BMC Med Inform Decis Mak

https://doi.org/10.1186/s12911-026-03806-5

Article in Press

Personalis: medical software for autoimmune diseases

Received: 29 July 2025

Accepted: 27 August 2026

Published online: 01 September 2026

Cite this article as: Meier P., Kruta J., Bopp N. *et al.* Personalis: medical software for autoimmune diseases. *BMC Med Inform Decis Mak* (2026). https://doi.org/10.1186/s12911-026-03806-5

Patrick Meier, Jan Kruta, Nicolas Bopp, Thierry Martin, Raphael Carapito, Marta Rizzi, Reinhard E. Voll, Andreas Schwarting, Anne-Sophie Korganow, Erik Schkommodau & Enkelejda Miho

Personalis: Medical Software for Autoimmune Diseases

Patrick Meier[1], Jan Kruta[1], Nicolas Bopp[1], Thierry Martin[2], Raphael Carapito[2], Marta Rizzi[3], Reinhard E. Voll[3], Andreas Schwarting[4,5,6], Anne-Sophie Korganow[7], Erik Schkommodau[1], Enkelejda Miho[1,8,9,*]

[1] FHNW University of Applied Sciences and Arts Northwestern Switzerland, School of Life Sciences, Muttenz, Switzerland.
[2] Laboratoire d'Immunologie, Plateau Technique de Biologie, Pôle de Biologie, Nouvel Hôpital Civil, Strasbourg, France.
[3] Department of Rheumatology and Clinical Immunology, Medical Center – University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany.
[4] Acura Rheumatology Center Rhineland Palatinate, Bad Kreuznach, Germany.
[5] University Center of Autoimmunity, Johannes Gutenberg University, Mainz, Germany.
[6] Division of Rheumatology and Clinical Immunology, University Medical Center, Johannes Gutenberg University, Mainz, Germany.
[7] Department of Clinical Immunology and Internal Medicine, National Reference Center for Autoimmune Diseases, University Hospitals of Strasbourg, Strasbourg, France.
[8] aiNET GmbH, Basel, Switzerland.
[9] SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland.
* Corresponding author: Prof. Dr. Enkelejda Miho (enkelejda.miho@fhnw.ch)

Keywords: medical software, CDSS, autoimmune diseases, diagnostics, machine learning, electronic health records, laboratory tests

Abstract

Background: The early and accurate detection and diagnosis of diseases play a crucial role in impeding disease progression and guiding appropriate treatment selection. Complex disease such as autoimmune diseases (ADs), often present nonspecific symptoms that can overlap also with other conditions, leading to misdiagnosis. Medical software, known as clinical decision support systems (CDSS), is used by clinicians to categorize patients based on specific criteria. However, existing CDSS implementations, generally developed within individual hospitals, are often limited to specific data types, and reflect questionnaires that do not take advantage of machine learning (ML) methods for detection and prediction of complex patterns. **Methods:** To address this gap, we developed Personalis as a proof-of-concept medical software that integrates various modalities of medical datasets and applies ML models to target hospital datasets and disease-target specific prediction tasks within a unified software environment. The platform has been run with real-world data to investigate the potential of this proof-of-concept medical software in providing early clinical support for the personalized prediction of specific autoimmune diseases in individual patients. These intermediate results have been implemented to redesign the front-end to enhance user-experience, and serve the conveyed clinical needs and expectations. **Results**: Personalis is a proof-of-concept medical software that demonstrated the feasibility of integrating various

data modalities and machine learning methods to support clinicians in the diagnosis, treatment or management of individual patients with autoimmune diseases. By leveraging ML methods, Personalis provided prediction with a certain accuracy of a specific autoimmune disease for an individual patient, highlighting factors that contributed to the prediction results. Explainable machine learning made the models' decisions understandable to humans, and the resulting factors were further used to provide additional clinical insights beyond the disease prediction accuracy metric. The platform was designed considering user-experience and human factors to enable actionable clinical practical adoption. Its results may provide hints on the prognosis assessment by selecting a specific patient record. **Conclusions**: Personalis provides a proof-of-concept medical software for integrating heterogeneous clinical data with configurable ML-based prediction and interpretation workflows for autoimmune diseases. It offers a mechanistic overview of the parameters that mostly influence the prediction of the machine learning models, therefore providing interpretable results and mechanistic insights essential to model transparency. The potential of the medical software Personalis is be extended to several diseases, and support in cases of challenging differential diagnoses.

INTRODUCTION

Autoimmunity encompasses a broad spectrum of syndromes and diseases in which the immune system erroneously targets and attacks healthy cells and tissues. This complex phenomenon manifests in various human organs and tissues, giving rise to a diverse range of ADs. ADs can exhibit systemic manifestations, as observed in systemic lupus erythematosus (SLE) and rheumatoid arthritis (RA), or display organ-specific involvement, as seen in diseases like Graves' disease and type 1 diabetes. These conditions represent intricate immunological dysregulation, and necessitate comprehensive understanding and targeted therapeutic interventions for effective disease management[1]. It can be particularly difficult to diagnose autoimmune disorders precisely and early due to their diverse clinical manifestations depending on the site of the autoimmune attack to any of the organs or tissues in the body[2]. Given the rise in prevalence and incidence of such disorders globally, the chronic nature of autoimmunity, and the associated high burden of monitoring and treatment due to the personal course of disease and

personalized response to therapeutics, novel technologies to care require immediate attention[3]. Current challenges such as dissecting the heterogeneity of ADs at the molecular and genetic levels, and identifying biomarkers for individualized diagnosis and treatment[4], can be addressed by leveraging modern digital technologies. These techniques, including machine learning (ML), could support clinicians in practice and research with patients who present signs and symptoms from various potential diseases/etiologies, and that may require a distinct and sometimes divergent approach to management[5]. Digital tools such as medical software, commonly known as clinical decision support (CDSS) platforms, have the potential to alleviate the diagnostic odyssey and support specialists as well as unfamiliar general practitioners, especially in cases of limited access to specialist care and for patients with rare autoimmune disorders[6].

