Engineered Science
DOI: https://dx.doi.org/10.30919/es1836

Architecture, Implementation and Experimental Study of the Health Passport Intellectualized Software Package for the Diagnosis of Clinical and Hematological Syndromes

Indira Uvaliyeva,[1] Zhanat Idrisheva,[1,*] David Borozenets,[1] Shynar Tezekpayeva,[1] Zarına Khassenova[1,*] and Zhanar Beldeubayeva[2,*]

Abstract

The article presents the architecture, software implementation, and experimental validation of an intellectualized software package developed within the framework of the "Health Passport" digital platform for automated diagnosis of clinical and hematological syndromes, in particular, anemias of various morphological and biochemical nature. The proposed system is implemented in the form of a modular software and hardware complex that covers the entire analysis cycle: from the collection and pre-processing of laboratory data to the calculation of diagnostic indices, morphological classification, probabilistic forecasting and the formation of a diagnostic conclusion. The key modules of the platform include: Data Acquisition (structured data import), Data Preprocessing (data cleaning and normalization), Mathematical Evaluation (calculation of MCV, MCH, MCHC indices and aggregated anemia score), Morphological Classification (anemia presence, type and character determination), as well as Ensembling and Neural modules for building and training machine learning models and neural networks. As part of the experimental study, classification accuracy was achieved up to 98.1% (XGBoost) and 100% (neural network classifier with probabilistic output). The system demonstrated the stability, reproducibility and clinical interpretability of decisions. The results obtained indicate the high applied importance of the complex as a tool for standardizing laboratory diagnostics, reducing the burden on doctors and integrating into digital healthcare ecosystems.

Keywords: Digital Health Passport; Clinical and Hematological syndromes; Anemia; Machine Learning Classification; Intelligent Medical Software Architecture.
Received: 29 May 2025; Revised: 15 October 2025; Accepted: 21 October 2025
Article Type: Research article.

1. Introduction

Anemia, along with its accompanying clinical and hematological manifestations, continues to be one of the most common diseases on the planet, affecting over 1.6 billion people and posing a serious global threat to public health.[1] Affecting almost a third of the world's inhabitants, this pathology significantly increases morbidity and mortality, reduces work productivity, and also slows down psychomotor development, especially in children and women during pregnancy.[2] According to the reports of the World Health Organization (WHO), the prevalence of anemia is 24.8% among children under 5 years of age, 29.6% among pregnant women and 37.9% among women of fertile age, with the iron deficiency form dominating, accounting for more than half of all cases.[3] Clinically, anemia is expressed in various syndromes, from chronic fatigue, shortness of breath and palpitations to cognitive disorders, and in hematological terms it is characterized by a microcytic hypochromic structure of red blood cells and a drop in hemoglobin below 110 g/l in children and 120 g/l in women.[4] The problem is particularly acute in low- and middle-income countries, where the incidence of anemia varies from 40 to 60% in certain areas of sub-Saharan Africa and South Asia, mainly among the most vulnerable categories: women of fertile age, pregnant women, and infants under 5 years of age.[5,6] In these regions, anemia is aggravated by a complex of risk factors, including iron deficiency from poor nutrition, heavy blood loss during

[1]*D. Serikbaev East Kazakhstan Technical University, 19 D. Serikbayev street, Oskemen, 070004, Kazakhstan*
[2]*S. Seifullin Kazakh Agrotechnical Research University, 62 Zhenis avenue, Astana, Akmola, 010011, Kazakhstan*
[*]*E-mail*: zhidrisheva@edu.ektu.kz (Zhanat Idrisheva),
zthasenova@mail.ru (Zarına Khassenova),
zh.beldeubayeva@gmail.com (Zhanar Beldeubayeva)

menstruation or parasitic diseases (such as hookworm), chronic inflammation, as well as malaria and other infections that suppress the formation of red blood cells.[7,8] In children, it provokes stunting, weakened immune defenses, and lifelong cognitive impairments, with a 5-10—point drop in IQ and a 20-30% decrease in school performance.[9] Among pregnant women, anemia increases the likelihood of preeclampsia, premature birth, and low birth weight, increasing perinatal mortality by 2-3 times.[10]

The economic damage caused by anemia is enormous: according to estimates by WHO and the World Bank, it annually leads to losses of 0.6% of the GDP of developing countries, which globally is equal to 25-50 billion US dollars, mainly due to a decrease in labor efficiency by 10-15% and an increase in medical costs.[6,11] In low-income countries, including India, Bangladesh, and Ethiopia, anemia affects up to 50% of women of fertile age, exacerbating gender inequality and slowing demographic shifts.[12]Moreover, in the era of climate change and urbanization, the situation is deteriorating: droughts and floods destroy agricultural land, limiting access to iron-rich foods, and moving to megacities changes the diet towards carbohydrates with low nutritional value.[13]

Although effective preventive measures such as iron supplements, flour fluoridation, and fortification programs are readily available, their implementation remains insufficient: in Africa, only less than 20% of pregnant women receive complete supplementation.[14] This highlights the importance of comprehensive strategies combining nutrition optimization, infection control, and modernization of health systems within the framework of the United Nations Sustainable Development Goals (SDGs 2 and SDGs 3).[15]A s a result, anemia appears not only as a medical, but also as a socio-economic catastrophe requiring urgent international measures to eradicate it by 2030.[16]

Diagnosis of anemia in clinical practice remains fragmented and depends heavily on subjective interpretation of laboratory values. Morphological differentiation of anemias - microcytic, macrocytic and normocytic - requires analysis of not one but a whole set of parameters: mean corpuscular volume (MCV), mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC), red blood cell count (RBC), hemoglobin (HGB), hematocrit (HCT), as well as biochemical and immunological markers, including ferritin (Fer) and vitamin B12. However, in real practice, interpretation is often limited to hemoglobin level only, which leads to diagnostic errors and suboptimal choice of therapy.

Iron deficiency anemia (IDA), for example, is an early marker of gastrointestinal cancers, particularly colorectal cancer. However, the results of a national survey among 325 primary care physicians in the United States showed that only 73.5% of respondents correctly interpreted the laboratory values of HDA, especially for borderline values of ferritin and transferrin saturation.[4] This indicates the existence of “gray areas” in manual diagnosis that require digitalization. Even in the presence of formalized diagnostic criteria, the accuracy and reproducibility of manual interpretation remain limited, especially in mixed forms of anemia. Therefore, there is an increasing need to implement intelligent systems that can integrate with electronic medical records and use a comprehensive interpretation model. At the same time with the increasing amount of available laboratory data, the relevance of machine learning (ML) and data mining techniques is increasing.[17] ML methods allow automating the classification process by learning from precedents and identifying hidden patterns. Classical algorithms such as logistic regression, support vector method (SVM), decision trees and k-nearest neighbors (KNN) algorithm are widely used in medical informatics. Several works.[18,19] have demonstrated their effectiveness in morphological classification of anemias with accuracy above 90%.[20] However, such models are sensitive to unbalanced classes and noise in the data, which limits their clinical applicability without adaptation.

Ensemble methods (bagging, boosting, stacking, voting) provide more robust and accurate results. Algorithms like XGBoost and LightGBM demonstrate accuracy up to 98-99%.[21-23] while allowing quantification of the contribution of each feature to the classification, which is critical for clinical interpretation. Stacking models combine the strengths of several basic algorithms using a meta-classifier, as demonstrated by models combining Random Forest, KNN and gradient bousting with logistic regression as the unifying layer.[24] For the task of predicting anemia in pregnant women, for example, homogeneous ensembles based on CatBoost, Random Forest and XGBoost have been used, showing accuracies ranging from 91.34% to 97.44% depending on the chosen class decomposition strategy.[25,26] This demonstrates the potential of ensemble approaches in the tasks of not only classification but also risk stratification.

Anemia, characterized by a decrease in hemoglobin or red blood cell count, is responsible for more than 115,000 maternal deaths and about 591,000 perinatal deaths worldwide each year. According to the World Health Organization classification, anemia in pregnant women is divided into mild (Hb 10-10.9 g/dL), moderate (Hb 7.0-9.9 g/dL), and severe (Hb < 7 g/dL). In a study,[27] the authors used homogeneous ensemble machine learning algorithms to estimate the prevalence of anemia among pregnant women in Ethiopia. The data were pre-processed to ensure their suitability for machine analysis and building a model that predicts the degree of

anemia. Decision tree, random forest, CatBoost and extreme gradient bousting (XGBoost) algorithms were applied to create the predictive model, both with and without class decomposition strategies ("one versus one" and "one versus all"). Without class decomposition, CatBoost showed the highest accuracy at 97.08%, followed by XGBoost at 94.26%, and Random Forest at 91.34%. When one-to-one decomposition was applied, the accuracy was 97.44 % for CatBoost, 95.21 % for XGBoost, and 94.4 % for Random Forest.

Modern research in the field of anemia diagnostics increasingly includes deep neural networks (NN),[28-30] recurrent[31,32] and convolutional architectures[33] trained on extended datasets. Neural networks are highly sensitive to complex and hidden patterns, but they are traditionally criticized for the "opacity" of solutions. In response, the explicable AI (XAI) direction is developing, including the methods SHAPE (SHapley Additive Explanations),[34] LIME (Local Interpretable Model-agnostic Explanations),[35] attention mechanisms, and visualization of feature contributions to prediction. Such tools make it possible to integrate neural network conclusions into the clinical process while maintaining the trust of the doctor.

Thus, the analysis of the literature shows a steady trend towards the transition from manual interpretation of hematological parameters to intelligent diagnostic systems based on machine learning and neural network modeling methods. The most effective solutions combine interpretability, adaptability and modular architecture, enabling integration into clinical practice, including electronic health passports and decision support systems. The present study is aimed at the development, software implementation and experimental validation of a modular hardware and software complex focused on the automated morphological classification of anemia and the formation of a primary diagnostic conclusion based on quantitative and probabilistic criteria. The proposed system, integrated into the digital Health Passport platform, combines mathematical modeling methods, ensemble machine learning algorithms, and visually customizable neural network architectures in a single clinical and analytical environment. The paper presents in detail the architecture of the platform, the algorithmic basis, experimental results and directions for further development.

