scientific reports

OPEN

Machine learning-based prediction of relapse in rheumatoid arthritis patients using data on ultrasound examination and blood test

Hidemasa Matsuo[1,7]✉, Mayumi Kamada[2,7], Akari Imamura[1], Madoka Shimizu[1], Maiko Inagaki[1], Yuko Tsuji[1], Motomu Hashimoto[3,4], Masao Tanaka[3], Hiromu Ito[3,5,6] & Yasutomo Fujii[1]

Recent effective therapies enable most rheumatoid arthritis (RA) patients to achieve remission; however, some patients experience relapse. We aimed to predict relapse in RA patients through machine learning (ML) using data on ultrasound (US) examination and blood test. Overall, 210 patients with RA in remission at baseline were dichotomized into remission (n = 150) and relapse (n = 60) based on the disease activity at 2-year follow-up. Three ML classifiers [Logistic Regression, Random Forest, and extreme gradient boosting (XGBoost)] and data on 73 features (14 US examination data, 54 blood test data, and five data on patient information) at baseline were used for predicting relapse. The best performance was obtained using the XGBoost classifier (area under the receiver operator characteristic curve (AUC) = 0.747), compared with Random Forest and Logistic Regression (AUC = 0.719 and 0.701, respectively). In the XGBoost classifier prediction, ten important features, including wrist/metatarsophalangeal superb microvascular imaging scores, were selected using the recursive feature elimination method. The performance was superior to that predicted by researcher-selected features, which are conventional prognostic markers. These results suggest that ML can provide an accurate prediction of relapse in RA patients, and the use of predictive algorithms may facilitate personalized treatment options.

Rheumatoid arthritis (RA), characterized by synovial inflammation that causes progressive joint damage and disability, is among the most frequent chronic inflammatory diseases[1]. Recent effective therapies enable most RA patients to achieve remission; however, some patients experience relapse[2,3]. Several clinical information and biological markers (e.g., gender, disease duration, age, C-reactive protein (CRP), erythrocyte sedimentation rate (ESR), rheumatoid factor (RF)) constitute prognostic factors in RA[4].

Ultrasound (US) is a non-invasive and sensitive method for detecting inflammatory soft tissue and early bone lesions; therefore, the technique is commonly used to assess RA's disease activity[5]. Several studies show that detecting synovitis with the US is associated with the disease progression of RA[6–11]. Recently, our group demonstrated the utility of superb microvascular imaging (SMI), a recent innovative type of Doppler US technology used for visualizing minute vessels with low blood flow velocity, in predicting relapse in RA patients[12]. Among joints generally assessed with the US, we showed that wrist and metatarsophalangeal (MTP) SMI scores might be vital for predicting RA relapse, depending on patients' baseline disease activity[12,13]. However, it is unclear whether combining US data and other features, including clinical and biological markers, improves relapse prediction.

Machine learning (ML) is a type of artificial intelligence that encompasses algorithmic methods that enable machines to solve problems. One of the advantages of ML is the ability to analyze diverse data types (e.g.,