Medical software plays a crucial role in supporting clinical practice to leverage data analytics on medical data, especially in cases of large-scale molecular and cellular datasets, and use Electronic Health Records (EHR) to provide valuable clinical insights on the patient health status[7,8]. CDSS encompass a wide range of applications in medical care, from minimizing drug-drug interactions, to providing dose guidance, and highlighting critical care values for blood sugar levels[9,10]. Furthermore, it has been demonstrated that their use enhanced diagnostic accuracy for specific diseases and improved adherence to guideline recommendations lead to favorable clinical outcomes. For instance, in adults with conditions like heart failure, reduced exacerbations, improved quality of life, better patient-reported outcomes, and lower mortality rates could be achieved[11]. These findings emphasize the effectiveness of CDSS in improving diagnostic precision, treatment adherence, and overall patient outcomes.

CDSS are categorized into different classes based on various criteria. The most common classification divides CDSS into knowledge-based and nonknowledge-based systems[12,13]. Knowledge-based systems rely on rules (e.g., if–then) to evaluate input data and generate corresponding outputs or actions that simulate human thinking. On the other hand, nonknowledge-based systems use artificial intelligence (AI) techniques, such as ML or statistical methods, to learn from past experiences for pattern recognition to make decisions, rather than relying solely on expert knowledge. By incorporating AI techniques, these

systems can identify patterns in clinical data[14]. Although they share similar components, slight variations in these components result in a wide range of possibilities. CDSS are a subset of Decision Support Systems (DSS) used in the medical field[15]. They are primarily distinguished based on: i) the timing of support provided (before, during, or after clinical decision-making), ii) whether they offer active alerts or respond passively to physician input, and iii) the underlying foundation on which they are built[8]. Despite the pressing need for high-quality decision support, CDSS encounter various challenges such as: the amount and complexity of data available, data management issues, regulatory compliance, and a lack of user-friendliness. These challenges need to be addressed to enhance acceptance in practice and provide validated support to physicians[16]. To address these challenges, we developed Personalis as a configurable proof-of-concept CDSS that integrates heterogeneous clinical datasets, dataset-specific preprocessing protocols, multiple ML classifiers, prediction workflows, and model-result visualization within a single software environment with a user and access management administration interface. Rather than implementing a single disease- or model-specific prediction pipeline, Personalis provides a common architecture in which different datasets, prediction disease targets, and ML models can be configured and applied through the same workflow.

METHODS

Data

Dataset I. The Lupus BioBank of the upper Rhine valley (LBBR) database contains an extensive collection of clinical and biological data from a cohort of 1300 patients diagnosed with diverse autoimmune disorders. This database encompasses vital information, including serum, plasma, DNA, and cell samples. The patients included in the database were identified as having conditions such as systemic lupus erythematosus (SLE), discoid Lupus erythematosus (DLE), and systemic sclerosis (SS). To compile this dataset, collaborative efforts were undertaken with 17 hospitals located within the upper Rhine valley region, which spans both France and Germany. The LBBR, contains approximately 1000 patients exclusively diagnosed with SLE who fulfil the ACR classification criteria. Additionally, 300 patients with

SLE and other autoimmune diseases, recruited through the RARENET project, contribute to the dataset's diversity and richness.

Dataset II. The dataset used in the Rheuma-Vor App (RAPP) consists of clinical information collected through a questionnaire completed by physicians as previously described[17]. This dataset includes data from 815 patients who presented symptoms related to rheumatic disease. The questionnaire consists of seventeen specific questions aimed at either confirming or ruling out suspicions of three autoimmune diseases: rheumatoid arthritis (RA), axial spondyloarthritis (SpA), and psoriatic arthritis (PsA). These diseases are prevalent inflammatory joint conditions that present significant diagnostic and treatment challenges for rheumatologists[18].

Dataset III. Bead-based multiplex assays were utilized to quantify the concentrations of cytokines, chemokines, immunoglobulins, and other biomarkers in order to elucidate their expression profiles in patients with SLE. The obtained serum data was subjected to analysis using the BioLegend LEGENDplex™ Data Analysis Software. The resulting outcomes are reported as concentrations expressed in picograms per milliliter (pg/mL), unless explicitly stated otherwise. The data presented in this study was derived from a sample size of 23 patients diagnosed with SLE.

Software Development

The software structure of the Personalis platform was implemented using the MEAN stack, an open-source JavaScript software stack comprising multiple individual software solutions. This stack offers the notable advantage of enabling the development of both server-side (e.g., the hospital data center) and client-side (the clinical interface) code using a unified programming language, JavaScript. By harnessing the capabilities of these contemporary, versatile, and open-source components, the MEAN stack emerges as an appropriate solution for the platform's development. **Supplementary Table S1** provides an overview of the key components in the stack, including Angular, Express.js, Node.js, and MongoDB[19]. The MEAN stack provides a unified and efficient environment for the development of sophisticated websites and web applications, streamlining workflow processes, and leveraging the developer-friendly nature of JavaScript, which encompasses support for object-oriented principles and asynchronous

programming paradigms. Moreover, the complete application can be executed within a fully modular and dockerized environment, accompanied by a secure NGINX reverse proxy server.

The document schema refers to JSON objects utilized to establish the structure and composition of documents within a collection. It serves the purpose of specifying a desired set of fields or configuring the contents of a field. In this project, a schema was meticulously crafted for each dataset, delineating the obligatory and permissible fields. Leveraging the capabilities of Mongoose, a data model can be generated based on the prescribed schema, aligning with the intended model representation. These models facilitate the creation and retrieval of documents from the MongoDB database, where an individual instance derived from a model is termed a document. Thus, models assume the responsibility of managing the generation and retrieval operations pertaining to documents within the MongoDB collection.