2. Materials and methods

2.1 Modeling of business processes for diagnosing clinical and hematological syndromes

The process of diagnosing CHS, including various forms of anemia, was formalized in the form of a sequential digital outline covering all key stages: from obtaining primary laboratory data to a formalized conclusion based on the results of interpretation.[36] This approach ensures continuity, reproducibility and automation of the diagnostic cycle.

The methodology for modeling business processes for diagnosing clinical and hematological syndromes, implemented within the framework of this study, is implemented in 5 stages, which are schematically presented in Fig. 1.

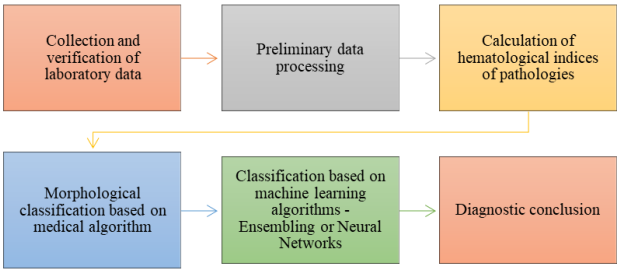

Fig. 1: Stages of clinical and hematological syndromes diagnostics in the Health Passport complex.

In the first stage - "Collection and verification of laboratory data" - clinical and laboratory information (hemoglobin, hematocrit, red blood cell count, ferritin, vitamin B12, etc.) is collected from hematology analyzers. The Data Acquisition module provides structured data entry, format control, and primary filtering of errors. At the second stage, the data is cleared of omissions and outliers, standardization of units of measurement and normalization (Data Preprocessing module). This step is critical for ensuring the correct operation of mathematical and machine models. At the "Calculation of hematological indices of pathologies" stage, on the basis of pre-processed data, derived risk indicators are calculated for all significant variables: HGB, HCT, RBC, MCV, MCH, MCHC as well as ferritin and B12 vitamin. At the fourth stage, the transformed data can be used to perform medical analysis using logical-threshold rules and fuzzy logic (Morphological Classification module), or to classify data using machine learning algorithms (Ensembling, Neural). The former implementation is based on morphological indicators and values of the M index, and is also refined using biochemical markers (ferritin, B12). module. In the latter, the user can forecast using ensemble or neural network models. The Ensembling module assembles model compositions (Boosting, Bagging, Voting, Stacking), while the Neural module provides a visual editor for neural networks. Both tools allow users to train and predict on new data. At the output, a structured output is formed, including:

- the presence or absence of anemia;
- cytometric type (micro-, normo-, or macrocytic);
- probabilistic estimates (in the case of using neural networks);
- concomitant characteristics (*e.g.*, the assumption of the B12-deficiency or iron-deficiency nature of anemia).

Each stage is logically and technically connected with the previous one. All activities are logged and auditable. Thus, a digital diagnostic business process has been implemented that complies with the principles of clinical validity, reproducibility and scalability.

Table 1: The role of specialists and DSS in the framework of the Health Passport software package.

Process Actors	Role and functions of actors
Technicians	Responsible for obtaining primary biomaterials and conducting hematological and biochemical analyses (HB, RBC, HCT, Fer, B12, etc.)
	Standardized output data tables are generated to be uploaded to the system through the Data Acquisition module
	Ensure quality control, compliance with asepsis standards, venipuncture techniques and standardization of procedures that affect the reliability of the results
Clinicians	The end users of diagnostic results
	Interpret the output of the system, form a clinical conclusion and therapeutic tactics
	The software package is used as a tool to *support the decision*, and not as a substitute for expert assessment
Intelligent System	Automates the processing of laboratory data, including preprocessing, index calculation, morphological and biochemical classification
	Provides an objective, reproducible and standardized interpretation of the results
	Provides a clinically oriented conclusion in the following format: the presence of anemia, its morphological and biochemical type, diagnostic confidence (in %)

2.2 Digital diagnostics of clinical and hematological syndromes

Digital diagnosis of clinical and hematological syndromes is based on a clear distribution of roles between clinical specialists, laboratory staff and an intelligent decision support system (DSS).[37] The role of these specialists and DSS is presented in Table 1.

Thus, the software package acts as an intelligent intermediary between the laboratory and the doctor, reducing the likelihood of errors associated with the human factor and speeding up the receipt of diagnostic solutions.

The transition to a digital model for the diagnosis of clinical and hematological syndromes is an objective necessity due to a number of key factors.

- Firstly, modern clinical diagnostic laboratories generate large amounts of data, which makes manual interpretation inefficient and error-prone.[38-40] Automation ensures the processing of information in real time without depending on the burden on personnel.
- Secondly, digitalization makes it possible to standardize the interpretation of laboratory test results, eliminating the subjective factor and increasing the reproducibility of diagnostic decisions.[41,42] A unified system of rules and formulas excludes inter-operator variability.
- Third, intelligent data processing systems provide the ability to quantify diagnostic confidence, especially in cases of mixed or borderline forms of anemia.[43,44] This expands the analytical capabilities of the physician and improves diagnostic accuracy.
- Fourth, the integration of the software package with electronic medical records and laboratory information systems makes it possible to build end-to-end digital circuits from the laboratory to the clinical solution, supporting the concept of a "smart hospital".[45]
- Fifth, the platform's modular structure and scalability allow it to adapt to a variety of settings, from outpatient appointments to telemedicine consultations and preventive screenings.

Thus, the digitalization of the diagnostic process is not just a technological improvement, but a necessary stage in the transition to evidence-based, objective and clinically reproducible medicine.

2.3 Formal diagnostic models

A key element of the intellectualized system of differential diagnosis of anemia is the use of formalized quantitative models based on the calculations of diagnostic indices and aggregated indicators. This enables reproducibility, interpretability, and automation of diagnostic decisions. Differential diagnosis of CHS is carried out on the basis of the calculation of basic hematological indices. Based on primary laboratory parameters (HGB, RBC, HCT), the following standardized indices are calculated: MCV, MCH, MCHC. The formulas for calculating these three standardized indices are as Eq. (1):

$$\begin{cases} MCV=\dfrac{HCT}{RBC} \\ MCH=\dfrac{HGB}{RBC} \\ MCHC=\dfrac{HGB}{HCT} \end{cases} \quad (1)$$

These indicators are universal markers of the morphological characteristics of erythrocytes and are used to subsequently determine the type of anemia. For a comprehensive assessment of the severity of anemic syndrome, an aggregated M indicator is used, which reflects the degree of deviation from the norm in six parameters. The calculation is carried out according to the weighted Eq. (2):

$$M=0{,}5\cdot mHGB+0{,}1\cdot mHCT+0{,}1\cdot mMCHC +0{,}1\cdot mMCH+0{,}1\cdot mMCV+0{,}1\cdot mRBC \quad (2)$$

where m_i is the degree of belonging of each indicator to the pathological range (fuzzified value from 0 to 1).

Thresholds and decision-making logic Eq. (3):

$$\begin{cases} M\geq 0{,}5 \text{ - anemia} \\ 0{,}2<M<0{,}5 \text{ - anemia possible} \\ M\leq 0{,}2 \text{ - no anemia} \end{cases} \quad (3)$$

This Eq. (3) is applied at the first stage of the analysis and determines the need for further morphological refinement.

Next, the procedure for determining the cytometric type of anemia is implemented. That is, if anemia is confirmed, classification is made by type:

- Microcytic anemia - decrease in MCH and MCV below the normal thresholds;
- Macrocytic anemia - MCH decreased, MCV increased;
- Normocytic anemia - index values within the normal range, but with deviations in biochemical markers (for example, reduced ferritin or B12);
- Mixed form - thresholds for more than one index are exceeded at the same time (for example, mmicro >0.5 and macro > 0.5).

At the final stage, biochemical information is used to clarify the nature of anemia by Eq. (4):

$$\begin{cases} \text{If Fer} < \text{threshold- iron deficiency risk} \\ \text{If B12} < \text{threshold - B12 deficiency risk} \\ \text{Else- chronic diseases risk} \end{cases} \quad (4)$$

These formalized criteria Eq. (4) are implemented in the MathematicalEvaluation module and passed to Morphological Classification for a final diagnostic conclusion.

2.4 Algorithmic base of classifiers

As part of the intellectualized diagnostic complex construction, a wide range of machine learning algorithms were used - both classical and ensembled ML, as well as neural network architectures. The approach is based on comparative modeling and selection of optimal solutions in terms of accuracy, interpretability and stability.

To solve the problems of binary and multiclass classification, the models presented in Table 2 were used.

To increase the generalizing ability of the models and resistance to noise, the following ensemble techniques were used:

- Voting (hard and soft voting) - aggregates the predictions of several models, choosing the most likely class;
- Bagging - training several models on different subsamples with subsequent averaging of their predictions;
- Boosting - sequential training of models with a focus on the errors of previous iterations (XGBoost, AdaBoost, LightGBM);
- Stacking is a multi-level architecture where the metamodel is trained on the outputs of the underlying models. Configuration with RF, DT and KNN at level 0 and Logistic Regression at level 1 was used.

Graphical models of these 4 algorithms for assembling machine learning methods are presented in Fig. 2.

All ensembles were tested using cross-validation and demonstrated comparable or better performance in accuracy, precision, recall and F1-score metrics compared to single models and traditional methodics.

Table 2: Models for solving problems of binary and multiclass classification of CHS.