[1]Department of Human Health Sciences, Graduate School of Medicine, Kyoto University, 53 Kawahara-cho, Shogoin, Sakyo-ku, Kyoto 606-8507, Japan. [2]Department of Biomedical Data Intelligence, Graduate School of Medicine, Kyoto University, Kyoto, Japan. [3]Department of Advanced Medicine for Rheumatic Diseases, Graduate School of Medicine, Kyoto University, Kyoto, Japan. [4]Department of Clinical Immunology, Graduate School of Medicine, Osaka City University, Osaka, Japan. [5]Department of Orthopaedic Surgery, Graduate School of Medicine, Kyoto University, Kyoto, Japan. [6]Department of Orthopaedic Surgery, Kurashiki Central Hospital, Okayama, Japan. [7]These authors contributed equally: Hidemasa Matsuo and Mayumi Kamada. ✉email: matsuo@kuhp.kyoto-u.ac.jp
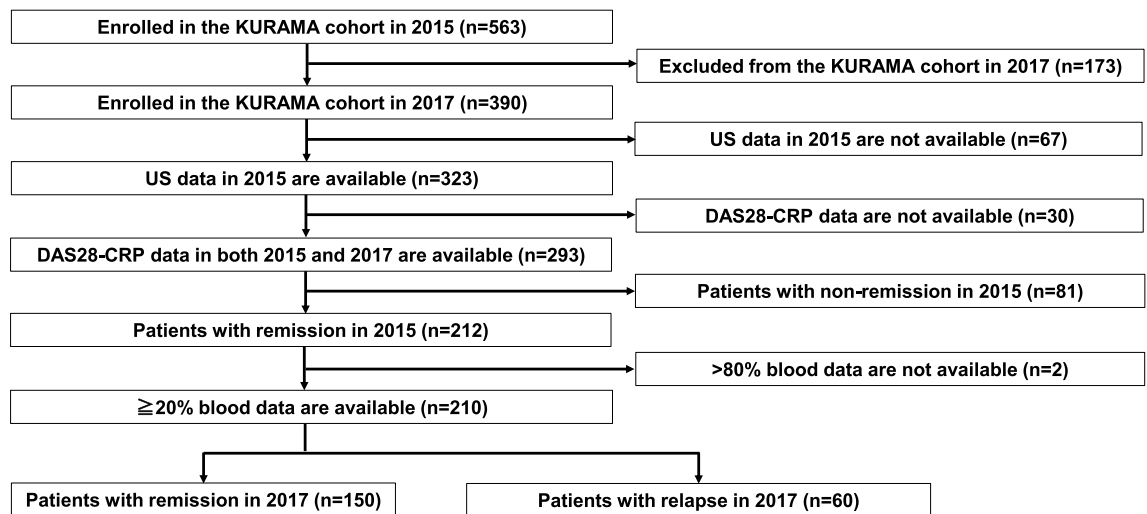
Figure 1. Flow chart depicting patient selection. From 563 RA patients enrolled in the KURAMA cohort in 2015, 390 patients with available follow-up data in 2017 were selected. Next, 323 patients whose US data were available in 2015 were selected. Of the 323 patients, DAS28-CRP data in 2015 and 2017 were available in 293 patients and 81 patients with non-remission (DAS28-CRP ≥ 2.3) in 2015 were excluded. Two of the 212 patients in remission (DAS28-CRP < 2.3) in 2015 lacking the most blood test data (> 80%) were excluded. Finally, the remaining 210 patients were divided into Group 1 (patients with remission in 2017, n = 150) and Group 2 (patients with relapse in 2017, n = 60). *KURAMA* Kyoto University Rheumatoid Arthritis Management Alliance, *US* ultrasound, *DAS28* disease activity score on 28 joints, *CRP* C-reactive protein.

demographic data, laboratory findings, and imaging data) and incorporate them into prognosis prediction[14]. It can uncover useful patterns of features for prediction that would be difficult or impossible for even well-trained individuals to identify[15]. There are hundreds of studies on the application of ML in autoimmune diseases and ML generally achieved promising predictive results[16,17]. In RA, several studies already show the prediction of prognosis using ML[17–23]. However, to our knowledge, no studies have used US data. This study aims at predicting relapse in patients with RA through ML using data on US examination and blood test. Our novel prediction model may lead to a better assessment of relapse risk and enable personalized treatment for RA patients.

Results

Prediction of relapse in RA patients using US examination data, blood test data, and all data.
First, we investigated whether a combination of US examination data and blood test data improves the prediction of relapse in RA patients (n = 210) enrolled in the Kyoto University Rheumatoid Arthritis Management Alliance (KURAMA) cohort. A flow chart depicting patient selection is shown in Fig. 1. Characteristics of patients with remission (n = 150) and relapse (n = 60) are shown in Table 1. In patients with relapse, several clinical and biological markers associated with RA disease activity (disease activity score on 28 joints-CRP (DAS28-CRP), simplified disease activity index (SDAI), clinical disease activity index (CDAI), Health Assessment Questionnaire (HAQ), and patient global assessment with visual analog scale (Pt-VAS)) were significantly higher than those in patients with remission. Using 14 data on US examination, 59 data on blood test (including five data on patient information), and all data (US examination data and blood test data), predictive performance was assessed by three ML classifiers [Logistic Regression, Random Forest, and the extreme gradient boosting (XGBoost)]. Supplementary Table 1 shows the detailed list of features. Consequently, in the Random Forest and XGBoost model, area under the receiver operator characteristic curve (AUCs) calculated using all data were higher than those calculated using US or blood test data (Table 2). The highest AUC (0.677) was obtained with Random Forest using all data. These results suggest that a combination of US examination data and blood test data better predicts relapse in RA patients.

