



## OPEN ACCESS

## EDITED BY

Javier Rodríguez-Carrio,  
University of Oviedo, Spain

## REVIEWED BY

Matsumoto Kotaro,  
Kurume University, Japan  
Diego Benavent,  
University Hospital La Paz, Spain  
A. H. Zwiderman,  
Academic Medical Center, Netherlands

## \*CORRESPONDENCE

Johannes Knitza  
johannes.knitza@uk-erlangen.de

†These authors have contributed  
equally to this work and share first  
authorship

‡These authors have contributed  
equally to this work and share last  
authorship

## SPECIALTY SECTION

This article was submitted to  
Rheumatology,  
a section of the journal  
Frontiers in Medicine

RECEIVED 26 May 2022

ACCEPTED 30 June 2022

PUBLISHED 22 July 2022

## CITATION

Knitza J, Janousek L, Kluge F,  
von der Decken CB, Kleinert S,  
Vorbrüggen W, Kleyer A, Simon D,  
Hueber AJ, Muehlensiepen F,  
Vuillerme N, Schett G, Eskofier BM,  
Welcker M and Bartz-Bazzanella P  
(2022) Machine learning-based  
improvement of an online  
rheumatology referral and triage  
system.

*Front. Med.* 9:954056.

doi: 10.3389/fmed.2022.954056

## COPYRIGHT

© 2022 Knitza, Janousek, Kluge, von  
der Decken, Kleinert, Vorbrüggen,  
Kleyer, Simon, Hueber, Muehlensiepen,  
Vuillerme, Schett, Eskofier, Welcker  
and Bartz-Bazzanella. This is an  
open-access article distributed under  
the terms of the [Creative Commons  
Attribution License \(CC BY\)](#). The use,  
distribution or reproduction in other  
forums is permitted, provided the  
original author(s) and the copyright  
owner(s) are credited and that the  
original publication in this journal is  
cited, in accordance with accepted  
academic practice. No use, distribution  
or reproduction is permitted which  
does not comply with these terms.

# Machine learning-based improvement of an online rheumatology referral and triage system

Johannes Knitza<sup>1,2,3\*†</sup>, Lena Janousek<sup>4†</sup>, Felix Kluge<sup>4†</sup>,  
Cay Benedikt von der Decken<sup>5,6,7</sup>, Stefan Kleinert<sup>7,8,9</sup>,  
Wolfgang Vorbrüggen<sup>7,10</sup>, Arnd Kleyer<sup>1,2</sup>, David Simon<sup>1,2</sup>,  
Axel J. Hueber<sup>1,11</sup>, Felix Muehlensiepen<sup>3,12</sup>,  
Nicolas Vuillerme<sup>3,13,14</sup>, Georg Schett<sup>1,2</sup>, Bjoern M. Eskofier<sup>4</sup>,  
Martin Welcker<sup>7,15‡</sup> and Peter Bartz-Bazzanella<sup>6,7‡</sup>

<sup>1</sup>Department of Internal Medicine 3, Friedrich-Alexander-University Erlangen-Nürnberg and Universitätsklinikum Erlangen, Erlangen, Germany, <sup>2</sup>Deutsches Zentrum für Immuntherapie (DZI), Friedrich-Alexander-University Erlangen-Nürnberg and Universitätsklinikum Erlangen, Erlangen, Germany, <sup>3</sup>Université Grenoble Alpes, AGEIS, Grenoble, France, <sup>4</sup>Machine Learning and Data Analytics Lab, Department of Artificial Intelligence in Biomedical Engineering (AIBE), Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany, <sup>5</sup>Medizinisches Versorgungszentrum Stolberg, Stolberg, Germany, <sup>6</sup>Klinik für Internistische Rheumatologie, Rhein-Maas-Klinikum, Würselen, Germany, <sup>7</sup>RheumaDatenRhePort (rhadar), Planegg, Germany, <sup>8</sup>Praxisgemeinschaft Rheumatologie-Nephrologie, Erlangen, Germany, <sup>9</sup>Medizinische Klinik 3, Rheumatology/Immunology, Universitätsklinikum Würzburg, Würzburg, Germany, <sup>10</sup>Verein zur Förderung der Rheumatologie e.V., Würselen, Germany, <sup>11</sup>Division of Rheumatology, Klinikum Nürnberg, Paracelsus Medical University, Nürnberg, Germany, <sup>12</sup>Faculty of Health Sciences, Center for Health Services Research, Brandenburg Medical School Theodor Fontane, Rüdersdorf, Germany, <sup>13</sup>Institut Universitaire de France, Paris, France, <sup>14</sup>LabCom Telecom4Health, Orange Labs and Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP-UGA, Grenoble, France, <sup>15</sup>MVZ für Rheumatologie Dr. Martin Welcker GmbH, Planegg, Germany

**Introduction:** Rheport is an online rheumatology referral system allowing automatic appointment triaging of new rheumatology patient referrals according to the respective probability of an inflammatory rheumatic disease (IRD). Previous research reported that Rheport was well accepted among IRD patients. Its accuracy was, however, limited, currently being based on an expert-based weighted sum score. This study aimed to evaluate whether machine learning (ML) models could improve this limited accuracy.