Machine Learning

The datasets served as the basis for a series of supervised machine-learning analyses. Model performance was assessed using cross-validation, allowing the predictive behavior of the classifiers to be evaluated across the available labelled samples. This approach allows for the comparison of the applicability of a model to a specific dataset, as models can perform differently depending on the provided data. Internal model performance was quantified from the cross-validation predictions using accuracy, precision, recall (sensitivity), F1-score, and ROC-AUC; a confusion matrix was additionally generated for each classification task. Binary and multiclass classification tasks were implemented depending on the prediction target. For multiclass tasks, including the RAPP disease classification, precision, recall, and F1-score were calculated using micro averaging, while ROC-AUC was calculated from class probabilities using a one-vs-one multiclass formulation. The prediction target was defined according to the selected dataset and analysis task and included both binary clinical outcomes and multiclass disease classifications.. The analysis employed supervised learning techniques, including random forest (RF), logistic regression (LR), stochastic gradient descent classifier (SGD), neural network (NN), and support vector machine (SVM), combined with dataset-specific preprocessing within a scikit-learn pipeline. Numeric predictors were processed using four-nearest-neighbor imputation followed by standard scaling.

Categorical predictors were imputed using a constant missing-value category and subsequently one-hot encoded, with previously unseen categories ignored. These preprocessing steps were included within the machine-learning pipeline and were therefore fitted separately within each cross-validation training fold. Class imbalance was addressed using class_weight="balanced" for the RF, LR, SGD, and SVM classifiers, whereas no explicit class-imbalance correction was applied to the NN classifier. No over- or undersampling procedures were used. For internal performance estimation, all available labelled samples contributed to the cross-validation procedure, with each sample predicted by a model fitted without that sample. After cross-validation, the final pipeline was fitted on all available labelled samples for subsequent patient-level prediction. Model performance was internally assessed using out-of-fold predictions generated by cross-validation over the available labelled samples. To ensure the inclusion of as much relevant information as possible, all applicable records in the database that can contribute towards a result are utilized once a model is trained. This approach accommodates the variations in sample sizes across the datasets and maximizes the utilization of available training data for accurate classification.

To provide a standardized quantitative assessment of the implemented classifier families, one representative prediction target was evaluated for each dataset: discoid lupus for LBBR, disease class (None, RA, PsA, or SpA) for RAPP, and discoid lupus for the Cytokine dataset. Samples without an observed target value were excluded. Fixed predictor sets were used, excluding the prediction target, administrative identifiers, and variables identified as direct target-derived or diagnostic proxies; the predictor and exclusion lists and hyperparameter search spaces used for the standardized evaluation are provided in Supplementary Table S2. Model performance was assessed using nested stratified cross-validation. For LBBR and RAPP, five outer folds and five inner folds were used, whereas three outer and two inner folds were used for the smaller Cytokine dataset. Hyperparameters were selected by accuracy within the inner cross-validation using only the respective outer-training data. All preprocessing steps, including missing-value imputation, scaling, and categorical encoding, were fitted within the training folds. Performance metrics were calculated from the pooled outer-fold out-of-fold predictions, ensuring

that each evaluated sample was predicted by a model that had not been trained on that sample. Accuracy, precision, sensitivity/recall, specificity, F1-score, and ROC-AUC were calculated; for the multiclass RAPP task, macro-averaged metrics were used. A DummyClassifier based on class priors was included as an evaluation-only baseline. Ninety-five percent confidence intervals were obtained from 2,000 class-stratified bootstrap resamples of the pooled out-of-fold predictions. For each bootstrap sample, the respective performance metric was recalculated, and the 2.5th and 97.5th percentiles of the resulting distribution were used as the confidence limits.

RESULTS

We developed the web-based clinical decision support platform "Personalis" for the integration and analysis of heterogeneous medical datasets in autoimmune diseases. The platform combines data management, configurable machine-learning workflows, patient-level prediction, and visualization of model outputs within a unified software environment (**Figure 1**). The objective of the platform is to support the exploration and interpretation of patient data and to provide a flexible framework for clinical decision-support research.

Figure 1. Schematic overview of the Personalis platform and its workflows. The diagram illustrates the main components involved in platform implementation and use, including data integration and management, user and administrator interactions, and the machine-learning workflow. The arrow indicates the simplified relationships and flow of information between the respective platform components and clinician stakeholders.

A non-relational database integrates various clinical data modalities

In the development of the platform prototype, we have integrated MongoDB, a renowned NoSQL database, to make the most of its capabilities in storing and managing data efficiently. Through the strategic use of MongoDB's NoSQL architecture, we have harnessed its inherent strengths, leveraging a flexible and adaptable data structure that deviates from the limitations imposed by traditional SQL databases. This pivotal decision has granted the platform a remarkable degree of versatility, as the MongoDB's non-relational data structure facilitates seamless scalability through the integration of additional servers to accommodate the ever-expanding and evolving datasets. The diversity of data available for the platform encompasses a wide array of sources, ranging from patient-generated clinical data collected during hospitalization to human genetic sequencing data. Furthermore, the collaborative nature of the platform invites continuous contributions, ensuring that novel datasets are consistently incorporated into the database over time. To ensure optimal organization and accessibility, we have devised a meticulous preprocessing pipeline that harmonizes and standardizes the disparate data sources, resulting in the establishment of a comprehensive spreadsheet format. This curated dataset is securely stored on a dedicated server, serving as a permanent backup repository, safeguarding its integrity and availability. Empowered by this robust infrastructure, the platform opens new frontiers for exploration and analysis, leveraging advanced ML models to unlock valuable clinical insights and propel advancements in personalized healthcare.