Prediction of relapse in RA patients using researcher-selected or RFE-selected features.
Next, we applied the recursive feature elimination (RFE) selection algorithm to the prediction to remove weak features and improve the prediction performance. We also selected ten features (gender, disease duration, age, wrist SMI score, MTP SMI score, ESR (1 h), CRP, RF, anti-CCP, and MMP-3) typically associated with disease activity and prognosis in RA patients and compared the results. The best Logistic Regression and Random Forest models utilized 20 and ten RFE-selected features for the best XGBoost model (Supplementary Table 2). RFE-selected features are shown in Supplementary Table 3. AUCs, accuracies, precisions, recalls, and F1-scores were higher in the prediction using RFE-selected features than that using researcher-selected features (Table 3). Among the three ML models, XGBoost showed the highest prediction result (AUC = 0.747, Fig. 2), and the AUC was also higher than the prediction using all data (Tables 2, 3). In the prediction by XGBoost, ten features, including four US examination data, five blood text data, and a piece of patient information were selected (Table 4). These results suggest that RFE-selected features are suitable for prediction in ML, compared with researcher-selected

	Median (range) or n (%)				
	Remission (n = 150)		Relapse (n = 60)		P-value
Age, years	63.8	(28.1–81.4)	66.8	(20.4–91.4)	0.18#
Female, n (%)	123	(82.0)	49	(81.7)	1.00§
Disease duration, years	7.1	(0.9–71.6)	9.9	(1.2–41.7)	0.06#
RF positive, n (%)	111	(74.0)	50	(83.3)	0.20§
Anti-CCP positive, n (%)	109	(72.7)	46	(76.7)	0.61§
CRP, mg/dl	0.1	(0.0–3.6)	0.1	(0.0–3.6)	0.51#
DAS28-CRP	1.3	(1.0–2.3)	1.6	(1.0–2.3)	0.007#
SDAI	2.5	(0.0–8.8)	3.5	(0.2–10.4)	0.0009#
CDAI	2.2	(0.0–8.8)	3.2	(0.1–9.8)	0.002#
HAQ	0.3	(0.0–1.9)	0.5	(0.0–2.5)	0.007#
Pt-VAS (mm)	10.0	(0.0–70.0)	13.0	(0.0–80.0)	0.03#
Use of glucocorticoid, n (%)	27	(18.0)	15	(25.0)	0.26§
Use of methotrexate, n (%)	116	(77.3)	34	(56.7)	0.004§
Use of biologics, n (%)	74	(49.3)	26	(43.3)	0.45§
TNF inhibitors, n (%)	45	(30.0)	15	(25.0)	0.50§
Tocilizumab, n (%)	18	(12.0)	6	(10.0)	0.81§
Abatacept, n (%)	10	(6.7)	4	(6.7)	1.00§
Tofacitinib, n (%)	1	(0.7)	1	(1.7)	0.49§

Table 1. Patient characteristics (n = 210). The groups are defined as follows: Remission: Patients with remission in both 2015 and 2017. Relapse: Patients with remission in 2015 and relapse in 2017. *RF* rheumatoid factor, *CCP* cyclic citrullinated peptide antibody, *CRP* C-reactive protein, *DAS28* disease activity score on 28 joints, *SDAI* simplified disease activity index, *CDAI* clinical disease activity index, *HAQ* Health Assessment Questionnaire, *Pt-VAS* patient global assessment with visual analog scale. #Mann–Whitney U test; §Fisher's exact test.

	AUC		
Model	US	Blood	All
Logistic Regression	0.659	0.571	0.645
Random Forest	0.621	0.615	0.677
XGBoost	0.650	0.577	0.664

Table 2. AUCs for predicting relapse in RA patients using US examination data, blood test data, or all data calculated by each model. *RA* rheumatoid arthritis, *US* ultrasound, *AUC* area under the curve.

	AUC		Accuracy		Precision		Recall		F1-Score	
Model	Researcher	RFE	Researcher	RFE	Researcher	RFE	Researcher	RFE	Researcher	RFE
Logistic Regression	0.643	0.701	0.629	0.667	0.576	0.626	0.590	0.647	0.571	0.625
Random Forest	0.658	0.719	0.695	0.729	0.619	0.691	0.562	0.595	0.552	0.596
XGBoost	0.590	0.747	0.610	0.776	0.528	0.735	0.532	0.703	0.524	0.706

Table 3. Prediction results of relapse in RA patients using researcher/RFE-selected features calculated by each model. *RA* rheumatoid arthritis, *RFE* recursive feature elimination, *AUC* area under the curve.

features. XGBoost relapse prediction shows the best performance, and all features (US, blood, and patient information) may be essential.