**Materials and methods:** Data from a national rheumatology registry (RHADAR) was used to train and test nine different ML models to correctly classify IRD patients. Diagnostic performance was compared of ML models and the current algorithm was compared using the area under the receiver operating curve (AUROC). Feature importance was investigated using shapley additive explanation (SHAP).

**Results:** A complete data set of 2265 patients was used to train and test ML models. 30.5% of patients were diagnosed with an IRD, 69.3% were female. The diagnostic accuracy of the current Rheport algorithm (AUROC of 0.534) could be improved with all ML models, (AUROC ranging between

0.630 and 0.737). Targeting a sensitivity of 90%, the logistic regression model could double current specificity (17% vs. 33%). Finger joint pain, inflammatory marker levels, psoriasis, symptom duration and female sex were the five most important features of the best performing logistic regression model for IRD classification.

**Conclusion:** In summary, ML could improve the accuracy of a currently used rheumatology online referral system. Including further laboratory parameters and enabling individual feature importance adaption could increase accuracy and lead to broader usage.

#### KEYWORDS

artificial intelligence, machine learning, rheumatology, triage, symptom checker, digital health, decision support system (DSS)

## Introduction

Rheumatology services are facing an increasing demand of referrals while the number of rheumatologists is steadily declining (1). This shortage of rheumatologists results in a long delay for patients from symptom onset to rheumatology appointment, diagnosis and start of therapy. To reduce irreversible damage caused by uncontrolled inflammatory disease, rheumatologists have to triage patients. Despite various triage and screening strategies (2), the large majority of patients referred to rheumatologists end up not having an inflammatory rheumatic disease (IRD) (3, 4). In contrast to emergency medicine (5), no objective, rheumatology triage criteria exists and this lack of standardized triage decisions hampers quality of care. The European Alliance of Associations for Rheumatology (EULAR) recently emphasized the added value of patient preassessment by telehealth to improve the referral process to rheumatology and help prioritization of people with suspected IRD (6).

Rheport is an online rheumatology referral system, currently used in Germany to automatically triage appointments of new rheumatology patient referrals according to the respective probability of an IRD (7). An objective, weighted sum score is used to calculate the individual IRD probability. The tool can be used by patients themselves or referring physicians. Recent studies showed that Rheport was well accepted and perceived as easy to use by patients (8), however, the diagnostic accuracy was limited (4). Machine learning has been successfully used in various disciplines to increase diagnostic accuracy (9–11). In rheumatology, expert-level performance has recently been achieved using deep learning for detection of radiographic sacroiliitis (12) and individual risk of disease flares could be predicted in patients with

rheumatoid arthritis using advanced machine learning (10). To our knowledge, no study has investigated the potential of machine learning to improve the triage of patients referred to rheumatology centers.

The aim of this study therefore was to evaluate whether machine learning could improve the triage accuracy (detection of inflammatory rheumatic diseases) of the online rheumatology self-referral system Rheport.

## Materials and methods

### Rheport questionnaire

Rheport<sup>1</sup> was implemented at seven German rheumatology centers to automatically triage appointments of new rheumatology patient referrals according to the respective probability of an inflammatory rheumatic disease (IRD). The fixed Rheport patient questionnaire used in this study (version 1.0), consists of 23 questions, including basic health information, typical rheumatic symptoms, and laboratory parameters [C-reactive protein (CRP), erythrocyte sedimentation rate (ESR)]. Every question is assigned a weight in percentage and every possible answer is assigned a separate factor. All weights and factors are based on expert knowledge from rheumatologists. The sub-score of a question is calculated by multiplying the given answers factor by the weight of the question. Adding up all sub-scores of the 23 questions results in the total score of the patient. An example of a completed questionnaire is presented in **Table 1**.

<sup>1</sup> [www.rheport.de](http://www.rheport.de)

TABLE 1 Example of a questionnaire summary report including the respective score calculation.

Question	Answers	Factor	Weight	Sub-score
Gender	Female	0	0%	0.00
Age	<60 years	0	2%	0.00
Weight loss	No	0	3%	0.00
Duration of complaints	6 months, <12 months	3	5%	0.15
Preceding injury	No	5	1%	0.05
Preceding infection	No	0	1%	0.00
Preceding tick sting	No	0	0%	0.00
Referring physician	General practitioner	0	1%	0.00
Lab results	No lab results available	0	10%	0.00
Family history	No	0	0%	0.00
Joint pain	With movement	1	10%	0.10
Joint swelling	Big toe	10	10%	1.00
Finger swelling	Whole finger	10	4.5%	0.45
Duration of joint swelling	>6 weeks, <half a year	10	2.5%	0.25
Joint stiffness	All day long	1	7.5%	0.075
Headache	Neck and back of head	1	7.5%	0.075
Lower back pain	No	0	7.5%	0.00
Other pain	No	0	7.5%	0.00
Pain related limitation of movement	No	0	2.5%	0.00
Muscle weakness	No	0	10%	0.00
General symptoms	Disrupted sleep, often tired	0	2.5%	0.00
Other symptoms	Fever > 38°C	4	2.5%	0.10
Comorbidities	Psoriasis	4	2.5%	0.10
	Total score			2.35

TABLE 2 Rheport's triage levels, respective total score thresholds and appointment time frame.