Considering the sensitive nature of the data managed by the platform, stringent measures were implemented in collaboration with experts from the Swiss Institute of Bioinformatics to fortify data security and protect against unauthorized access and misuse. By leveraging secure connection protocols and implementing user authentication mechanisms, access to the platform is granted only through approved user accounts administered by an authorized administrator. Additionally, to enhance data privacy, all user data is encrypted within the system. Sensitive information such as passwords is stored in a hashed format, mitigating the risk of data breaches and ensuring the confidentiality of user information. These measures ensure a high level of confidentiality and privacy for the data, mitigating risks and

safeguarding against potential breaches. Additionally, by implementing the offline hosting capability, Personalis can be hosted on a local server without reliance on external cloud providers or similar services. This further enhances the platform's security and data privacy, providing an additional layer of protection for sensitive information.

A resilient and efficient back-end infrastructure

For the realization of the Personalis platform, a strategic combination of Express.js and Node.js technologies were leveraged to create a robust back-end infrastructure. Express.js, a prominent back-end web framework for Node.js, was employed to streamline the process of web application development by providing a more user-friendly interface compared to Node.js itself. By harnessing the extensive functionalities offered by Express.js, we were able to construct a powerful web server, establish efficient routing mechanisms, and seamlessly integrate middleware to intercept and manipulate incoming requests before proceeding with further processing.

Node.js, on the other hand, served as the underlying server-side platform, enabling the execution of JavaScript code beyond the confines of a web browser environment. Renowned for its asynchronous and event-driven architecture, Node.js offered a scalable runtime environment tailored for building network applications. Within this project, Node.js played a pivotal role in handling incoming requests, generating appropriate responses based on the nature of the request, executing server-side commands, and facilitating seamless interactions with both the database and script files. By utilizing Express.js and Node.js in a harmonious symbiosis, we have optimally harnessed the capabilities of these technologies to establish a resilient and efficient back-end infrastructure for the Personalis platform. This strategic integration empowers the platform to deliver exceptional performance, scalability, and interactivity, enabling seamless communication between the front-end and back-end components while ensuring optimal utilization of resources and facilitating a seamless user experience.

User authentication ensures proper data management and robust security

The platform ensures robust security through user authentication and the implementation of secure web protocols to mitigate unauthorized access, including the utilization of a reverse proxy that selectively

directs valid client requests to the backend (**Figure 2, Supplementary Table 2**). Furthermore, security measures are enhanced by segregating data storage and presentation within the browser and utilizing REST for information transfer. The adoption of a non-relational database enables scalability and flexibility in data capacity. In cases where datasets undergo modifications, necessary adjustments to the database or collections can be easily accommodated with minimal effort. These measures collectively enhance the platform's security, adaptability, and data management capabilities.

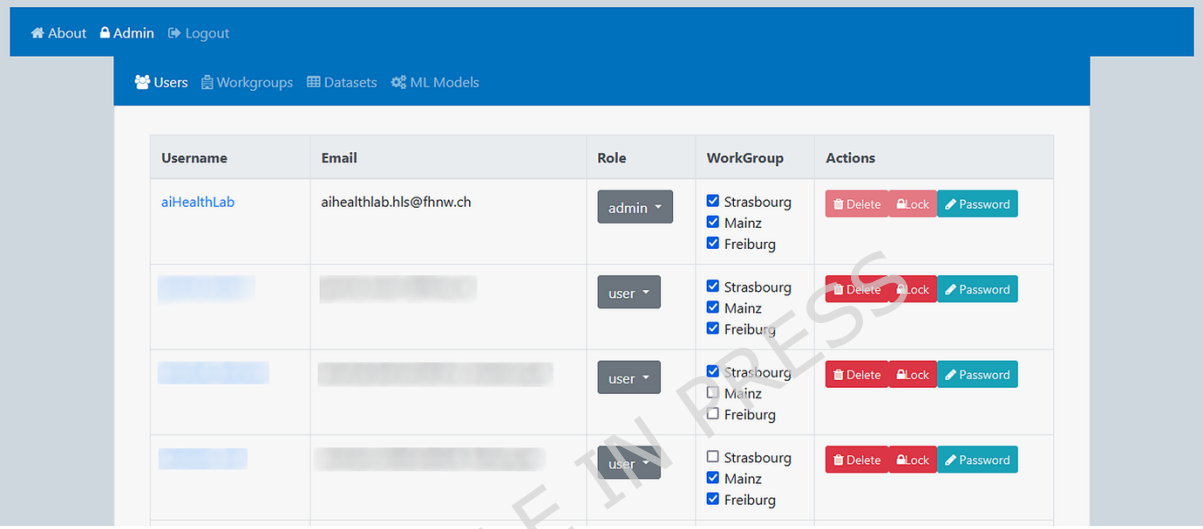

Figure 2. User and access management administration interface of the Personalis platform. The administrator view provides access to user profiles and their assigned workgroups and roles, which determine the corresponding access permissions within the platform. The navigation panel also provides access to the administration of workgroups, datasets, and machine-learning models. Supplementary Table S3 summarizes the actions available to the respective user roles.