The importance of features selected by RFE for predicting relapse in RA patients. Finally, the importance of features selected by RFE in XGBoost was examined. Wrist and MTP SMI scores were the top two features, followed by four blood test features (lymphocyte count, ESR, platelet count, and alanine aminotransferase) (Fig. 3A). Furthermore, the ten features' value was compared between patients with remission and relapse (Fig. 3B). Consequently, wrist and MTP SMI scores were significantly higher in patients with relapse. However, alanine aminotransferase (ALT) and height were significantly lower in patients with relapse. There were no significant differences in the remaining six features. The comparison result of all features between patients with
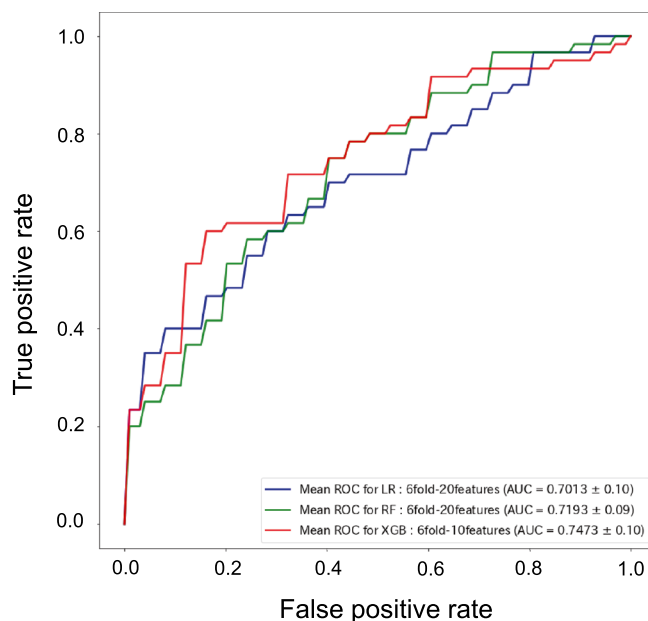
Figure 2. ROC curves for predicting relapse in RA patients. ROC curves of Logistic Regression, Random Forest, and XGBoost for predicting relapse in RA patients are shown. *ROC* receiver operating characteristics, *RA* rheumatoid arthritis.

Selected by	Features
Researcher	Gender, Disease duration, Age, Wrist SMI score, MTP SMI score, ESR (1 h), CRP, RF, anti-CCP, MMP-3
RFE	Height, Wrist SMI score, MTP SMI score, Lisfranc GS score, Cuneonavicular GS score, LYMPH, ESR (1 h), PLT, ALT, CRE

Table 4. Comparison of the researcher/RFE-selected features. In RFE-selected features, those selected in XGBoost were shown. *SMI* superb microvascular imaging, *ESR* erythrocyte sedimentation rate, *CRP* C-reactive protein, *RF* rheumatoid factor, *CCP* cyclic citrullinated peptide antibody, *MMP-3* matrix metalloproteinase-3, *RFE* recursive feature elimination, *MTP* metatarsophalangeal, *LYMPH* lymphocyte count, *PLT* platelet count, *ALT* alanine aminotransferase, *GS* gray scale, *CRE* creatinine.

remission is shown in Supplementary Table 4. To confirm and visualize the characteristics of the selected features, t-distributed Stochastic Neighbor Embedding (tSNE)[24] was applied to the standardized input data (Fig. 3C). The RFE selected features were diverse in the embedding space, and it implies the prediction of relapse is made by combining SMI scores with various features. These results suggest that all type of features, especially US data, are important for predicting relapse in RA patients. In addition to the features with significant differences between patients with remission and relapse, those with no significant differences may also contribute to the prediction.