Triage levels	Total score thresholds	Appointment time frame
1 Very urgent	>4.0	Within 1 week
2 Urgent	2.4–4.0	Within 2 weeks
3 Intermediate	1.0–2.4	Within 1 month
4 IRD unlikely	<1	Transfer (back) to GP

Rheport consists of 4 different triage levels, based on total score thresholds (Table 2). Total scores lower 1 are classified as unlikely to have an IRD (non-IRD) and these patients are transferred (back) to their treating general physician. Patients with a minimum total score of 1 may book an appointment, being classified as IRD, at a participating rheumatology center. With increasing total scores patients have access to earlier appointments. The cut-offs were established after reviewing scores of the first 255 consecutive patients. IRD patients with scores < 1 (8/255) presented with very mild symptoms and later presentation would not have led to a decisive deterioration in prognosis in these case. Based on this cut-off, Rheport had a negative

predictive value (NPV) of 86.4% and a positive predictive value (PPV) of 34.2%.

## Rheport dataset and data pre-processing

Prior to completing the questionnaire, users need to register and actively consent that their data is uploaded pseudonymized to the RheumaDatenRhePort (RHADAR) registry (7). RHADAR is a German real-world rheumatology registry including adult patients. Participating rheumatologists are encouraged to add their final diagnosis to the RHADAR registry. We only included fully completed questionnaires with matched added final diagnosis by the treating rheumatologist. Patients having consented until August 17th, 2020, were included in this analysis. Descriptive statistical analyses were carried out using to describe gender, age, inflammatory marker-, and family history-status, based on patient questionnaires and final diagnosis according to treating rheumatologist (Table 3).

Most ML-models work with numerical data only (13). All the features in the dataset were non-numerical and had to be transcribed. For ordinal features (CRP, ESR, duration of joint swelling and joint complaints) a value was assigned for

TABLE 3 Diagnostic categories.

**Diagnostic categories n, (%)**

Inflammatory rheumatic disease	690 (30.5)
Rheumatoid arthritis	339 (15.0)
Psoriatic arthritis	103 (4.5)
Polymyalgia rheumatica	84 (3.7)
Axial spondyloarthritis	56 (2.5)
Undifferentiated arthritis	44 (1.9)
Reactive arthritis	21 (0.9)
Systemic lupus erythematosus	18 (0.8)
Crystal arthropathies	12 (0.5)
Systemic sclerosis	5 (0.2)
Inflammatory idiopathic myositis	3 (0.1)
Behcet's disease	2 (0.1)
ANCA-associated vasculitis	2 (0.1)
Giant cell arteritis	1 (0.04)
Non-inflammatory rheumatic disease	1,575 (69.5)

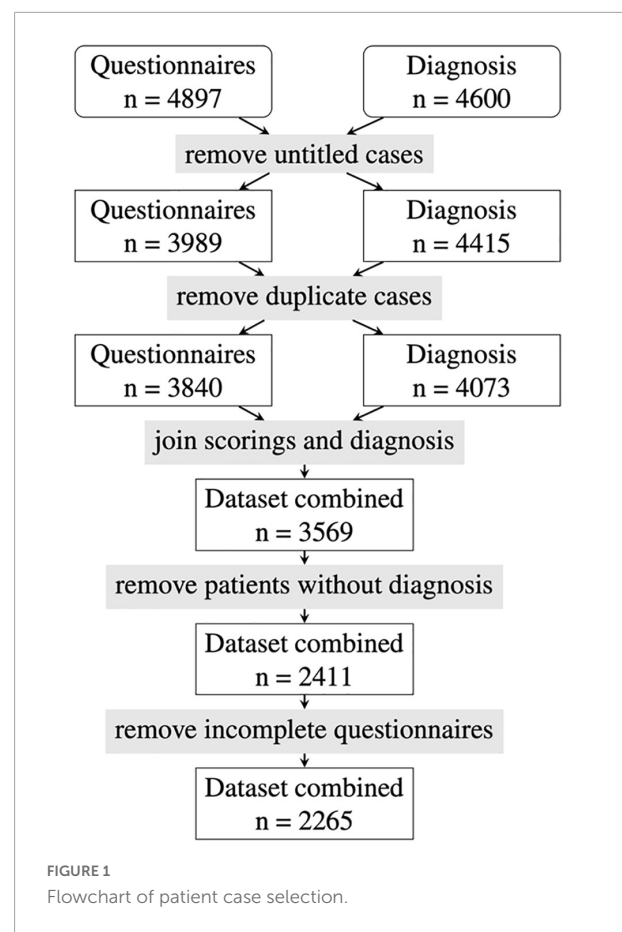
every category. For the remaining nominal features one-hot-encoding was used. For every category of the feature, a new dummy column was created. The entries in these columns were either 0 or 1.