Machine learning model evaluation shows prediction of autoimmune diseases

Personalis implements multiple supervised machine-learning classifiers to generate dataset- and disease target-specific predictions from the medical data integrated into the platform (**Figure 3, Supplementary Table S3**). To quantitatively assess the implemented classifier families, we performed an internal nested cross-validation evaluation using one representative prediction task from each retrospective dataset (**Figure 4, Supplementary Table S4**). For LBBR, model performance showed a clear trade-off between

sensitivity and specificity. Random forest (RF) models achieved an accuracy of 0.800 and a specificity of 0.948; however, sensitivity was 0.216. Logistic regression (LR) showed lower accuracy (0.726) and specificity (0.808) but higher sensitivity (0.403) compared to the other models. The LBBR accuracy values should be interpreted in the context of class imbalance, as the `DummyClassifier` function achieved an accuracy of 0.797 while showing no sensitivity (zero) for all positive class. For the multiclass RAPP task, RF achieved an accuracy of 0.538 and a macro ROC-AUC of 0.756, compared with 0.481 and 0.495, respectively, for the `DummyClassifier`. Performance estimates for the Cytokine dataset were substantially more uncertain because only 18 samples were available; for example, support vector machines (SVM) achieved an accuracy of 0.667 (95% CI 0.444–0.889) and a ROC-AUC of 0.639 (95% CI 0.375–0.875). Overall, the evaluation illustrates that the classifier performance and the balance between different performance measures varied substantially across datasets and models. These results show the strong diagnostic utility potential of the Personalis medical software, and its' ability to accommodate several machine learning models. However, they also demonstrate ML limitations primarily related to the small dataset constrains and target class imbalance.

To ensure flexibility and adaptability, we have designed the platform to allow users to create and customize their own ML models according to the selection of RF, LR, stochastic gradient descent (SGD), neural networks (NN) and SVM models. By providing a user-friendly interface, clinicians and researchers can define the classifiers and target variables that align with their specific research objectives and medical domain (**Figure 4**). This empowers users to tailor the models to their unique needs and explore the potential relationships and patterns within their data.

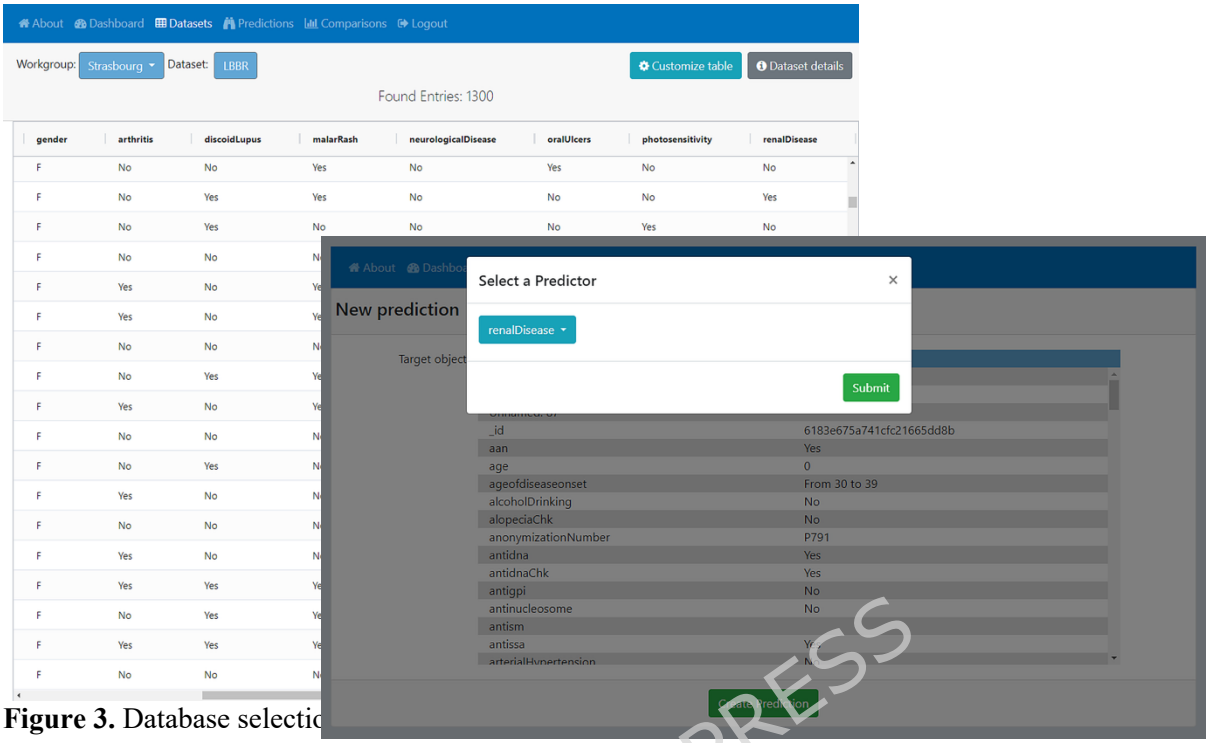

Figure 3. Database selectio
predictors for autoimmune disorders. The interface allows users to select the dataset and the prediction target, and to define the predictor variables used for the machine-learning analysis. The displayed options illustrate how dataset-specific prediction tasks can be configured before model training for patient-level prediction.

To facilitate efficient analysis and generate predictions, we have developed a robust infrastructure that incorporates pre-processing and storage of the medical data on a dedicated server. This ensures that the data is transformed and prepared in a standardized format, optimizing the compatibility with the ML models. By employing best practices in data pre-processing, such as data cleaning, normalization, and feature extraction, we strived to enhance the quality and reliability of the predictions generated by the platform. Once the ML analysis is performed, the results are seamlessly transmitted back to the user. In order to enhance the interpretability and accessibility of these results, we have implemented a detailed visualization module within the application as guided by design principles and intermediate user testing interviews. This module provides a comprehensive overview of the predictions, highlighting important features and potential differences in the expression of individual values. This visualization provides users with information on the model prediction and the features contributing to the corresponding output,

thereby supporting interpretation of the model results. These are actionable insights that can guide further ML analyses and clinical investigations; therefore, they empower users, inform their clinical decision-making, and have the potential contribute to improved patient outcomes.