Discussion

We studied relapse prediction in RA patients through ML using data on US examination and blood test. A combination of US examination and blood test data showed higher AUCs than those calculated using individual data. The result is not surprising because the input of more features generally improves prediction. Next, we used RFE to remove weak features and improve the prediction performance. The prediction using RFE-selected features showed higher performance than that using researcher-selected features, although the number of selected features was the same. The result suggests that RFE uncovered an optimal combination of features for better prediction. Among the ten features selected by RFE in XGBoost, wrist and MTP SMI scores were the top two vital features, suggesting that US data significantly improved prediction of relapse in RA patients. Three features (wrist and MTP SMI scores and ESR) were also included in the researcher-selected features. Wrist and MTP SMI scores were reported as prognostic factors[12,13], and ESR is one of RA's most fundamental inflammation markers[1,4]. In the remaining seven features, ALT and height were significantly lower in patients with relapse. ALT is not well-characterized as a prognostic factor, but the elevation is a marker of liver toxicity in RA treatment[25]. There is a possibility that patients with lower ALT may receive lower-intensity therapies, contributing to higher relapse risk. Height is also uncommon as a prognostic factor in RA; however, there is a study that adult height is inversely associated with disease activity[26], which is compatible with the result. There were no significant differences in six features between patients with remission and relapse. The comparison is a univariate analysis of the total cohort. Therefore, information on the association among features and prognostic significance in patient subgroups is lacking. Further studies on the importance of these features, including underlying biological mechanisms, are
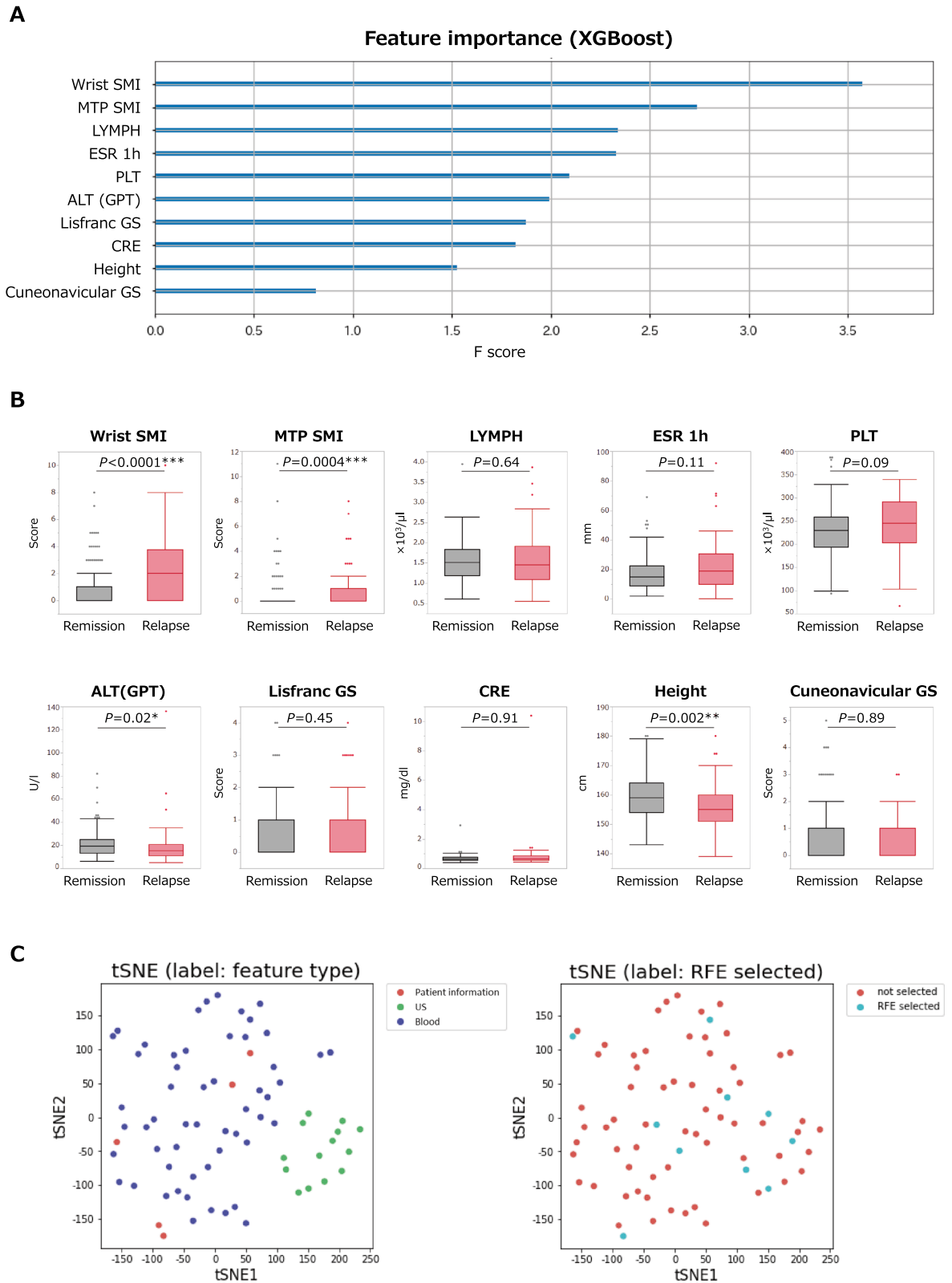
A

B

C

Figure 3. Feature importance for predicting relapse and comparison of each feature between RA patients with remission and relapse. (**A**) Importance of features for predicting relapse calculated by XGBoost model. (**B**) Comparison of each feature between RA patients with remission and relapse. (**C**) Visualization of the characteristics of the selected features in XGBoost model using tSNE. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$. *SMI* superb microvascular imaging, *MTP* metatarsophalangeal, *LYMPH* lymphocyte count, *ESR* erythrocyte sedimentation rate, *PLT* platelet count, *ALT* alanine aminotransferase, *GPT* glutamic pyruvic transaminase, *GS* gray scale, *CRE* creatinine, *US* ultrasound.

required. Among 63 features which were not selected by RFE in XGBoost, several features also had statistical difference in the value between patients with remission and relapse (Supplementary Table 4). This raises the possibility that other relative features with more importance are alternatively selected by RFE.