To address the unequal IRD distribution in the dataset [and general population referred to rheumatologists (3, 4)], we used the Python library imbalanced-learn (14). This library provides different methods for over- and undersampling during training and testing of the ML models, such as SMOTE (15).

The dataset was used for training nine different ML-models, applying the Python library scikit-learn (16) including: K-nearest neighbor, decision tree, support vector machine, neural network, logistic regression, random forest, bagging classifier, gradient boosting, and AdaBoost.

## Model training, testing, and feature importance analysis

Using a validation dataset and setting it aside to test the model after training, results in less training data. Cross-validation avoids this problem and we therefore used a threefold cross-validation approach, where 2 parts acted as the training data while 1 part was used for testing. This was done 3 times, so that every part of the data once acted as the validation data. The prediction error is the average of the 3 computed errors. Cross-validation is also used for hyper-parameter tuning. A combination of both is called nested cross-validation. Without nested cross-validation, the same data that would be used for tuning the hyper-parameters and to estimate the model performance, which might result in overfitting. To avoid this, nested cross-validation uses an inner loop for tuning the hyper-parameters and an outer loop for



generating the prediction error. The best performing model was selected based on the mean area under the receiver operating characteristics curve (AUC).

Diagnostic accuracy for the current algorithm and best performing model was further evaluated referring to sensitivity, specificity, NPV. Different clinical target sensitivities were explored (90 and 95%). For the best performing ML model, feature importance was investigated using shapley additive explanation (SHAP) (17, 18). This analysis indicates to which extent and in which direction (pro IRD vs. against IRD) a certain feature influences the ML model.

We used the TRIPOD checklist (19) (Supplementary Material 1) to enable transparent reporting of our study results.

## Results

### Patient dataset

A dataset of 4,897 Rheport questionnaires from seven German rheumatology centers with verified diagnosis was available. After removing invalid data (test vignettes, untitled cases) and duplicate data, another 1,158 patients with missing

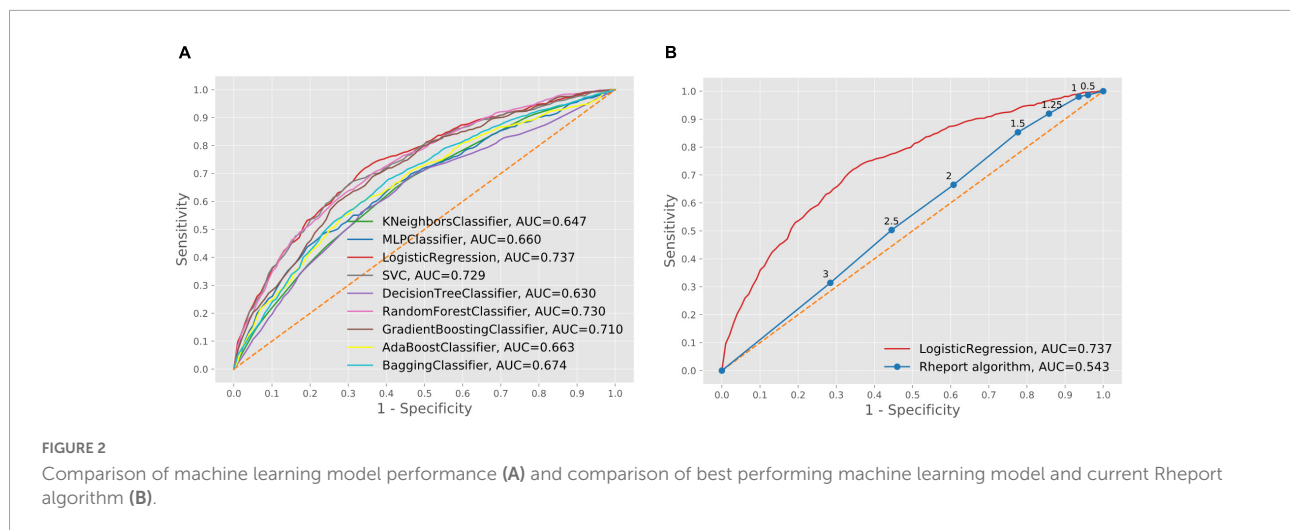


FIGURE 2

Comparison of machine learning model performance (A) and comparison of best performing machine learning model and current Rheport algorithm (B).

TABLE 4 Specificity, PPV, NPV of current Rheport algorithm and best performing machine learning model according to targeted sensitivity.

Targeted sensitivity (%)	Current Rheport algorithm			Logistic regression		
	Specificity (%)	PPV (%)	NPV (%)	Specificity (%)	PPV (%)	NPV (%)
95	10	33	81	20	34	89
90	17	34	78	33	37	88
70	36	34	72	67	48	84

PPV, positive predictive value; NPV, negative predictive value.

final diagnosis and 146 incomplete questionnaires had to be removed. The final dataset consisted of 2,265 patients (see Figure 1).