Model	Rationale for application for research tasks
Logistic Regression	A simple and interpretable method that uses linear class separation and is suitable for basic feature sensitivity assessment
Support Vector Method (SVM)	Effective in high-dimensional problems and works well with small but informative samples
k Nearest Neighbors Method (kNN)	Intuitive model based on the observation distance metric
Decision Tree	An easy-to-interpret algorithm that forms a hierarchy of diagnostic rules
Random Forest	An ensemble of decisive trees, resistant to overtraining and effectively works with heterogeneous features
Gradient boosting (XGBoost)	One of the most productive algorithms on structured data, demonstrated the highest accuracy (up to 98.1%) in diagnosing anemia

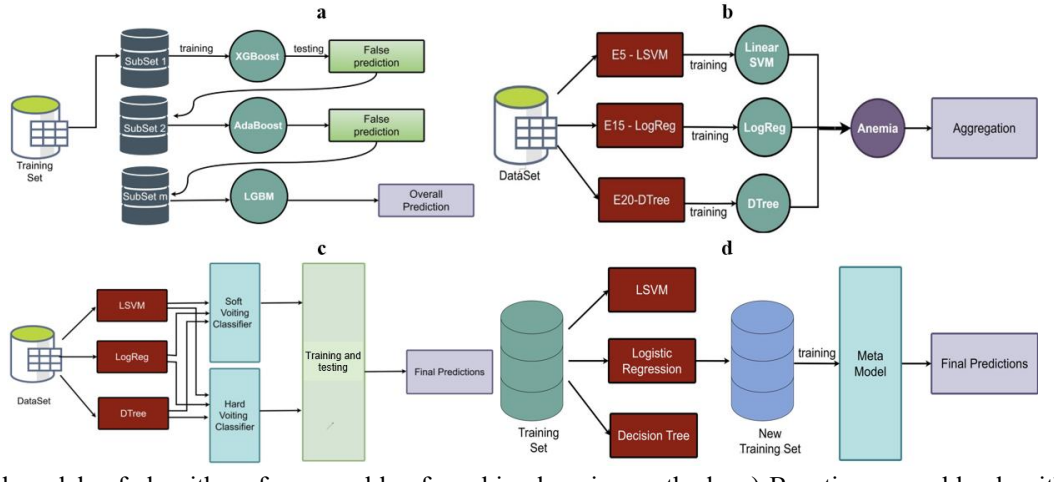

Fig. 2: Graphical models of algorithms for ensemble of machine learning methods: a) Boosting ensemble algorithm; b) Bagging ensemble algorithm; c) Voting ensemble algorithm; d) Stacking ensemble algorithm.

3. Results and discussion

3.1 Implementation of the software package architectural model

The architectural model of the software and hardware complex for diagnosing clinical and hematological syndromes for health care describes the structure and interaction between various components of the system. It includes software and hardware components that work together to collect, process, and analyze medical data.

The architectural model is implemented as a modular distributed system with centralized data management and a clear delineation of functional roles between components. The architecture is built on the principles of a microservice approach and is focused on ensuring high scalability, fault tolerance and compatibility with external information systems. The diagnostic platform is built as a modular and scalable software and hardware complex designed for intelligent decision-making support in the detection of clinical and hematological syndromes. The architecture of the system is based on the principles of interoperability, reproducibility, automation and clinical validity.

The platform is implemented on the principle of service-oriented architecture (SOA) with a clear division of functionality into modules. Each stage of the diagnostic process - from the acquisition of laboratory data to the final interpretation - is separated into a separate logically complete component. Interoperability between modules is provided through standardized APIs and a centralized relational database.

Both the classic client-server model (access via web interface and REST API) and the ability to deploy in the cloud infrastructure are supported, which allows the system to be used both at stationary clinical workplaces and on mobile devices while maintaining the requirements for security and user authentication.

The architecture is modular, that is, each component performs a separate function, which facilitates the maintenance and development of the system. All data must be protected using encryption and access control. The software package has the following key modules and functional blocks:

- Database;
- Data Acquisition analytical module;
- Data Preprocessing analytical module;
- Mathematical Evaluation computational model;
- Analytical Morphological Classification;
- Intelligent Ensembling module
- Analytical Neural module.

A centralized database based on PostgreSQL ensures the integrity, consistency and traceability of all data - from raw input to classification output. All modules are connected to a single DBMS and use secure RESTful APIs for data exchange.

The user interface is implemented as a web application and provides access to all the functions of the platform: loading data, setting up models, running calculations, and viewing results. The interface is adapted to the work of medical personnel and provides intuitive interaction with each stage of the diagnostic process. Fig. 3 shows the general layout of the architectural model.

The server infrastructure of the model is designed for storing and processing data. It includes multiple tiers of servers (databases, web servers, and applications). Data storage systems are used for long-term storage of large amounts of data by CHS. The infrastructure ensures high availability and fault tolerance, enabling continuous access to diagnostic tools. Scalable cloud-based or on-premise deployment options can be implemented depending on institutional needs. Client devices such as doctors' workstations and mobile devices are used to interact with the Health Passport software package. These devices allow clinicians to access patient records, enter new clinical data, and review analytical outputs directly at the point of care. This ensures seamless workflow integration and real-time decision-making support.

The Health Passport software package has an intuitive

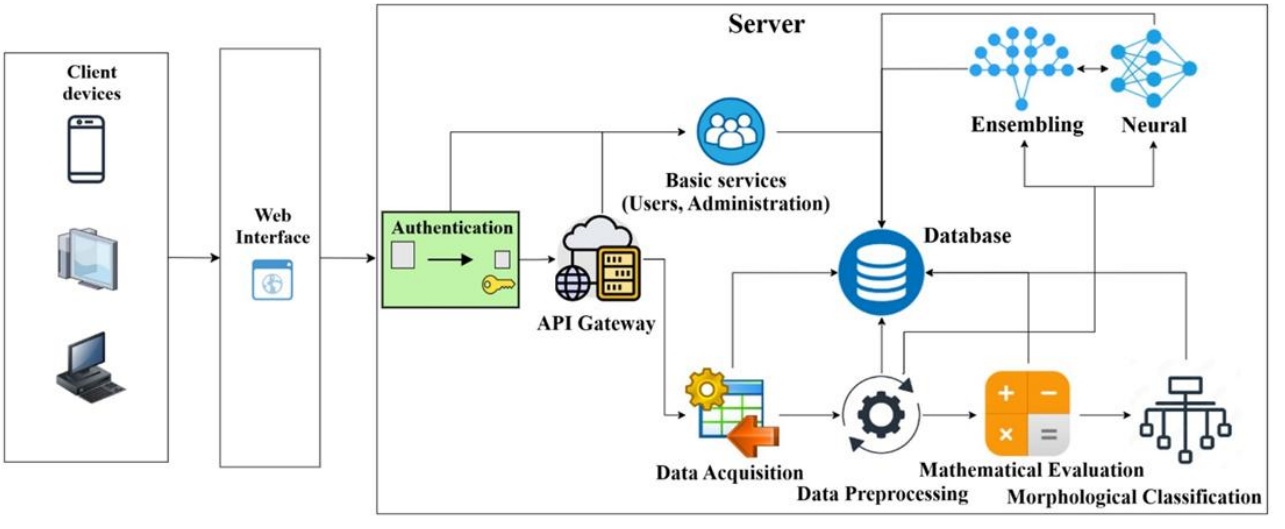

Fig. 3: Architectural model of the Health Passport software package.

interface for doctors that allows them to enter data, receive test results and recommendations. The user interface includes visualization tools such as trend charts and heatmaps, improving the interpretability of diagnostic outputs. It is optimized for use in clinical environments with minimal training required. To ensure information security, all modules of the architectural model contain appropriate security measures, including data encryption, access control, and monitoring of user activity. Compliance with healthcare data protection standards is ensured. Audit logs are maintained to track system usage and prevent unauthorized access.

The model is a client-server model with a microservice architecture. That is, client devices (doctors) interact with the server, which processes requests and provides data. Each module can be implemented as a separate service, making it easier to manage, update, and scale the system. This architectural approach improves system modularity and fault isolation, allowing independent deployment and maintenance of each functional block. It also facilitates integration of new analytical modules or AI models without affecting the core system. It is planned to consider the integration of the software complex with other medical information systems and standards. Such integration may include HL7, FHIR, and DICOM standards, enabling interoperability with hospital EHR systems and diagnostic devices. This would enhance data exchange and promote standardized workflows across clinical institutions.

Thus, the architecture of the platform provides an end-to-end, automated and reproducible cycle of analysis of clinical and hematological data with the ability to scale, integrate and adapt to specific clinical tasks. Its modular and extensible design makes it suitable for deployment in both centralized hospitals and remote healthcare settings, supporting personalized and population-wide diagnostics alike.

3.2 Modular implementation of the software package

The modular architecture of the Health Passport platform provides a logical decomposition of the diagnostic process into autonomous computational components. Each module is responsible for a clearly defined stage in the clinical and hematological evaluation pipeline, enabling transparency, traceability, and reusability of individual operations. This architectural solution is especially important in the medical domain, where reproducibility, explainability, and data integrity are critical. The entry point for system interaction is the main interface window, which serves as the control panel for both technical and medical users. From this central hub, users can access various diagnostic modules—each designed to process, evaluate, and visualize specific types of data or perform essential operations. Specialized components perform morphological classification of anemia subtypes based on calculated indices, while machine learning and neural network modules provide predictive analytics and probabilistic inference. Importantly, the architecture is built around a secure, centralized database that ensures reliable storage of input parameters, intermediate calculations, and final diagnostic outputs. The database supports real-time communication with frontend elements, allowing clinicians to view, compare, and validate results through an intuitive graphical user interface.

Security, user authentication, and role-based access control are also core parts of the system architecture. This makes the platform suitable for deployment not only in research or laboratory settings but also in clinical environments where strict compliance with data privacy regulations is mandatory. Furthermore, the microservice-based structure allows for modular updates and scalability without affecting the core analytical capabilities of the system. Fig.4 illustrates the main window of the Health Passport software package, where each module is represented as an interactive card, allowing users to launch tasks, manage patient records, and initiate diagnostic workflows with minimal training or technical background

To ensure modularity, transparency, and scalability in the diagnostic workflow, the Health Passport software package is

Fig. 4: Main window of the Health Passport software package.