Among the three ML models, XGBoost, a scalable, distributed gradient-boosted decision tree ML library, achieved the best performance (AUC = 0.747). The model has gained much attention recently due to its superior performance[27,28], which is compatible with the prediction results in this study. Because the decision tree-based model is adequate for data sets containing various features, Random Forest and XGBoost showed more accuracy than Logistic Regression for mixed data. XGBoost algorithm selects one feature when there is a high correlation between variables, whereas Random Forest randomly selects a feature and learns the correlations of different features across the model. Therefore, XGBoost was considered more accurate in feature selection because it could select a smaller number and more efficient features. In our previous study analyzing almost the same cohort without using ML, the highest AUC was 0.67 for predicting relapse[12], suggesting that ML using US examination and blood test data improved prediction results. This study's sample size (n = 210) is typical among previous studies on ML applications to autoimmune diseases[17]. However, larger sample size could improve prediction. In this study, the follow-up period was 2 years, and the results may vary according to follow-up duration. Therefore, the results should be validated in studies conducted in larger populations with multiple follow-up times. Recent studies showed the possible application of ML to the measurement of US/X-ray images[29,30]. A combination of such technologies and our ML model can be a promising approach for convenient and better prediction of relapse.

In conclusion, we established an improved model for predicting relapse in RA patients through ML. The combination of data on US examination and blood test was a unique approach of this study, and US data were shown to be essential for prediction. The findings may lead to a better assessment of relapse risk and enable the selection of personalized treatment strategies for RA patients.

Methods

Patients.   From 563 RA patients enrolled in the KURAMA cohort in 2015, 390 patients with available follow-up data in 2017 were selected (Fig. 1). Next, 323 patients whose US data were available in 2015 were selected. Of the 323 patients, DAS28-CRP data in 2015 and 2017 were available in 293 patients, and 81 patients with non-remission (DAS28-CRP ≥ 2.3) in 2015 were excluded. Two of the 212 patients in remission (DAS28-CRP < 2.3) in 2015 lacking the most blood test data (> 80%) were excluded. Subsequently, the remaining 210 patients were divided into "Remission" (patients with remission in 2017, n = 150) and "Relapse" (patients with relapse in 2017, n = 60).

Data collection.   The US was examined using an Aplio500 (Canon Medical Systems) fitted with a 12 MHz linear probe (18L7). Bilateral joints (second through fifth metacarpophalangeal (MCP), radial wrist, ulnar wrist, second through fifth MTP, Lisfranc, cuneonavicular, Chopart, and ankle) were examined as described previously[31]. The scanning technique and interpretation of lesions were based on Outcomes Measures in Rheumatology (OMERACT)[32]. The former of the two SMI modes (color-coded and monochrome SMI) was used for this study. Regions of interest for SMI were fixed at the same size and depth for each joint type. Under the established four-point scale (0–3) semi-quantitative scoring system[33], gray scale (GS) and SMI scores were determined on-site by at least two of five sonographers with 1–9 years of experience, and agreement was obtained in weekly meetings attended by all five sonographers. The scores for each group of joints were summed as follows: MCP, bilateral second through fifth MCP; wrist, bilateral radial, and ulnar joints; MTP, bilateral second through fifth MTP; Lisfranc, bilateral Lisfranc joints; Cuneonavicular, bilateral Cuneonavicular joints; Chopart, bilateral Chopart joints; ankle, bilateral ankle joints. There were no missing data in the US examination.

Patient information and blood test data in 2015 were also collected from the KURAMA cohort. Supplementary Table 1 shows the list of features. Cases with more than 80% of the missing features were eliminated. Missing values were complemented with each feature's median value. For replacing missing values on patients' height and weight, median values were calculated by gender. In total, 14 US examination data, 54 blood test data, and five data on patient information were available for analysis. For convenience, data on patient information were included in blood test data in this study.

Prediction models.   Three ML classifiers (Logistic Regression, Random Forest, and XGBoost) were employed to predict RA patients' relapse. The logistic regression model is a generalized linear model and traditional approach for binary classification on clinical prediction. Random Forest is an ensemble algorithm that combines multiple decision trees to build a robust model[34]. It is widely used because of its high interpretability of prediction results. XGBoost is also a decision tree-based ensemble algorithm and achieves more accurate prediction utilizing gradient boosting[35].

Predictive performance was assessed using the mean AUC by nested stratified six-fold cross-validation (CV). The inner loop, consisting of a three-fold CV, was used to select hyper-parameters by grid-search. The class balance option was set for all models to deal with imbalanced data.

For feature selection, we employed RFE, a method for extracting subsets of features that contribute to prediction performance by recursive processing. Since RFE allows us to set the size of the final feature subset, we varied the value within [5, 10, 20, 30, 50], finally selecting the number of features that showed the best AUC. Analyses and model constructions were performed with Python 3.8 packages (Scikit-learn 0.23 and XGBoost 1.1.1).