Table 3 displays the final diagnostic categories according to the treating rheumatologists. 690/2,265 (30.5%) of the patients were diagnosed with an IRD, the most common IRD being rheumatoid arthritis 339/690 (49.1%). Self-reported patient characteristics according to the Rheport questionnaire are listed in Supplementary Material 2. The majority of patients were female (69.3%) and younger than 60 years of age (76.9%). Only a minority reported that symptoms started less than 6 weeks ago (9.0%). The majority of patients had previously seen a general physician (85.1%). Only a minority of patients 839/2,265 (37.0%) entered lab results (CRP, ESG) completing the Rheport questionnaire. The large majority of patients complained of being tired (80.0%). The most commonly cited comorbidities were osteoarthritis (37.2%) and obesity (31.5%).

## Performance of machine learning models

The current Rheport algorithm (AUROC of 0.534) could be improved with all ML models, starting with decision tree (AUROC of 0.630) to the three best performing models, including support vector machine (AUROC of 0.729), random

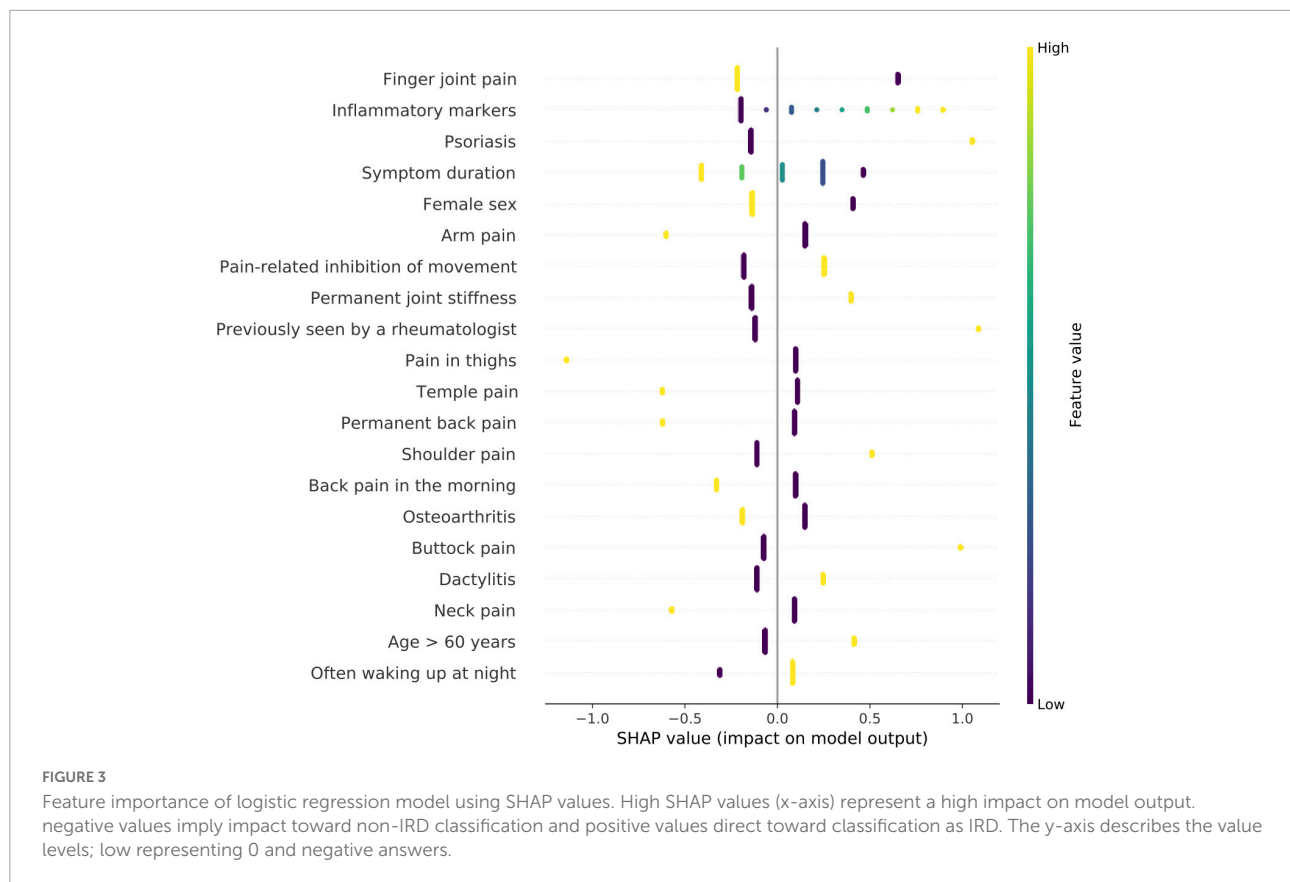
forest (AUROC of 0.730) and logistic regression (AUC of 0.737) (Figure 2). Targeting high sensitivity values of 90 and 95% as a screening test, using the logistic regression model could double specificity (17% vs. 33% and 10% vs. 20%), respectively, resulting in negative predictive values of 78% vs. 88% and 81% vs. 89%, respectively (Table 4).