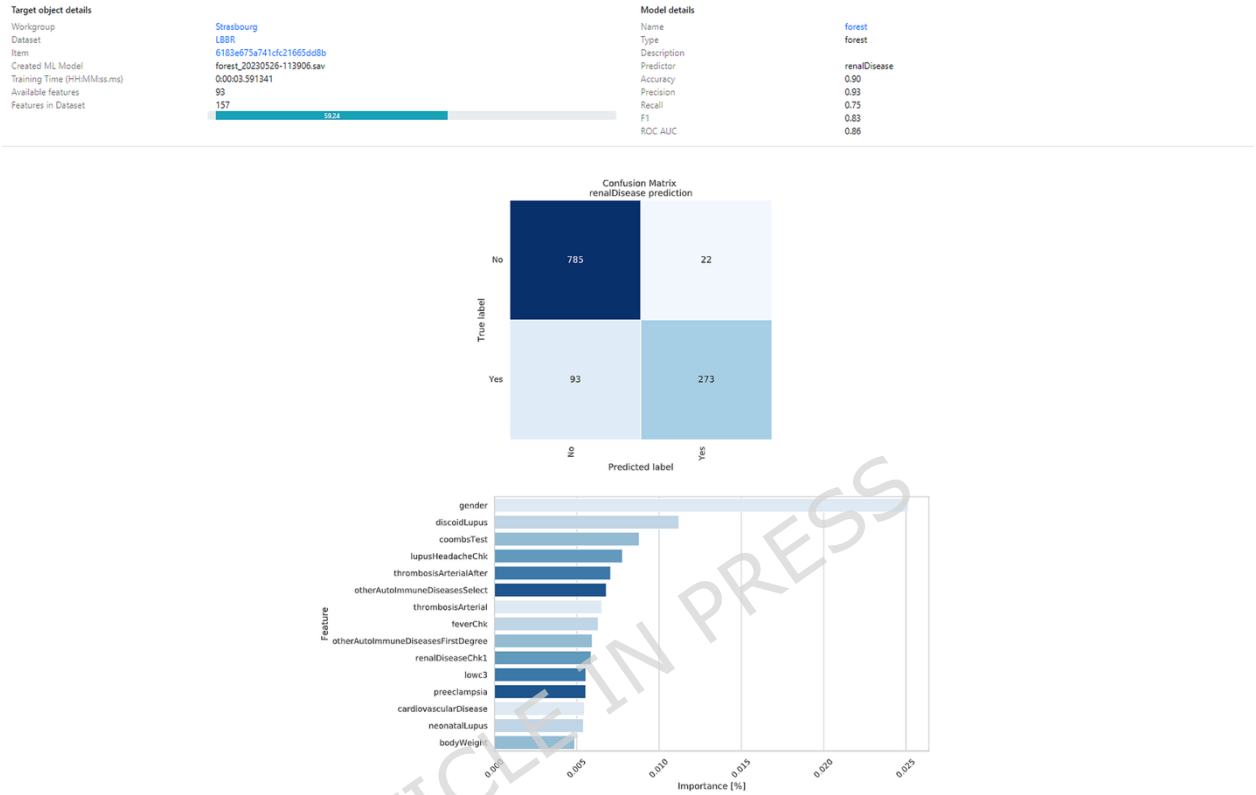

Figure 4. Prediction result interface of the Personalis software. This view displays the selected dataset and machine learning model, and the resulting patient-level prediction and model's evaluation metrics. Additional panels provide information on model confidence and features' importance to support interpretation of the generated disease prediction.

Discussion

The accurate identification and diagnosis of autoimmune disorders pose a significant challenge due to their complex symptoms and diverse data types employed in the diagnostic process. CDSS emerged as a crucial solution in addressing this challenge by providing essential assistance and guidance to healthcare professionals. CDSS harnesses specific data types and use advanced algorithms, including ML, to effectively analyze intricate patient data, thereby aiding in the prediction, diagnosis, and treatment of ADs[20,21]. Integrating medical software into clinical practice yields numerous benefits, such as improved

decision-making, enhanced diagnostic accuracy, and optimized patient care within the realm of ADs. Moreover, the implementation of new data using CDSS, ML, and data schemata was made feasible through the innovative Personalis software. This platform enables the analysis of complex data for disease prediction, and ensures rapid access to the desired information and results. However, the current implementation has several methodological limitations. Preprocessing is dataset-specific and includes fixed missing-value handling strategies that may require adaptation when applied to new clinical data sources. Class imbalance is addressed using model-specific strategies, with balanced class weights applied to RF, LR, SGD, and SVM, while the NN is trained using its standard unweighted configuration. Feature handling relies on dataset-specific rule-based filtering rather than a dedicated statistical feature-selection procedure. On the other hand, the platform supports optional grid search cross-validation `GridSearchCV`-based hyperparameter tuning. Model performance is currently estimated using internal cross-validation, while validation in independent held-out or external cohorts remains an important next step for assessing model generalizability. The case-study internal evaluation showed a substantial variation in performance across datasets and metrics. In LBBR, relatively high accuracy and specificity coincided with low sensitivity, while the class-prior baseline achieved similar accuracy, highlighting the limitations of accuracy alone for this imbalanced data, and the importance of additional clinical data. The RAPP dataset showed a clearer discrimination beyond the baseline, particularly with the RF model, whereas cytokine performance estimates were highly uncertain because of the small sample size. These findings highlight the usefulness of the flexibility to select multiple performance measures and baseline comparisons when interpreting a model's performance. Despite these limitations, its ability to seamlessly integrate new data distinguishes it from existing solutions[20,21] and sets it as a proof-of-concept in medical software for autoimmune diseases. A specific advantage of the application to the autoimmune diseases case study is that the Personalis software captures a clinical signal; in fact, most machine learning limitations of the present prototype is the limited sample size of some datasets, particularly the biomarker dataset. Larger and independent cohorts will be needed to obtain more robust estimates of model performance and generalizability.