Ethical statements.   This study was conducted following the principles set down in the Declaration of Helsinki and was approved by the ethics committee of Kyoto University (R0357). All patients provided written informed consent.

References

1. Smolen, J. S., Aletaha, D. & McInnes, I. B. Rheumatoid arthritis. *Lancet* **388**, 2023–2038 (2016).
2. Goekoop-Ruiterman, Y. P. & Huizinga, T. W. Rheumatoid arthritis: Can we achieve true drug-free remission in patients with RA? *Nat. Rev. Rheumatol.* **6**, 68–70 (2010).
3. Aga, A. B. *et al.* Time trends in disease activity, response and remission rates in rheumatoid arthritis during the past decade: Results from the NOR-DMARD study 2000–2010. *Ann. Rheum. Dis.* **74**, 381–388 (2015).
4. van der Helm-van Mil, A. H. Risk estimation in rheumatoid arthritis: From bench to bedside. *Nat. Rev. Rheumatol.* **10**, 171–180 (2014).
5. Ohrndorf, S. & Backhaus, M. Advances in sonographic scoring of rheumatoid arthritis. *Ann. Rheum. Dis.* **72**, ii69–ii75 (2013).
6. Scirè, C. A. *et al.* Ultrasonographic evaluation of joint involvement in early rheumatoid arthritis in clinical remission: Power Doppler signal predicts short-term relapse. *Rheumatology (Oxford)* **48**, 1092–1097 (2009).
7. Peluso, G. *et al.* Clinical and ultrasonographic remission determines different chances of relapse in early and long standing rheumatoid arthritis. *Ann. Rheum. Dis.* **70**, 172–175 (2011).
8. Foltz, V. *et al.* Power Doppler ultrasound, but not low-field magnetic resonance imaging, predicts relapse and radiographic disease progression in rheumatoid arthritis patients with low levels of disease activity. *Arthritis Rheum.* **64**, 67–76 (2012).
9. Iwamoto, T. *et al.* Prediction of relapse after discontinuation of biologic agents by ultrasonographic assessment in patients with rheumatoid arthritis in clinical remission: High predictive values of total gray-scale and power Doppler scores that represent residual synovial inflammation before discontinuation. *Arthritis Care Res.* **66**, 1576–1581 (2014).
10. Nguyen, H. *et al.* Prevalence of ultrasound-detected residual synovitis and risk of relapse and structural progression in rheumatoid arthritis patients in clinical remission: A systematic review and meta-analysis. *Rheumatology (Oxford)* **53**, 2110–2118 (2014).
11. Kawashiri, S. Y. *et al.* Ultrasound-detected bone erosion is a relapse risk factor after discontinuation of biologic disease-modifying antirheumatic drugs in patients with rheumatoid arthritis whose ultrasound power Doppler synovitis activity and clinical disease activity are well controlled. *Arthritis Res. Ther.* **19**, 108 (2017).
12. Matsuo, H. *et al.* Prediction of recurrence and remission using superb microvascular imaging in rheumatoid arthritis. *J. Med. Ultrason. (2001)* **47**, 131–138 (2020).
13. Matsuo, H. *et al.* Positive rate and prognostic significance of the superb microvascular imaging signal in joints of rheumatoid arthritis patients in remission with normal C-reactive protein levels and erythrocyte sedimentation rates. *J. Med. Ultrason. (2001)* **48**, 353–359 (2021).
14. Ngiam, K. Y. & Khor, I. W. Big data and machine learning algorithms for health-care delivery. *Lancet Oncol.* **20**, e262–e273 (2019).
15. Goecks, J., Jalili, V., Heiser, L. M. & Gray, J. W. How machine learning will transform biomedicine. *Cell* **181**, 92–101 (2020).
16. Kingsmore, K. M., Puglisi, C. E., Grammer, A. C. & Lipsky, P. E. An introduction to machine learning and analysis of its use in rheumatic diseases. *Nat. Rev. Rheumatol.* **17**, 710–730 (2021).
17. Stafford, I. S. *et al.* A systematic review of the applications of artificial intelligence and machine learning in autoimmune diseases. *NPJ Digit. Med.* **3**, 30 (2020).
18. Luque-Tévar, M. *et al.* Integrative clinical, molecular, and computational analysis identify novel biomarkers and differential profiles of anti-TNF response in rheumatoid arthritis. *Front. Immunol.* **12**, 631662 (2021).
19. Kalweit, M. *et al.* Personalized prediction of disease activity in patients with rheumatoid arthritis using an adaptive deep neural network. *PLoS One* **16**, e0252289 (2021).
20. Yoosuf, N. *et al.* Early prediction of clinical response to anti-TNF treatment using multi-omics and machine learning in rheumatoid arthritis. *Rheumatology (Oxford)* https://doi.org/10.1093/rheumatology/keab521 (2021).
21. Vodencarevic, A. *et al.* Advanced machine learning for predicting individual risk of flares in rheumatoid arthritis patients tapering biologic drugs. *Arthritis Res. Ther.* **23**, 67 (2021).
22. Koo, B. S. *et al.* Machine learning model for identifying important clinical features for predicting remission in patients with rheumatoid arthritis treated with biologics. *Arthritis Res. Ther.* **23**, 178 (2021).
23. Johansson, F. D. *et al.* Predicting response to tocilizumab monotherapy in rheumatoid arthritis: A real-world data analysis using machine learning. *J. Rheumatol.* **48**, 1364–1370 (2021).
24. van der Maaten, L. J. P. & Hinton, G. E. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
25. Karlsson Sundbaum, J. *et al.* Methotrexate treatment in rheumatoid arthritis and elevated liver enzymes: A long-term follow-up of predictors, surveillance, and outcome in clinical practice. *Int. J. Rheum. Dis.* **22**, 1226–1232 (2019).
26. Chen, Y., Yu, Z., Packham, J. C. & Mattey, D. L. Influence of adult height on rheumatoid arthritis: Association with disease activity, impairment of joint function and overall disability. *PLoS One* **8**, e64862 (2013).
27. Zhao, Y. *et al.* Ensemble learning predicts multiple sclerosis disease course in the SUMMIT study. *NPJ Digit. Med.* **3**, 135 (2020).
28. Morid, M. A., Lau, M. & Del Fiol, G. Predictive analytics for step-up therapy: Supervised or semi-supervised learning?. *J. Biomed. Inform.* **119**, 103842 (2021).
29. Fiorentino, M. C. *et al.* A deep-learning framework for metacarpal-head cartilage-thickness estimation in ultrasound rheumatological images. *Comput. Biol. Med.* **141**, 105117 (2022).
30. Rohrbach, J., Reinhard, T., Sick, T. & Dürr, O. Bone erosion scoring for rheumatoid arthritis with deep convolutional neural networks. *Comput. Electr. Eng.* **78**, 472–481 (2019).
31. Naredo, E. *et al.* Ultrasound joint inflammation in rheumatoid arthritis in clinical remission: How many and which joints should be assessed?. *Arthritis Care Res. (Hoboken)* **65**, 512–517 (2013).
32. Backhaus, M. *et al.* Guidelines for musculoskeletal ultrasound in rheumatology. *Ann. Rheum. Dis.* **60**, 641–649 (2001).
33. Szkudlarek, M. *et al.* Interobserver agreement in ultrasonography of the finger and toe joints in rheumatoid arthritis. *Arthritis Rheum.* **48**, 955–962 (2003).
34. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
35. Chen, T. & Carlos, G. XGBoost: A Scalable Tree Boosting System. *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 785–794. https://doi.org/10.1145/2939672.2939785 (2016).