## Feature importance

The feature analysis included a total of 111 Rheport questionnaire features. Figure 3 lists the 20 most important features for the best performing model (logistic regression). The SHAP analysis revealed finger joint pain, followed by inflammatory markers, psoriasis (as an underlying comorbidity), symptom duration and female sex as the five most important features. Reported finger joint pain directed the model toward Non-IRD classification, whereas underlying psoriasis and elevated inflammatory markers were top features directing toward IRD.

## Discussion

To our knowledge, we could show for the first time, that machine learning can improve the diagnostic



accuracy of an online rheumatology referral system. The best ML model, namely logistic regression, promisingly improved the current Rheport algorithm (AUC of 0.737 vs. 0.543).

The subjectivity of current rheumatology triage decisions was recently highlighted by a study reporting an increase of inappropriate rheumatology referrals of 14.3% when a new rheumatologist was hired (20). The importance of pre-appointment management to improve efficacy of rheumatology consultations has been reported already by Harrington and Walsh (21). Having to make all prior patient records available for granting an appointment was key for a successful implementation. However, currently used hand written referral letters often lack this crucial information. Wong et al. recently reported that only 55% of referral letters included medical history, 51% laboratory parameters, and 34% imaging reports (22), depriving rheumatologists of crucial information. Deprived of information on prior external imaging and laboratory workup, even experienced rheumatologists are not able to correctly classify patients as IRD or non-IRD in the majority of cases (23). The most impactful features of the logistic regression model also highlight the need for having access to prior medical records and complete patient information, including current symptoms and basic laboratory parameters. Digitalization can

significantly reduce the burden of making this information available prior to an appointment. In our opinion, digitally supported online rheumatology referrals could standardize and significantly improve triage decisions. The ability to communicate with referring physicians and to give them feedback could further improve referrals (21) and allow case resolutions after e-consultations, saving face-to-face visits of up to 20% (24). Displaying top feature importance using heat maps additional to the final Rheport score could effectively support rheumatologists in making an informed triage decision. Similarly, rheumatologists might be reluctant to implement a triage software, if feature importance cannot be adapted according the specific center.

Importantly the feature analysis revealed some counterintuitive differences. Whereas increased inflammatory markers lead to an IRD classification, presence of finger joint pain lead to a non-IRD classification. One reason for this could be high proportion of patients with osteoarthritis (37.2%).

To improve the performance of Rheport, we believe that adding mandatory laboratory parameters and imaging results, similar to what rheumatologists need to make a correct decision (23), is vital. Solely relying on patient-reported symptoms, the accuracy of symptom checkers

is relatively poor regarding correct IRD detection, with a sensitivity ranging between 14% (25), 19% (26, 27), and 54% (4). In a first analysis, comparing different symptom checker in rheumatology, including 34 patients, Powley et al. (27) showed only 19% patients with “inflammatory arthritis” were given the diagnosis of RA or PsA. Similarly, Proft et al. demonstrated, that using an online symptom checker-based self-referral tool for axial spondyloarthritis patients resulted in a correctly identified proportion of 19.4% (26). In a previous single-center study (4), we compared Rheport to an artificial intelligence-based symptom checker Ada, resulting in a similarly limited sensitivity and specificity of 53.7, 51.8, and 53.7 and 63.6%, respectively.

A strength of the study is the real-world and multicenter nature and size of the dataset used. A strength of Rheport is that only rheumatologists can add their final diagnosis. This represents extra work for the participating rheumatologists but adds high quality data, enabling a continuous improvement of the system (7). This study has some limitations. The majority of patients did not report any laboratory results. This might have influenced the results, especially regarding the high feature importance. Schneider et al. recently reported the added value of including laboratory parameters for machine learning based classification of inflammatory bowel disease in children (9). The high feature importance of inflammatory markers suggests that adding additional laboratory data such as antibody status could improve the accuracy of Rheport. As patients seem eager to use self-sampling (28, 29), future research should investigate the integration of self-sampling into online symptom checkers/self-referral tools. In a next study we will analyze whether it makes a difference if the information is entered by patients themselves or referring physicians. Despite the multicenter nature, the results are limited to one country and further validation studies are needed. Especially for more rare IRDs, also only scarcely included in the dataset, the accuracy is likely to be worse. New patient data and matching final diagnoses enable perpetual improvement of the algorithm. Implementing machine learning predictive models into clinical routine raises additional questions such as trust and liability.