Table 3: Key modules and functional blocks of the health passport software package.

Module / Unit	Purpose
Database	Centralized storage of all stages of the diagnostic process. Structured by entities: users, transactions, source data, computed indexes, classification results, etc. Ensures data traceability, protection, and consistency.
Data Acquisition	Responsible for the structured import of laboratory data from external sources (*e.g.*, hematology analyzers). Provides format validation, structure definition, transaction logging, and visual preview of data before loading.
Data Preprocessing	Performs intelligent data cleanup: outlier removal, missing value correction, numeric normalization, format casting, and categorical feature encoding. Prepares data for analytical modules.
Mathematical Evaluation	It implements the calculation of derived diagnostic indices (MCV, MCH, MCHC, anemia M index, micro/macro/normocytic coefficients, etc.) based on formalized mathematical models.
Morphological Classification	Performs morphological classification of anemias based on threshold values of diagnostic indices. Uses the rules of logical interpretation and fuzzy logic. Includes biochemical clarification of the diagnosis using ferritin and vitamin B12.
Ensembling	Responsible for building, training, and applying ensemble machine learning models (Boosting, Bagging, Voting, Stacking). Provides configuration of the architecture, selection of models and target variables.
Neural	Visual editor of neural networks. Allows to configure input and output parameters, configure hidden layers, select activation functions and hyperparameters, as well as train the model and make predictions. Shares the dataset pool with Ensembling.

organized into several key functional units. Each module plays a specific role in the processing, transformation, and interpretation of clinical and hematological data. The modular structure follows best practices in software architecture by adhering to the separation of concerns principle, ensuring that each stage of the diagnostic pipeline can be independently validated, updated, or replaced without affecting the entire system. This approach not only improves maintainability but also enables targeted enhancements, such as integrating new machine learning models or adapting to evolving clinical guidelines. From initial data acquisition and cleaning to mathematical evaluation, morphological classification, and advanced analytics via ensemble and neural models, the system forms a cohesive end-to-end environment for intelligent anemia diagnostics. The synergy between these components enables both high-throughput processing and fine-grained clinical decision support. The key modules and their functions are summarized in Table 3 below.

a) Clinical and laboratory data collection module. The Data Acquisition module is designed for structured collection of clinical and laboratory data from external sources, including hematology analyzers such as Mindray BC-5800. Mindray BC-5800 is a mid-range automatic hematology analyzer designed for quantitative and qualitative analysis of blood cells using laser light scattering flow cytometry technology. The device provides an extended set of parameters (up to 29 indicators), including leukocyte differentiation into five populations, which allows it to be used in clinical diagnostic laboratories of various profiles. This analyzer supports standardized data exchange protocols (*e.g.*, ASTM and HL7), enabling direct integration with laboratory information systems and clinical registries. High measurement speed (90 tests per hour), as well as convenient, intuitive control, optimizing the working procedure of blood analysis. In the Data Acquisition module, format validation, transaction logging and storage of input data to the database are performed. The user can upload a data file (XLSX/CSV), view its structure, delete previously uploaded transactions, and proceed to processing. Raw and processed data preview modes are supported. Each transaction is stored in a database table with a unique identifier, timestamp, and association with a specific user. The data is passed to the Data Preprocessing module through the associated transaction ID.

b) Data Preprocessing Module. The Data Preprocessing module provides data cleansing: removing rows with missing or incorrect values (*e.g.*, missing age or gender), removing outliers, normalizing numeric features, and standardizing formats. The windows of this module display the loaded transactions and the status of their processing. The user can start preprocessing, compare input and processed data, and control the correctness of filtering. Implemented preview of changes in the table. The processed data is written to a separate table while maintaining the link to the original transaction. A uniform format suitable for submission to analytical modules is formed.

c) Mathematical Evaluation Module. In the Mathematical Evaluation module, the derived hematological indices are automatically calculated: mMCV, mMCH, mMCHC, mRBC, mHGB, as well as the aggregate anemia index (M). The formulas are implemented on the basis of clinically verified calculation models. The module interface displays a list of processed transactions, allows to start a calculation, view intermediate and final values, and verify calculation formulas. The results of the calculations are stored in a specialized table with separate storage of raw indices and aggregated coefficients. This data is further used in the classification modules.

d) Morphological Classification Module. The Morphological Classification module implements threshold and fuzzy interpretation of diagnostic indices. It gives a conclusion about the morphological type of anemia: microcytic, macrocytic, normocytic, or mixed. If necessary, biomarker clarification (ferritin, B12) is used. The user can view the classification results, including the values of all subindices (Mmicro,

Mmacro, etc.) and the rationale for the diagnosis. Export and storage of results in a separate table associated with the original transaction and calculated indices is supported. The nature of anemia is fixed in terms of binary affiliation and probabilistic expression.

e) Ensembling Module. The Ensembling module is designed for the visual design of model ensembles. Supports ensemble type selection (Boosting, Bagging, Voting, Stacking), hyperparameter setting, target and input variable selection. Provides learning and prediction. This module provides three tabs: ensemble architecture, dataset loading, training, and forecasting. The user can edit, save and run models, see quality metrics (accuracy, F1, etc.). Models, configurations, and trained weights are stored in the database. Related predicted datasets are labeled and displayed separately, indicating the model that performed the prediction.

f) Neural Module. The Neural module is a graphical editor of neural networks. The user manually defines the network architecture, including the number of layers and neurons, activation functions, and training parameters. Visual control of training is provided. The following tabs are available in this module: "Architecture", "Hyperparameters", "Dataset Upload", "Train and Predict". The following functions are supported in the module: saving the model, loading data, starting training, log console, viewing accuracy. The network configuration, trained weights, and training history are stored in a related table. The output is stored as probabilistic predictions (*e.g.*, the probability of anemia and its morphological subtypes).

3.3 Technology stack

The development and implementation of the Health Passport software package is based on the use of a modern and flexible technological stack that provides high performance, scalability, security and convenience of user interaction.

JavaScript/TypeScript (Node.js) was used for both client and server development, with React.js applied in the front-end part, allowing for a reactive and interactive web interface. Machine learning modules are implemented in Python, using the following libraries:

- Scikit-learn - implementation of basic algorithms for classification, regression and ensemble;
- XGBoost and LightGBM - gradient boosting for high-precision structured classification tasks;
- TensorFlow and Keras - for building, training and implementing neural network models of various architectures, including deep and convolutional networks;
- Pandas, NumPy, Matplotlib, Seaborn - for data analysis, preprocessing and visualization.

The central repository of information is the PostgreSQL relational database, designed using Sequelize ORM. The system implements the following measures to protect personal data:

- access rights control;
- hashing of authentication passwords;
- logging of transactions and user operations;
- data partitioning;
- support for connection-level encryption (SSL/TLS) and database configuration.

The system complies with the requirements for the protection of personal data and can be adapted to the internal standards of medical institutions.

Visual interaction is implemented through an interactive web interface, including dynamically updated components, visual model editors and reports in tabular and graphical form. Interaction with the backend is provided through a RESTful API, which allows to:

- automate workflows;
- integrate the system with external medical systems and screening portals;
- Enable interaction with mobile and cloud clients.

Thus, the technological stack of the system is focused not only on scientific reliability and computational efficiency, but also on practical applicability in a clinical environment, with an emphasis on safety and user comfort. To assess the performance, stability and reproducibility of all components of the "Health Passport" intelligent software package, step-by-step testing of each module was carried out on real and synthetically generated clinical and hematological datasets.

3.4 Results of testing the modules of the software package

The Data Acquisition module plays a critical role in the functioning of the Health Passport software package by serving as the initial entry point for clinical and laboratory information. This module is specifically tailored to support flexible and intuitive input of patient hematological parameters, enabling healthcare professionals to seamlessly feed the system with essential diagnostic data. Its robust design allows for compatibility with a wide variety of table formats, including structured spreadsheets, manual entry forms, and integrated electronic health record exports. The module includes intelligent field validation to prevent incorrect or incomplete data submissions, significantly reducing the risk of diagnostic errors due to input inconsistencies. A particularly useful feature of the Data Acquisition module is its multilingual support and simple interface, which lowers the entry barrier for medical staff across different regions and levels of digital literacy. The software is capable of interpreting both numeric and text-based categorical inputs, automatically converting them into standardized forms suitable for downstream processing. It also accommodates missing values and flags them for review or imputation, thus maintaining data integrity. Moreover, the module stores historical records in a searchable format, allowing clinicians to access and analyze past diagnostic results for individual patients over time, thereby supporting longitudinal tracking and continuity of care.

This functionality not only facilitates the initial data entry but also enables efficient patient management by linking current inputs with previous diagnoses. Such integration is

vital for evidence-based decision-making and monitoring disease progression or treatment response. The user interface is divided into several key sections, as illustrated in Fig. 5.

Panel (a) presents the input fields for diagnostic parameters such as hemoglobin, hematocrit, red blood cell count, ferritin, and vitamin B12 levels. Panel (b) displays the patient search bar used to retrieve diagnostic history, making it easy to track individual medical cases and enhance the accuracy of future predictions.

The Patient Card serves as a centralized interface for viewing the diagnostic history and demographic profile of each patient. It provides a structured and clinically intuitive layout, enabling healthcare professionals to analyze a patient's hematological data across multiple time points. This functionality supports longitudinal tracking of diagnostic outcomes, treatment effects, and disease progression—an essential feature for chronic or recurrent conditions such as anemia.

When a clinician selects a specific patient from the Data Acquisition search window (as shown earlier), the system retrieves and displays all available diagnostic records in a side-by-side format. Each record includes key blood test indicators such as hemoglobin, hematocrit, red blood cell count, ferritin, and B12 levels. Additionally, calculated blood indices like MCH, MCHC, and MCV are automatically displayed to aid interpretation.