Acknowledgements

Author contributions

H.M. and M.K. designed the research, analyzed the data, and wrote the manuscript. A.I. helped with data analysis. M.K., A.I., M.S., M.I., and Y.T collected the US data. M.H., M.T., H.I., and Y.F. supervised the research and commented on the research direction. All authors reviewed the manuscript.

Competing interests

The department in which M.H. and H.I. retired from and M.T. currently works received financial support from Nagahama city, Toyooka city, and five pharmaceutical companies (Mitsubishi Tanabe Pharma, Chugai Pharmaceutical, UCB Japan, Ayumi, and Asahi Kasei Pharma). M.H. receives research grants and/or speaker fees from Bristol-Myers, Eisai, Eli Lilly, Mitsubishi Tanabe, and Novartis Pharma. H.I. receives research grants from Bristol-Myers, Eisai, Mochida, and Taisho. The KURAMA cohort study is supported by a grant from Daiichi-Sankyo. M.T. has received research grants and/or speaker fees from AbbVie GK, Asahi Kasei Pharma Corp., Astellas Pharma Inc., Ayumi Pharmaceutical Corp., Bristol-Myers Squibb, Chugai Pharmaceutical Co., Ltd., Eisai Co., Ltd., Eli Lilly Japan K.K., Pfizer Inc., UCB Japan Co., Ltd., Janssen Pharmaceutical K.K., Mitsubishi Tanabe Pharma Corp., Novartis Pharma K.K., Taisho Pharma Co., Ltd. The present study was investigator-initiated. These companies played no role in study design, data collection or analysis, writing of the manuscript, or the decision to submit for publication. H.M., M.K., A.I., M.S., M.I., Y.T, and Y.F. declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-11361-y.

Correspondence and requests for materials should be addressed to H.M.

Reprints and permissions information is available at www.nature.com/reprints.

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