## Conclusion

This study suggests that machine learning models can improve the diagnostic accuracy of online self-referral systems, based on patient reported data. Current accuracy is limited by the predominantly patient-reported subjective data. Routine collection of electronic patient-reported data of newly referred patients could enable an improved and standardized rheumatology triage strategy.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

## Author contributions

JK, FK, PB-B, and MW: conceptualization. JK, FK, and LJ: methodology and formal analysis. JK, LJ, and CD: data curation. LJ and JK: writing—original draft preparation. JK, FK, CD, SK, WV, AK, DS, AH, FM, NV, GS, BE, MW, and PB-B: writing—review and editing. LJ: visualization. GS, NV, BE, PB-B, and MW: supervision. MW: funding acquisition. All authors have read and agreed to the published version of the manuscript.

## Funding

This study was funded by the RHADAR GbR (A Network of Rheumatologists), Bahnhofstr. 32, 82152 Planegg, Germany. RHADAR GbR received honoraria from UCB Pharma GmbH, Sandoz Deutschland/Hexal AG, Lilly GmbH, and Galapagos Biopharma Germany GmbH, and research support from Novartis Pharma GmbH. The funders were not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication. This work was supported by the Deutsche Forschungsgemeinschaft (DFG—FOR 2886 PANDORA—Z01/Z01/C1/A03 to JK, AK, DS, and GS and DFG—Heisenberg ES 434/8-1 to BE).

## Acknowledgments

We thank all the patients, and especially Nils Körber (by Nils Körber and Joachim Elgas GbR, Erlangen) who developed RheumaDok and the RHADAR database. We would also like to thank all the rheumatological specialist assistants supporting the clinical documentation in the participating centers. The present work is part of the PhD thesis of the first author JK (AGEIS, Université Grenoble Alpes, Grenoble, France).

## Conflict of interest

Qinum and RheumaDatenRhePort developed and hold rights for Rheport. WV, CD, and PB-B were involved in the development of Rheport. JK was a member of the scientific board of RheumaDatenRhePort. WV, CD, SK, PB-B, and MW were members of RheumaDatenRhePort GbR. RHADAR GbR received honoraria from UCB Pharma GmbH, Sandoz Deutschland/Hexal AG, Lilly GmbH, and Galapagos Biopharma Germany GmbH, and research support from Novartis Pharma GmbH.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