The results are then classified by the system using embedded algorithms that determine both the presence and type of anemia. Labels such as "Microcytic type," "Normocytic type," and "Iron-deficiency anemia" are generated based on index thresholds and clinical rules. This structured presentation allows physicians to quickly assess the patient's condition and verify consistency across multiple test entries. Furthermore, each diagnosis entry can be individually deleted if needed, offering flexibility in data management. Patient-specific metadata such as name, year of birth, and gender is clearly displayed in the header, facilitating accurate recordkeeping. This intelligent visualization, shown in Fig. 6, highlights the system's ability to provide not just raw laboratory data, but also clinically relevant summaries and decision support outputs for each patient in an accessible format.

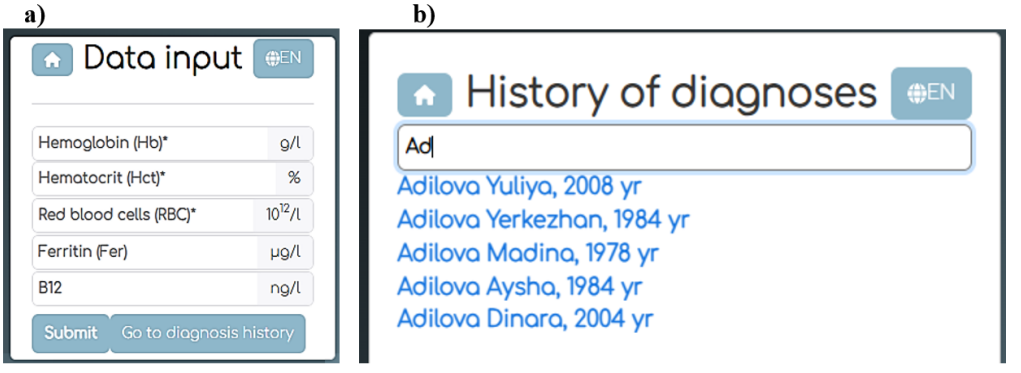

Fig. 5: The Data Acquisition module window: a) the window for entering the data to be diagnosed; b) Patient Card Search Box.

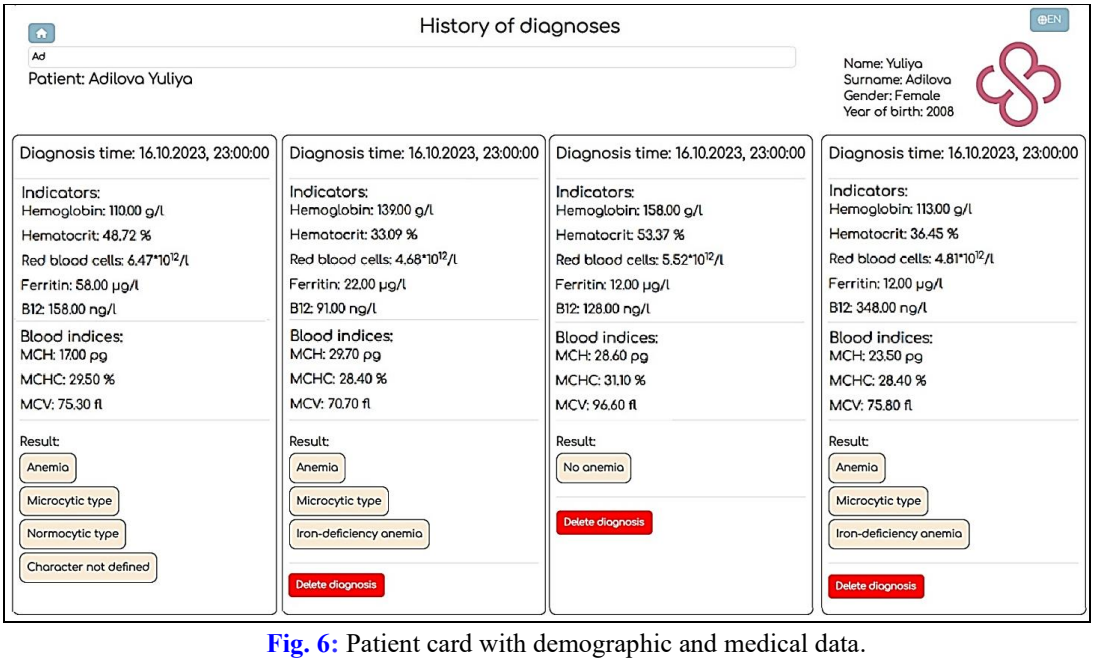

Fig. 6: Patient card with demographic and medical data.

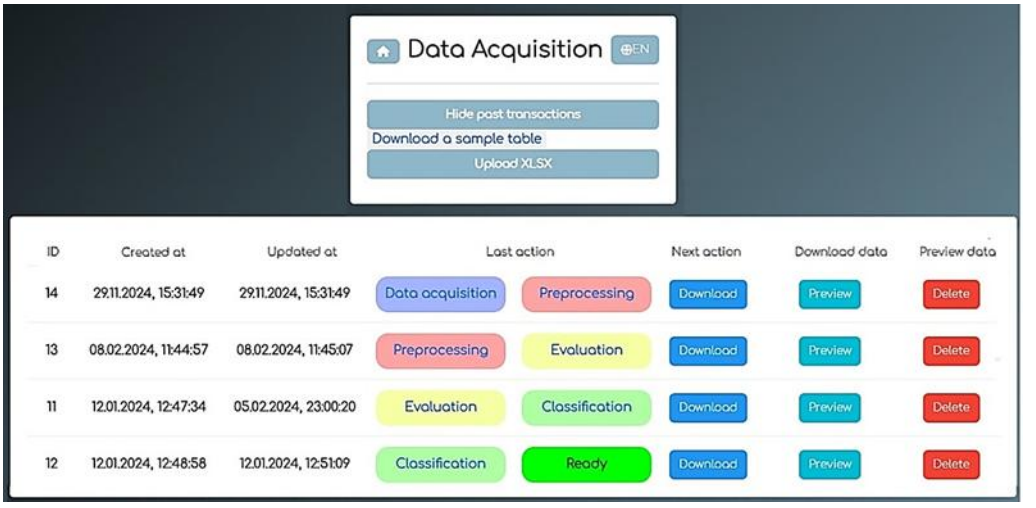

Fig. 7: Main window and list of past transactions with their functions.

	id [PK] integer	trans_id integer	gender boolean	age integer	HB double precision	RBC double precision	HCT double precision	Fer double precision	B12 double precision	notes text	age_group character varying (255)	createdAt timestamp with time zone	updatedAt timestamp with time zone
1	6960	26	true	25	149	6.563876652	0.549396476	42	403	10.17.2023	19-30	1970-01-01 06:00:45.3+06	1970-01-01 06:00:45.3+06
2	6961	26	true	15	147	6.681818182	0.664840909	56	456	10.17.2023	0-18	1970-01-01 06:00:45.3+06	1970-01-01 06:00:45.3+06
3	6962	26	true	10	116	5.201793722	0.387533632	68	411	10.17.2023	0-18	1970-01-01 06:00:45.3+06	1970-01-01 06:00:45.3+06
4	6963	26	false	17	141	4.747474747	0.357010101	45	107	10.17.2023	0-18	1970-01-01 06:00:45.3+06	1970-01-01 06:00:45.3+06
5	6964	26	false	10	167	6.072727273	0.564763636	14	70	10.17.2023	0-18	1970-01-01 06:00:45.3+06	1970-01-01 06:00:45.3+06
6	6965	26	true	92	134	5.317460317	0.509944444	42	209	10.17.2023	51+	1970-01-01 06:00:45.3+06	1970-01-01 06:00:45.3+06
7	6966	26	false	17	147	5.08650519	0.355038062	45	401	10.17.2023	0-18	1970-01-01 06:00:45.3+06	1970-01-01 06:00:45.3+06
8	6967	26	true	50	159	6.543209877	0.598703704	36	144	10.17.2023	51+	1970-01-01 06:00:45.3+06	1970-01-01 06:00:45.3+06
9	6968	26	false	11	159	6.625	0.5545125	48	486	10.17.2023	0-18	1970-01-01 06:00:45.3+06	1970-01-01 06:00:45.3+06
10	6969	26	false	25	154	6.260162602	0.590333333	12	325	10.17.2023	19-30	1970-01-01 06:00:45.3+06	1970-01-01 06:00:45.3+06
11	6970	26	false	32	162	9.418604651	0.738418605	61	183	10.17.2023	31-50	1970-01-01 06:00:45.3+06	1970-01-01 06:00:45.3+06
12	6971	26	false	3	118	7.239263804	0.569730061	41	159	10.17.2023	0-18	1970-01-01 06:00:45.3+06	1970-01-01 06:00:45.3+06
13	6972	26	true	5	110	4.230769231	0.418423077	32	224	10.17.2023	0-18	1970-01-01 06:00:45.3+06	1970-01-01 06:00:45.3+06
14	6973	26	true	25	147	6.282051282	0.535858974	42	101	10.17.2023	19-30	1970-01-01 06:00:45.3+06	1970-01-01 06:00:45.3+06
15	6974	26	true	13	143	5.933609959	0.576746888	69	287	10.17.2023	0-18	1970-01-01 06:00:45.3+06	1970-01-01 06:00:45.3+06
16	6975	26	false	45	110	6.470588235	0.487235294	58	158	10.17.2023	31-50	1970-01-01 06:00:45.3+06	1970-01-01 06:00:45.3+06
17	6976	26	false	21	149	8.465909091	0.859289773	57	477	10.17.2023	19-30	1970-01-01 06:00:45.3+06	1970-01-01 06:00:45.3+06
18	6977	26	true	31	145	5.823293173	0.476345382	33	165	10.17.2023	31-50	1970-01-01 06:00:45.3+06	1970-01-01 06:00:45.3+06
19	6978	26	false	105	128	6.037735849	0.442566038	38	161	10.17.2023	51+	1970-01-01 06:00:45.3+06	1970-01-01 06:00:45.3+06
20	6979	26	false	56	152	5.409252669	0.418135231	37	259	10.17.2023	51+	1970-01-01 06:00:45.3+06	1970-01-01 06:00:45.3+06
21	6980	26	false	36	167	8.028846154	0.683254808	13	176	10.17.2023	31-50	1970-01-01 06:00:45.3+06	1970-01-01 06:00:45.3+06
22	6981	26	true	59	113	4.232209738	0.392325843	66	350	10.17.2023	51+	1970-01-01 06:00:45.3+06	1970-01-01 06:00:45.3+06
23	6982	26	false	30	153	8.225806452	0.794612903	43	240	10.17.2023	31-50	1970-01-01 06:00:45.3+06	1970-01-01 06:00:45.3+06
24	6983	26	true	81	144	5.12455516	0.477608541	45	441	10.17.2023	51+	1970-01-01 06:00:45.3+06	1970-01-01 06:00:45.3+06

Fig. 8: Table with processed data.