Additionally, the Personalis user-friendly interface, reflects research results on user-experience and human factors embedded in its design. These are generally overseen by scientifically metric-heavily focused solutions: it empowers users to perform complex analyses with ease, further facilitating efficient clinical decision-making and patient management. In clinical practice, physicians encounter demanding situations while treating a diverse patient population with varying conditions, comorbidities, and treatment plans. To effectively assist in these scenarios, support tools must exhibit exceptional user-friendliness, adaptability to individual settings, high interoperability, and robust risk management similar to medical devices. However, the frequent use of alert systems can result in alert fatigue, leading to increased disregard of suggestions by users. This fatigue engenders a negative perception that impedes the acceptance of potentially more user-friendly technologies[20,22]. Personalis implements a clear user-driven flow of information.

The implementation of CDSS in healthcare settings raises also foreseeable legal challenges, including adherence to professional standards, testing requirements, liabilities, and data protection regulations. Overcoming these challenges requires translational research with explicit patient consent, establishing clear legal responsibilities, and informing patients about data usage and storage choices. Additionally, anticipating and adapting to future regulatory aspects is crucial for addressing the legal challenges associated with CDSS implementation[23,24]. Furthermore, a significant challenge lies in effectively harnessing unstructured and incomprehensive data, which constitutes a substantial proportion of the data found in clinical care documents[25]. These are limitations needed to address for further adaptations of the Personalis prototype in order to achieve regulatory-ready, accurate and reliable results across datasets.

Conclusion

The development of the Personalis medical software platform represents an advancement in the field of CDSS for the diagnosis and treatment of ADs. By integrating diverse medical datasets with configurable ML models, Personalis enables dataset- and target-specific prediction workflows for patient-level diagnostics and provides a framework for exploring their potential use in clinical decision support of several diseases. This prototype represents a proof-of-concept for medical software. Its inter-disciplinary

format ranging from integrating various clinical data types, from various hospital sources (from various languages, legal and IT systems), satisfying main principles of software engineering, applying multiple machine learning models as case-studies, and, at last, integrating user-experience research results as a layer reflected in the ultimate design, can limit an in-depth and detailed presentation of specific aspects of this study–such as data handling, or specific prediction results of one or the other ML model. Despite the ambitious breadth of this CDSS platform, its nonknowledge-based approach, coupled with its user-friendly interface, empowers healthcare professionals to customize and create their own ML models, enhancing flexibility and adaptability.

One of the major hurdles for using CDSS in the past was the complex implementation and adaptation of the system to different datasets. However, this challenge has been overcome through the utilization of modern frameworks and data schemata. Personalis is a personalized medical software platform designed to support the analysis and interpretation of patient data within a clinical decision-support framework. The platform enables easy access to large datasets from various sources, which can be seamlessly combined onto a single software. With a few clicks, the platform outputs predictions, provides a list of transparent weighted features where these predictions were drawn from, and can provide support to physicians in patient diagnoses, and treatment strategies. The implementation of the software as a web-based application, makes it accessible across multiple devices; its adaptability and optimization capabilities make it customizable as needed in healthcare facilities.

Through rigorous data management practices, including the use of a scalable MongoDB database for structured data storage and the implementation of robust security measures, such as user authentication protocols and encrypted communication, Personalis ensures the confidentiality and integrity of patient data. Nevertheless, the results of its use remain largely dependent on the clinical question, data quality, and hospital settings. The platform's unified open-source stack solution, powered by Express.js and Node.js technologies, facilitates seamless communication between front-end and back-end components, ensuring optimal performance and scalability. The data preprocessing and standardized formatting, combined with the detailed visualization of results, enables Personalis to provide clinicians and

researchers with actionable insights to support clinical decision-making and improve patient outcomes. Overall, the Personalis platform may represent a valuable tool in the ongoing efforts to enhance the detection, diagnosis, and treatment of complex ADs.

Declarations

Abbreviations

CDSS: clinical decision support system

DSS: decision support system

AI: artificial intelligence

ML: machine learning

EHR: electronic health record

RAPP: Rheuma-Vor App

LBBR: Lupus BioBank of the upper Rhine valley

RA: rheumatoid arthritis

SpA: axial spondyloarthritis

PsA: psoriatic arthritis

SLE: lupus erythematosus

DLE: discoid Lupus erythematosus

SS: systemic sclerosis

RF: random forest

LR: logistic regression

SGD: stochastic gradient descent classifier

NN: neural network

SVM: support vector machine

AUC: area under the curve

MEAN: MongoDB, Express.js, Angular, Node.js

SQL: Structured Query Language

REST: Representational State Transfer

Ethics Approval and Consent to Participate

The LBBR database was approved by the national data protection commission, Commission Nationale Informatique et Liberté (CNIL). Patients gave informed consent before inclusion. The RAPP study design was reviewed by the ethics committees of the state medical association of RhinelandPalatinate (837.260.17). All experiments were performed in accordance with relevant local guidelines and regulations and the Declaration of Helsinki.

Consent for Publication

Not Applicable.

Availability of Data and Materials

The datasets analyzed in the current study are available from the corresponding author upon reasonable request.

Competing Interests

EM owns shares in aiNET GmbH. Other authors declare no competing interests.

Funding

This work has been financially supported by the Interreg Europe 2018 grant Interreg V-Projekt Nr. WO 2.2 "PERSONALIS" and Schweizerische Eidgenossenschaft (NRP), Kanton Basel-Stadt, Kanton Basel-Landschaft, and Kanton Solothurn to Prof. Dr. Enkelejda Miho.

Author Contributions

EM, TM, RC, MR, RV and AS conceived and designed the study. TM, RC, MR, RV and AS provided necessary data. PM and EM developed the platform’s prototype. JK maintained the data science pipelines and run the machine learning models. NB maintained the software. EM and PM wrote the first draft of the manuscript. All authors contributed to writing the manuscript. All authors participated in scientific and development discussions.

