the specific health checkup data.

REFERENCES

[1] N. Hanabusa et al., 2021 annual dialysis data report, JSDT renal data registry, *Nihon Toseki Igakkai Zasshi*, vol.55, no.12, pp.665-723, 2022 (in Japanese).
[2] H. Ilyas et al., Chronic kidney disease diagnosis using decision tree algorithms, *BMC Nephrology*, vol.22, 273, 2021.
[3] B. Khan et al., An empirical evaluation of machine learning techniques for chronic kidney disease prophecy, *IEEE Access*, vol.8, pp.55012-55022, 2020.
[4] J. Qin et al., A machine learning methodology for diagnosing chronic kidney disease, *IEEE Access*, vol.8, pp.20991-21002, 2020.

[5] *UCI Machine Learning Repository: Chronic_Kidney_Disease Data Set*, https://archive.ics.uci.edu/ml/datasets/chronic_kidney_disease, Accessed on April 30, 2023.
[6] P. Ventrella, G. Delgrossi, G. Ferrario, M. Righetti and M. Masseroli, Supervised machine learning for the assessment of Chronic Kidney Disease advancement, *Computer Methods and Programs in Biomedicine*, vol.209, 106329, 2021.
[7] Japanese Society of Nephrology, Evidence-based clinical practice guideline for CKD 2018, *The Japanese Journal of Nephrology*, vol.60, no.8, pp.1037-1193, 2018 (in Japanese).
[8] M. Miyazaki et al., Extraction of high-risk groups for exacerbation of nephropathy from specific health checkup data and automatic creation of comments using computer and machine learning [Translated from Japanese], *Journal of the Japan Diabetes Society*, vol.65, no.suppl., p.193, 2022 (in Japanese).
[9] S. Matsuo et al., Revised equations for estimated GFR from serum creatinine in Japan, *American Journal of Kidney Diseases*, vol.53, no.6, pp.982-992, 2009.
[10] Y. Yamaguchi et al., Classification of renal function trends in time series data using machine learning, *Journal of Japan Society for Fuzzy Theory and Intelligent Informatics*, vol.35, no.1, pp.511-516, 2023 (in Japanese).
[11] T. Chen and C. Guestrin, XGBoost: A scalable tree boosting system, *Proc. of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.785-794, 2016.
[12] G. Ke et al., LightGBM: A highly efficient gradient boosting decision tree, *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, pp.3149-3157, 2017.
[13] T. Akiba, S. Sano, T. Yanase, T. Ohta and M. Koyama, Optuna: A next-generation hyperparameter optimization framework, *arXiv Preprint*, arXiv: 1907.10902, 2019.