Fig. 7 shows the module's main interface, which displays a list of loaded transactions. The user can view the status, go to the preview, or delete the transaction if necessary. Both loaded data in its original form and its extended format after automatic calculations are supported.

The module then visualizes the processed data using pre-calculated indicators, allowing for a quick assessment of the table structure's correctness and the next step. Viewing the raw data allows the user to verify the original values before pre-processing, as shown in Fig. S1 in the supplementary section. The transaction deletion mechanism is also tested, as shown in Fig. 7, where all associated information is securely deleted. During the data pre-processing stage, the system's ability to effectively handle outliers, omissions, and format and logical errors was confirmed in the Data Preprocessing module. After running the procedure, a table is generated in which the values are normalized, encoded, and standardized. Fig. S2 in the supplementary section shows the software package window, allowing users to examine the data before completing the loading process.

Fig. 8 shows a summary table that can be fed into the analytical modules without the need for manual rework.

The «Mathematical Evaluation» module is a key component of the Health Passport software, responsible for performing analytical operations on preprocessed clinical and hematological data. Its primary function is the calculation of diagnostic indices, which serve as intermediate and final results in the anemia classification process. This module's interface is designed for simplicity and ease of use. Each session is tracked with a unique identifier and timestamp, ensuring full traceability of the data creation time, last update, and actions performed. The user can easily monitor the progress of the workflow thanks to clearly defined steps such as «Preprocessing», «Evaluation», and «Classification», which are dynamically highlighted depending on the current processing stage. For each record, the user is presented with action buttons: «Preview» for viewing the processed dataset, and «Evaluation» for running or repeating the mathematical calculation of diagnostic scores and parameters. The main view of the module with a list of transactions and their

functions is shown in Fig. S3 in the supplementary section.

The modular structure of the software package ensures reproducibility of results, which is extremely important in clinical settings, by maintaining fixed pipelines and preserving the integrity of intermediate calculations. Furthermore, the system demonstrates resilience to borderline cases, such as missing or extreme values, thanks to integrated validity checks and backup entry strategies developed during the preprocessing stage. The user interface ensures transparency and consistency, allowing medical professionals to focus on diagnostic interpretation rather than technical execution. Thus, the mathematical evaluation module represents a reliable and user-friendly tool for evidence-based hematological analysis, providing physicians with actionable insights while ensuring scientific rigor and operational traceability.

After the mathematical evaluation process is completed, the system generates a comprehensive preview table containing the calculated diagnostic indicators. This table provides a detailed breakdown of hematological parameters including MCH, MCHC, MCV, as well as derived indices like MCV/MCH ratio, MCV/MCHC ratio, and MCV index, is shown in Fig. S4 in the supplementary section. These metrics are essential for classifying anemia subtypes such as microcytic, macrocytic, and normocytic forms. The table also includes a series of binary flags and percentage-based scores that quantify the likelihood of specific anemia categories. These diagnostic outputs are automatically generated based on predefined thresholds and rules rooted in clinical hematology. Each row in the preview corresponds to a patient record, with data fields that include personal annotations (*e.g.*, «Female, age 20») and the results of calculated indices. Visual cues such as progress rings or color-coded highlights are used to indicate abnormal or borderline values, for example, a red circle around a percentage value suggests a flagged metric that may require further clinical interpretation. This interactive and dynamic visualization enables users to quickly assess hematologic status, identify pathological patterns, and verify the internal consistency of the evaluation results. The ability to interpret results at both the individual and population level enhances the system's applicability in clinical decision-making and longitudinal patient monitoring.

Fig. 9 demonstrates a full-scale export of processed data in tabular form, featuring both original clinical inputs and derived indices. This output is presented in Excel format, offering medical professionals and researchers a transparent and verifiable way to analyze the results of automated hematological assessments. Each row corresponds to an individual patient and includes demographic fields such as gender, age, and notes, followed by primary laboratory parameters like HGB, RBC, HCT, Fer, and vitamin B12 levels.

The Fig. 9 also includes calculated values for key red blood cell indices: MCH, MCHC, and MCV, as well as their complex ratios (*e.g.*, MCV/MCH, MCH/MCHC), which help identify specific anemia types, such as microcytic, macrocytic, and normocytic. Flags for various diagnostic rules are clearly marked with binary codes (0 or 1), supporting rule-based classification and input generation for machine learning