Acknowledgments

We acknowledge and extend our gratitude to the FHNW Life Sciences Institute of Medical Engineering and Medical Informatics for their generous support in facilitating this research endeavor. We would also like to express our gratitude to Erand Smakaj and Lmar Marie Babrak for their dedicated support and valuable work to establish the first prototype while at FHNW, Boglarka Racz for the user-experience and user-interface design research during her master thesis, Anna Raselli for the investigation of the requirements in the current Swiss and European regulatory frameworks towards the approval of the Personalis CDSS as a stand-alone software as a medical device during her master thesis, Andreas Seelig for continuous server support, and Teyfik Agac for his preliminary literature search on CDSS as part of his work as a scientific assistant. Their input and support were fundamental throughout the research for the PERSONALIS software.

References

1 eBioMedicine. Multifaceted autoimmunity: new challenges and new approaches. *eBioMedicine* 2023; **88**. DOI:10.1016/j.ebiom.2023.104474.

2 Chinnathmbi A, Alyousef MA, Mofeez FIA, Alsenaid AI, Alanzi MS, Al-Madani SO. Novel Approaches to Autoimmune Diseases–A Review of New Studies. *Biosciences Biotechnology Research Asia* 2016; **13**: 1421–8.

3 Miller FW. The increasing prevalence of autoimmunity and autoimmune diseases: an urgent call to action for improved understanding, diagnosis, treatment, and prevention. *Current Opinion in Immunology* 2023; **80**: 102266.

4 Felten R, Sagez F, Gavand P-E, *et al.* 10 most important contemporary challenges in the management of SLE. *Lupus Science & Medicine* 2019; **6**: e000303.

5 Pisetsky DS. Pathogenesis of autoimmune disease. *Nat Rev Nephrol* 2023; **19**: 509–24.

6 Rheumatology TL. Long-awaited action on rare diseases. *The Lancet Rheumatology* 2022; **4**: e229.

7 Adane K, Muluye D, Abebe M. Processing medical data: a systematic review. *Archives of Public Health* 2013; **71**: 27.

8 Miller RA, Waitman LR, Chen S, Rosenbloom ST. Decision Support During Inpatient Care Provider Order Entry: The Vanderbilt Experience. In: Berner ES, ed. Clinical Decision Support Systems. New York, NY: Springer New York, 2007: 215–48.

9 Kuperman GJ, Teich JM, Tanasijevic MJ, *et al.* Improving Response to Critical Laboratory Results with Automation: Results of a Randomized Controlled Trial. *J Am Med Inform Assoc* 1999; **6**: 512–22.

10 Kubben P, Dumontier M, Dekker A, editors. Fundamentals of Clinical Data Science. Cham (CH): Springer, 2019 http://www.ncbi.nlm.nih.gov/books/NBK543527/ (accessed Aug 15, 2025).

11 Sunjaya AP, Ansari S, Jenkins CR. A systematic review on the effectiveness and impact of clinical decision support systems for breathlessness. *npj Prim Care Respir Med* 2022; **32**: 29.

12 Berner ES, editor. Clinical Decision Support Systems. New York, NY: Springer, 2007 DOI:10.1007/978-0-387-38319-4.

13 Osheroff JA, Teich J, Levick D, *et al.* Improving Outcomes with Clinical Decision Support: An Implementer's Guide, Second Edition, 2nd edn. New York: HIMSS Publishing, 2012 DOI:10.4324/9780367806125.

14 Deo RC. Machine Learning in Medicine. https://www.ahajournals.org/doi/10.1161/CIRCULATIONAHA.115.001593 (accessed Aug 15, 2025).

15 Jiang Y, Qiu B, Xu C, Li C. The Research of Clinical Decision Support System Based on Three-Layer Knowledge Base Model. *Journal of Healthcare Engineering* 2017; **2017**: 6535286.

16 Sittig DF, Wright A, Osheroff JA, *et al.* Grand challenges in clinical decision support. *Journal of Biomedical Informatics* 2008; **41**: 387–92.

17 Dreher M, Witte T, Hoeper K, *et al.* Rheuma-VOR study: optimising healthcare of rheumatic diseases by multiprofessional coordinating centres. *Annals of the Rheumatic Diseases* 2024; **83**: 184–93.

18 Kerola A, Rollefstad S, Kazemi A, *et al.* Psoriatic arthritis, axial spondyloarthritis and rheumatoid arthritis in Norway: nationwide prevalence and use of biologic agents. *Scandinavian Journal of Rheumatology* 2023; **52**: 42–50.

19 Jain P. Comparative analysis of MEAN stack and MERN stack. 2021; **8**.

20 Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. *npj Digit Med* 2020; **3**: 17.

21 Hak F, Guimarães T, Santos M. Towards effective clinical decision support systems: A systematic review. *PLOS ONE* 2022; **17**: e0272846.

22 Dramburg S, Marchante Fernández M, Potapova E, Matricardi PM. The Potential of Clinical Decision Support Systems for Prevention, Diagnosis, and Monitoring of Allergic Diseases. *Front Immunol* 2020; **11**: 2116.

23 Mitchell C, Ploem C. Legal challenges for the implementation of advanced clinical digital decision support systems in Europe. *J Clin Transl Res* 2018; **3**: 424–30.

24 Laka M, Carter D, Milazzo A, Merlin T. Challenges and opportunities in implementing clinical decision support systems (CDSS) at scale: Interviews with Australian policymakers. *Health Policy and Technology* 2022; **11**: 100652.

25 Zikos D, DeLellis N. CDSS-RM: a clinical decision support system reference model. *BMC Medical Research Methodology* 2018; **18**: 137.