- Krusche M, Sewerin P, Kleyer A, Mucke J, Vossen D, Morf H. Rheumadocs und arbeitskreis junge rheumatologie (AGJR). [Specialist training quo vadis?]. *Z Rheumatol.* (2019) 78:692–7. doi: 10.1007/s00393-019-00690-5
- Benesova K, Lorenz H-M, Lion V, Voigt A, Krause A, Sander O, et al. [Early recognition and screening consultation: a necessary way to improve early detection and treatment in rheumatology? : overview of the early recognition and screening consultation models for rheumatic and musculoskeletal diseases in Germany]. *Z Rheumatol.* (2019) 78:722–42. doi: 10.1007/s00393-019-0683-y
- Feuchtenberger M, Nigg AP, Kraus MR, Schäfer A. Rate of proven rheumatic diseases in a large collective of referrals to an outpatient rheumatology clinic under routine conditions. *Clin Med Insights Arthritis Musculoskelet Disord.* (2016) 9:181–7. doi: 10.4137/CMAMD.S40361
- Knitza J, Mohn J, Bergmann C, Kampylafka E, Hagen M, Bohr D, et al. Accuracy, patient-perceived usability, and acceptance of two symptom checkers (Ada and Rheport) in rheumatology: interim results from a randomized controlled crossover trial. *Arthritis Res Ther.* (2021) 23:112. doi: 10.1186/s13075-021-02498-8
- Krey J. [Triage in emergency departments. Comparative evaluation of 4 international triage systems]. *Med Klin Intensivmed Notfmed.* (2016) 111:124–33. doi: 10.1007/s00063-015-0069-0
- de Thurah A, Bosch P, Marques A, Meissner Y, Mukhtyar CB, Knitza J, et al. 2022 EULAR points to consider for remote care in rheumatic and musculoskeletal diseases. *Ann Rheum Dis.* (2022). doi: 10.1136/annrheumdis-2022-222341
- Kleinert S, Bartz-Bazzanella P, von der Decken C, Knitza J, Witte T, Fekete SP, et al. A real-world rheumatology registry and research consortium: the German RheumaDatenRhePort (RHADAR) Registry. *J Med Internet Res.* (2021) 23:e28164. doi: 10.2196/28164
- Knitza J, Muehlensiepen F, Ignatyev Y, Fuchs F, Mohn J, Simon D, et al. Patient's perception of digital symptom assessment technologies in rheumatology: results from a multicentre study. *Front Public Health.* (2022) 10:844669. doi: 10.3389/fpubh.2022.844669
- Schneider N, Sohrabi K, Schneider H, Zimmer K-P, Fischer B, de Laffolie J. CEDATA-GPGE study group. machine learning classification of inflammatory bowel disease in children based on a large real-world pediatric cohort CEDATA-GPGE® registry. *Front Med.* (2021) 8:666190. doi: 10.3389/fmed.2021.666190
- Vodencarevic A, Tascilar K, Hartmann E, Reiser M, Hueber AJ, Haschka J, et al. Advanced machine learning for predicting individual risk of flares in rheumatoid arthritis patients tapering biologic drugs. *Arthritis Res Ther.* (2021) 23:67. doi: 10.1186/s13075-021-02439-5
- Liu X, Faes L, Kale AU, Wagner SK, Fu DJ, Bruynseels A, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digit Health.* (2019) 1:e271–97. doi: 10.1016/S2589-7500(19)30123-2
- Bressem KK, Vahldiek JL, Adams L, Niehues SM, Haibel H, Rodriguez VR, et al. Deep learning for detection of radiographic sacroiliitis: achieving expert-level performance. *Arthritis Res Ther.* (2021) 23:106. doi: 10.1186/s13075-021-02484-0
- Emile SH, Ghareeb W, Elfeki H, El Sorogy M, Fouad A, Elrefai M. Development and validation of an artificial intelligence-based model to predict gastroesophageal reflux disease after sleeve gastrectomy. *Obes Surg.* (2022). doi: 10.1007/s11695-022-06112-x
- Lemaître G, Nogueira F, Aridas CK. Imbalanced-learn: a python toolbox to tackle the curse of imbalanced datasets in machine learning. *J Mach Learn Res.* (2017) 18:559–63.
- Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res.* (2002) 16:321–57. doi: 10.1613/jair.953
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: MACHINE LEARNING IN PYTHON. *J Mach Learn Res.* (2011) 12:2825–30.
- Shapley LS. A value for n-person games. In: Kuhn HW, Tucker AW editors. *Contributions to the Theory of Games (AM-28)*. (Vol. II), (Princeton, NJ: Princeton University Press) (2016). p. 307–18. doi: 10.1515/9781400881970-018
- Lundberg S, Lee S-I. A Unified Approach to Interpreting Model Predictions. *arXiv.* (2017) [Preprint]. arXiv:170507874,
- Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *BMC Med.* (2015) 13:1. doi: 10.1186/s12916-014-0241-z
- van den Broek-Altenburg E, Atherly A, Cheney N, Fama T. Understanding the factors that affect the appropriateness of rheumatology referrals. *BMC Health Serv Res.* (2021) 21:1124. doi: 10.1186/s12913-021-07036-5
- Harrington JT, Walsh MB. Pre-appointment management of new patient referrals in rheumatology: a key strategy for improving health care delivery. *Arthritis Rheum.* (2001) 45:295–300. doi: 10.1002/1529-0131(200106)45:33.0.CO;2-3
- Wong J, Tu K, Bernatsky S, Jaakkimainen L, Thorne JC, Ahluwalia V, et al. Quality and continuity of information between primary care physicians and rheumatologists. *BMC Rheumatol.* (2019) 3:1. doi: 10.1186/s41927-019-0067-6
- Ehrenstein B, Pongratz G, Fleck M, Hartung W. The ability of rheumatologists blinded to prior workup to diagnose rheumatoid arthritis only by clinical assessment: a cross-sectional study. *Rheumatology (Oxford).* (2018) 57:1592–601. doi: 10.1093/rheumatology/k ey127
- Pego-Reigosa JM, Peña-Gil C, Rodríguez-Lorenzo D, Altabás-González I, Pérez-Gómez N, Guzmán-Castro JH, et al. Analysis of the implementation of an innovative IT solution to improve waiting times, communication with primary

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmed.2022.954056/full#supplementary-material>



care and efficiency in Rheumatology. *BMC Health Serv Res.* (2022) 22:60. doi: 10.1186/s12913-021-07455-4

25. Knevel R, Knitza J, Hensvold A, Circiumaru A, Bruce T, Evans S, et al. Rheumatic?-a digital diagnostic decision support tool for individuals suspecting rheumatic diseases: a multicenter pilot validation study. *Front Med (Lausanne).* (2022) 9:774945. doi: 10.3389/fmed.2022.774945

26. Proft F, Spiller L, Redeker I, Protopopov M, Rodriguez VR, Muche B, et al. Comparison of an online self-referral tool with a physician-based referral strategy for early recognition of patients with a high probability of axial spa. *Semin Arthritis Rheum.* (2020) 50:1015–21. doi: 10.1016/j.semarthrit.2020.07.018

27. Powley L, McLroy G, Simons G, Raza K. Are online symptoms checkers useful for patients with inflammatory arthritis? *BMC Musculoskelet Disord.* (2016) 17:362. doi: 10.1186/s12891-016-1189-2

28. Morf H, Krusche M, Knitza J. Patient self-sampling: a cornerstone of future rheumatology care? *Rheumatol Int.* (2021) 41:1187–8. doi: 10.1007/s00296-021-04853-z

29. Kernder A, Morf H, Klemm P, Vossen D, Haase I, Mucke J, et al. Digital rheumatology in the era of COVID-19: results of a national patient and physician survey. *RMD Open.* (2021) 7:e001548. doi: 10.1136/rmdopen-2020-001548