id	gender	age	HB	RBC	HCT	Fer	B12	notes	age_group	MCH	MCHC	MCV	MCH_mac	MCV_mac	MCH_micr	MCV_micr	mMCH	mMCHC	mMCV	mHB	mRBC	mHCT	M	mMacro	mMicro	mNormo	mAHZ	mZhDA	mB12
6224	1	25	149	6.56388	0.5494	42	403	10.17.202	19-30	22.7	27.1207	83.7	0	0	0.50588	0	0.50588	1	0	0	0	0	0.15059	0	0.25294	0.50588	0.1	0	0
6225	1	15	147	6.68182	0.66484	56	456	10.17.202	0-18	22	22.1106	99.5	0	0.13235	0.58824	0	0.58824	1	0	0	0	0	0.15882	0.06618	0.29412	0.58824	0.8	0	0
6226	1	10	116	5.20179	0.38753	68	411	10.17.202	0-18	22.3	29.9329	74.5	0	0	0.55294	0.34375	0.55294	0.51678	0.34375	0.9	0	0	0.59135	0	0.44835	0.55294	1	0	0
6227	0	17	141	4.74747	0.35701	45	107	10.17.202	0-18	29.7	39.4947	75.2	0	0	0	0.3	0	0	0.3	0	0	0.28737	0.05874	0	0.15	0	0.25	0	0.97667
6228	0	10	167	6.07273	0.56476	14	70	10.17.202	0-18	27.5	29.5699	93	0	0	0	0	0	0.60753	0	0	0	0	0.06075	0	0	0	0	1	1
6229	1	92	134	5.31746	0.50994	42	209	10.17.202	51+	25.2	26.2774	95.9	0	0.02647	0.21176	0	0.21176	1	0	0	0	0	0.12118	0.01324	0.10588	0.21176	0.1	0	0.63667
6230	0	17	147	5.08651	0.35504	45	401	10.17.202	0-18	28.9	41.404	69.8	0	0	0	0.6375	0	0	0.6375	0	0	0	0.09495	0	0.31875	0	0.25	0	0
6231	1	50	159	6.54321	0.5987	36	144	10.17.202	51+	24.3	26.5574	91.5	0	0	0.31765	0	0.31765	1	0	0	0	0	0.13176	0	0.15882	0.31765	0	0.2	0.85333
6232	0	11	159	6.625	0.55451	48	486	10.17.202	0-18	24	28.6738	83.7	0	0	0.35294	0	0.35294	0.83154	0	0	0	0	0.11845	0	0.17647	0.35294	0.4	0	0
6233	0	25	154	6.26016	0.59033	12	325	10.17.202	19-30	24.6	26.087	94.3	0	0	0.28235	0	0.28235	1	0	0	0	0	0.12824	0	0.14118	0.28235	0	1	0.25
6234	0	32	162	9.4186	0.73842	61	183	10.17.202	31-50	17.2	21.9388	78.4	0	0	1	0.1	1	1	0.1	0	0	0	0.21	0	0.55	1	1	0	0.72333
6235	0	3	118	7.23926	0.56973	41	159	10.17.202	0-18	16.3	20.7116	78.7	0	0	1	0.08125	1	1	0.08125	0.7	0	0	0.55813	0	0.54063	1	0.05	0	0.80333
6236	1	5	110	4.23077	0.41842	32	224	10.17.202	0-18	26	26.2892	98.9	0	0.11471	0.11765	0	0.11765	1	0	1	0	0	0.61176	0.05735	0.05882	0.11765	0	0.4	0.58667
6237	1	25	147	6.28205	0.53586	42	101	10.17.202	19-30	23.4	27.4326	85.3	0	0	0.42353	0	0.42353	1	0	0	0	0	0.14235	0	0.21176	0.42353	0.1	0	0.99667
6238	1	13	143	5.93361	0.57675	69	287	10.17.202	0-18	24.1	24.7942	97.2	0	0.06471	0.34118	0	0.34118	1	0	0	0	0	0.13412	0.03235	0.17059	0.34118	1	0	0.37667
6239	0	45	110	6.47059	0.48724	58	158	10.17.202	31-50	17	22.5764	75.3	0	0	1	0.29375	1	1	0.29375	1	0	0	0.72938	0	0.64688	1	0.9	0	0.80667
6240	0	21	149	8.46591	0.85929	57	477	10.17.202	19-30	17.6	17.3399	101.5	0	0.19118	1	0	1	1	0	0	0	0	0.2	0.09559	0.5	1	0.85	0	0
6241	1	31	145	5.82329	0.47635	33	165	10.17.202	31-50	24.9	30.4401	81.8	0	0	0.24706	0	0.24706	0.38998	0	0	0	0	0.0637	0	0.12353	0.24706	0	0.35	0.78333
6242	0	105	128	6.03774	0.44257	38	161	10.17.202	51+	21.2	28.9222	73.3	0	0	0.68235	0.41875	0.68235	0.76944	0.41875	0	0	0	0.18705	0	0.55055	0.68235	0	0.1	0.79667
6243	0	56	152	5.40925	0.41814	37	259	10.17.202	51+	28.1	36.3519	77.3	0	0	0	0.16875	0	0	0.16875	0	0	0	0.01688	0	0.08438	0	0	0.15	0.47
6244	0	36	167	8.02885	0.68325	13	176	10.17.202	31-50	20.8	24.4418	85.1	0	0	0.72941	0	0.72941	1	0	0	0	0	0.17294	0	0.36471	0.72941	0	1	0.74667
6245	1	59	113	4.23221	0.39233	66	350	10.17.202	51+	26.7	28.8026	92.7	0	0	0.03529	0	0.03529	0.79935	0	1	0	0	0.58346	0	0.01765	0.03529	1	0	0.16667
6246	0	30	153	8.22581	0.79461	43	240	10.17.202	31-50	18.6	19.2547	96.6	0	0.04706	0.98824	0	0.98824	1	0	0	0	0	0.19882	0.02353	0.49412	0.98824	0.15	0	0.53333
6247	1	81	144	5.12456	0.47761	45	441	10.17.202	51+	28.1	30.1502	93.2	0	0	0	0	0	0.46245	0	0	0	0	0.04624	0	0	0	0.25	0	0
6248	0	91	127	6.64921	0.56053	11	181	10.17.202	51+	19.1	22.6572	84.3	0	0	0.92941	0	0.92941	1	0	0	0	0	0.19294	0	0.46471	0.92941	0	1	0.73
6249	1	92	148	5.24823	0.37105	62	432	10.17.202	51+	28.2	39.8868	70.7	0	0	0	0.58125	0	0	0.58125	0	0	0.11188	0.06931	0	0.29063	0	1	0	0
6250	1	86	147	7.35	0.69384	39	233	10.17.202	51+	20	21.1864	94.4	0	0	0.82353	0	0.82353	1	0	0	0	0	0.18235	0	0.41176	0.82353	0	0.05	0.55667
6251	1	89	138	6.60287	0.64312	44	341	10.17.202	51+	20.9	21.4579	97.4	0	0.07059	0.71765	0	0.71765	1	0	0	0	0	0.17176	0.03529	0.35882	0.71765	0.2	0	0.19667
6252	1	43	164	8.77005	0.84543	41	374	10.17.202	31-50	18.7	19.3983	96.4	0	0.04118	0.97647	0	0.97647	1	0	0	0	0	0.19765	0.02059	0.48824	0.97647	0.05	0	0.08667
6253	0	3	104	3.51351	0.35486	66	493	10.17.202	0-18	29.6	29.3069	101	0	0.17647	0	0	0	0.67327	0	1	0.97297	0.31419	0.69604	0.08824	0	0	1	0	0
6254	1	107	167	7.13675	0.66015	66	317	10.17.202	51+	23.4	25.2973	92.5	0	0	0.42353	0	0.42353	1	0	0	0	0	0.14235	0	0.21176	0.42353	1	0	0.27667
6255	0	57	163	5.6993	0.56936	31	169	10.17.202	51+	28.6	28.6286	99.9	0	0.14412	0	0	0	0.84284	0	0	0	0	0.08428	0.07206	0	0	0	0.45	0.77
6256	1	1	115	5.37383	0.41164	15	285	10.17.202	0-18	21.4	27.9373	76.6	0	0	0.65882	0.2125	0.65882	1	0.2125	1	0	0	0.68713	0	0.43566	0.65882	0	1	0.38333
6257	0	90	112	6.05405	0.56848	54	277	10.17.202	51+	18.5	19.7018	93.9	0	0	1	0	1	1	0	1	0	0	0.7	0	0.5	1	0.7	0	0.41
6258	1	47	166	7.75701	0.60815	29	81	10.17.202	31-50	21.4	27.2959	78.4	0	0	0.65882	0.1	0.65882	1	0.1	0	0	0	0.17588	0	0.37941	0.65882	0	0.55	1
6259	1	97	137	7.3262	0.50844	42	332	10.17.202	51+	18.7	26.9452	69.4	0	0	0.97647	0.6625	0.97647	1	0.6625	0	0	0	0.2639	0	0.81949	0.97647	0.1	0	0.22667
6260	0	52	107	5.63158	0.43419	38	419	10.17.202	51+	19	24.6433	77.1	0	0	0.94118	0.18125	0.94118	1	0.18125	1	0	0	0.71224	0	0.56121	0.94118	0	0.1	0
6261	0	23	168	8.75	0.882	39	447	10.17.202	19-30	19.2	19.0476	100.8	0	0.17059	0.91765	0	0.91765	1	0	0	0	0	0.19176	0.08529	0.45882	0.91765	0	0.05	0
6262	0	70	118	6.34409	0.5088	46	278	10.17.202	51+	18.6	23.192	80.2	0	0	0.98824	0	0.98824	1	0	0.7	0	0	0.54882	0	0.49412	0.98824	0.3	0	0.40667
6263	0	28	109	3.63333	0.35788	50	479	10.17.202	19-30	30	30.4569	98.5	0	0.10294	0	0	0	0.38579	0	1	0.73333	0.27646	0.63956	0.05147	0	0	0.5	0	0
6264	0	104	152	5.62963	0.40308	20	347	10.17.202	51+	27	37.7095	71.6	0	0	2.1E-10	0.525	2.1E-10	0	0.525	0	0	0	0.0525	0	0.2625	2.1E-10	0	1	0.17667
6265	1	96	116	4.77366	0.4091	41	386	10.17.202	51+	24.3	28.3547	85.7	0	0	0.31765	0	0.31765	0.91132	0	0.9	0	0	0.5729	0	0.15882	0.31765	0.05	0	0.04667
6266	1	71	167	6.73387	0.62894	51	342	10.17.202	51+	24.8	26.5525	93.4	0	0	0.25882	0	0.25882	1	0	0	0	0	0.12588	0	0.12941	0.25882	0.55	0	0.19333
6267	0	95	128	6.30542	0.63117	68	133	10.17.202	51+	20.3	20.2797	100.1	0	0.15	0.78824	0	0.78824	1	0	0	0	0	0.17882	0.075	0.39412	0.78824	1	0	0.89
6268	1	75	144	7.82609	0.74817	32	381	10.17.202	51+	18.4	19.2469	95.6	0	0.01765	1	0	1	1	0	0	0	0	0.2	0.00882	0.5	1	0	0.4	0.06333
6269	1	16	117	6.3587	0.47563	62	94	10.17.202	0-18	18.4	24.5989	74.8	0	0	1	0.325	1	1	0.325	0.8	0	0	0.6325	0	0.6625	1	1	0	1
6270	0	90	165	7.53425	0.53568	44	147	10.17.202	51+	21.9	30.8017	71.1	0	0	0.6	0.55625	0.6	0.29958	0.55625	0	0	0	0.14558	0	0.57812	0.6	0.2	0	0.84333
6271	0	50	107	3.83513	0.33135	46	230	10.17.202	51+	27.9	32.2917	86.4	0	0	0	0	0	0	0	1	0.32975	0.60806	0.59378	0	0	0	0.3	0	0.56667
6272	1	75	143	8.17143	0.61286	47	195	10.17.202	51+	17.5	23.3333	75	0	0	1	0.3125	1	1	0.3125	0	0	0	0.23125	0	0.65625	1	0.35	0	0.68333
6273	0	76	163	8.27411	0.62387	43	178	10.17.202	51+	19.7	26.1273	75.4	0	0	0.85882	0.2875	0.85882	1	0.2875	0	0	0	0.21463	0	0.57316	0.85882	0.15	0	0.74
6274	1	19	126	5.86047	0.58429	25	215	10.17.202	19-30	21.5	21.5647	99.7	0	0.13824	0.64706	0	0.64706	1	0	0	0	0	0.16471	0.06912	0.32353	0.64706	0	0.75	0.61667
6275	0	2	139	6.74757	0.66464	61	108	10.17.202	0-18	20.6	20.9137	98.5	0	0.10294	0.75294	0	0.75294	1	0	0	0	0	0.17529	0.05147	0.37647	0.75294	1	0	0.97333
6276	1	102	134	6.23256	0.57776	11	388	10.17.202	51+	21.5	23.1931	92.7	0	0	0.64706	0	0.64706	1	0	0	0	0	0.16471	0	0.32353	0.64706	0	1	0.04
6277	1	108	137	5.82979	0.41158	11	93	10.17.202	51+	23.5	33.2861	70.6	0	0	0.41176	0.5875	0.41176	0	0.5875	0	0	0	0.09993	0	0.49963	0.41176	0	1	1

Fig. 9: Sample data with calculated indices and indicators.

models. The inclusion of numeric thresholds and normalized values ensures compatibility with downstream analytical workflows and improves clinical interpretation. This structured format is particularly valuable for validating the performance of the classification models and inspecting edge cases. The spreadsheet also serves as a foundation for model retraining, statistical analysis, and integration into electronic health record systems, ensuring continuity of care and clinical utility.

The final phase of data processing within the diagnostic platform involves the generation of a comprehensive table containing all calculated morphological and biochemical indicators necessary for anemia classification. This dataset, originally visualized in Fig. S5 and Fig. S6 of supplementary section, is structured to ensure traceability and transparency of every computed parameter. The table includes standard hematological indices such as MCH, MCHC, and MCV, along with their normalized values and derived binary indicators for microcytic, macrocytic, and normocytic anemia types. In addition, each row in the table contains flags identifying whether individual measurements fall within reference thresholds. For instance, binary markers like mMCH, mMCV, and mHCT are used to indicate deviations from norm values, supporting rapid morphological classification. Other important fields include biochemical indicators such as serum ferritin (Fer) and vitamin B12, which are essential for diagnosing iron deficiency anemia and megaloblastic anemia, respectively. The table also includes derived composite indices such as mNormo, mHaz, and mZhda, which represent more nuanced classifications by combining morphological and biochemical data. These values are normalized between 0 and 1, simplifying integration into machine learning models. Timestamps such as "created" and "updated" are included for auditability and reproducibility, ensuring that each calculation can be traced back to a specific version of the input data and model configuration. This level of granularity is crucial for clinical decision support systems, where the reliability and interpretability of results are paramount.

The Morphological Classification module confirmed the system's ability to accurately determine both the type of anemia and the degree of confidence in classification based on fuzzified values. Fig. 10 shows the main interface where it is possible to select a transaction and view the classification results.

The final classification output of the diagnostic system is presented in a structured tabular format that allows medical professionals to quickly review calculated anemia indicators and corresponding diagnostic outcomes (Fig. S7 in the supplementary section). Each row in this table represents a patient case and includes a series of derived features such as morphological indicators (mMCH, mMCV, mHCT), composite indices (M, mMicro, mNormo, mMacro), and biochemical data used to inform classification logic. The key element of this interface is the visual presentation of probabilities or normalized values (ranging from 0 to 1), highlighted with red outlines to denote high-confidence decisions or thresholds that were triggered in the logic. For instance, when the mMicro index exceeds a certain value and aligns with a low MCH or MCV, the system flags the case as microcytic anemia. The M index aggregates multiple features into a unified decision metric, providing an interpretable signal to the clinician. In addition to raw indices, the decision outcome such as “microcytic,” “normocytic,” or “macrocytic” anemia is embedded within the row through logic applied to threshold-based rules or model predictions. This output is not only used for patient-level analysis but also serves as a critical input for further training, validation, and refinement of the ensemble and neural classifiers. The structured nature of this data enables seamless integration with statistical evaluation tools and supports auditability, which is essential in medical diagnostics. The final values - both morphological and biochemical indicators are clearly presented in Fig. S8 in the supplementary section, which shows a combined display of the initial features, calculated indices and the final diagnosis.

Fig. 11 shows the final table with the results of the classification. The module was tested on a variety of clinical datasets and showed high reproducibility of decisions when changing the structure of the input data.

Fig. 10: Main view of the module with a list of transactions and their functions.

Data Output

	id [PK] integer	result_anemia boolean	result_macro boolean	result_micro boolean	result_normo boolean	result_AHZ boolean	result_ZhDA boolean	result_B12 boolean	createdAt timestamp with time zone	updatedAt timestamp with time zone
1	6960	false	false	false	false	false	false	false	2024-01-09 17:49:45.136+06	2024-01-09 17:49:45.136+06
2	6961	false	false	false	false	false	false	false	2024-01-09 17:49:45.136+06	2024-01-09 17:49:45.136+06
3	6962	true	false	true	true	true	false	false	2024-01-09 17:49:45.136+06	2024-01-09 17:49:45.136+06
4	6963	false	false	false	false	false	false	false	2024-01-09 17:49:45.136+06	2024-01-09 17:49:45.136+06
5	6964	false	false	false	false	false	false	false	2024-01-09 17:49:45.136+06	2024-01-09 17:49:45.136+06
6	6965	false	false	false	false	false	false	false	2024-01-09 17:49:45.136+06	2024-01-09 17:49:45.136+06
7	6966	false	false	false	false	false	false	false	2024-01-09 17:49:45.136+06	2024-01-09 17:49:45.136+06
8	6967	false	false	false	false	false	false	false	2024-01-09 17:49:45.136+06	2024-01-09 17:49:45.136+06
9	6968	false	false	false	false	false	false	false	2024-01-09 17:49:45.136+06	2024-01-09 17:49:45.136+06
10	6969	false	false	false	false	false	false	false	2024-01-09 17:49:45.136+06	2024-01-09 17:49:45.136+06
11	6970	true	false	true	true	true	false	false	2024-01-09 17:49:45.136+06	2024-01-09 17:49:45.136+06
12	6971	true	false	true	true	false	false	false	2024-01-09 17:49:45.136+06	2024-01-09 17:49:45.136+06
13	6972	true	false	false	false	false	false	false	2024-01-09 17:49:45.136+06	2024-01-09 17:49:45.136+06
14	6973	false	false	false	false	false	false	false	2024-01-09 17:49:45.136+06	2024-01-09 17:49:45.136+06
15	6974	false	false	false	false	false	false	false	2024-01-09 17:49:45.136+06	2024-01-09 17:49:45.136+06
16	6975	true	false	true	true	true	false	false	2024-01-09 17:49:45.136+06	2024-01-09 17:49:45.136+06
17	6976	false	false	false	false	false	false	false	2024-01-09 17:49:45.136+06	2024-01-09 17:49:45.136+06
18	6977	false	false	false	false	false	false	false	2024-01-09 17:49:45.136+06	2024-01-09 17:49:45.136+06
19	6978	false	false	false	false	false	false	false	2024-01-09 17:49:45.136+06	2024-01-09 17:49:45.136+06
20	6979	false	false	false	false	false	false	false	2024-01-09 17:49:45.136+06	2024-01-09 17:49:45.136+06
21	6980	false	false	false	false	false	false	false	2024-01-09 17:49:45.136+06	2024-01-09 17:49:45.136+06
22	6981	true	false	false	false	false	false	false	2024-01-09 17:49:45.136+06	2024-01-09 17:49:45.136+06
23	6982	false	false	false	false	false	false	false	2024-01-09 17:49:45.136+06	2024-01-09 17:49:45.136+06
24	6983	false	false	false	false	false	false	false	2024-01-09 17:49:45.136+06	2024-01-09 17:49:45.136+06

Fig. 11: Table with the results of the classification.

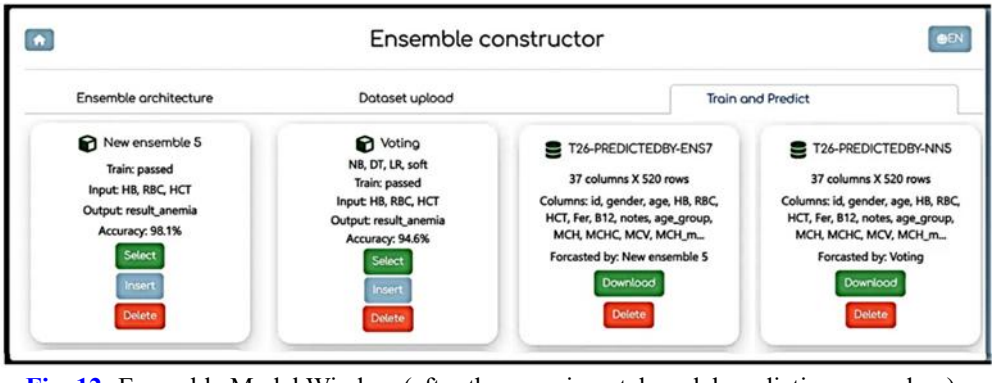

Fig. 12: Ensemble Model Window (after the experimental model prediction procedure).

The final output of the classification system is presented in the form of a Boolean table that captures the presence or absence of anemia and its specific subtypes for each patient entry. Each row represents an individual case identified by a unique ID, and the columns indicate the results of various classification flags. These include the general diagnosis (result_anemia), as well as subtype classifications such as result_micro (microcytic anemia), result_macro (macrocytic anemia), result_normo (normocytic anemia), and biochemical categories like result_B12, result_AHZ, and result_ZhDA (B12-deficiency, iron-deficiency, and mixed forms, respectively).

The results obtained confirmed the accuracy of the model's compliance with the morphological criteria used in clinical practice. Thus, the key advantages of the Morphological Classification module are the automation of the interpretation of laboratory data; quick determination of the type of anemia without the participation of an expert doctor; the ability to scale to large data sets.

3.5 Assessment of the clinical hematological syndromes' classification accuracy

As part of the experimental phase, the accuracy of the classifications of anemic states implemented in the Ensembling and Neural modules was quantified. The models were tested on consistent clinical and laboratory datasets that underwent pre-treatment and normalization steps.

The most stable results among ensemble approaches were shown by the XGBoost gradient boosting model, built on the input features of HGB, HCT and RBC. According to the results of cross-validation, the accuracy of the model on the training dataset was 98.1%, which is confirmed in Fig. 12, which demonstrates the card of the trained model and the metric of its quality. The model successfully classified both pronounced and borderline cases of anemia, while maintaining interpretability through access to the importance of traits and decision logic.

Thus, ensemble methods of machine learning have demonstrated high efficiency in solving problems of

multiclass classification of anemias. In particular, boosting algorithms (XGBoost and LightGBM) have shown not only high accuracy, but also significant interpretability due to the preservation of the significance of features, which is especially important in clinical trials. XGBoost efficiently detects nonlinear dependencies in data and is resistant to overfitting, while LightGBM provides a high learning rate that is useful for hyperparameter brute force. The bagging method was implemented on the basis of weak models (logistic regression, linear SVM and decision tree) trained on various bootstrap samples, which reduced model variance and increased resistance to overfitting. In turn, the voting method (Voting), implemented in two modes - hard and soft - provided a stable aggregation of predictions from the underlying models, increasing accuracy due to the consistency of decisions. Finally, Stacking combined the outputs of the underlying models (SVM, logistic regression, decision tree) using meta-classifiers, which improved the generalizability of the system by learning from inter-model patterns.

A neural network classifier built using a single feature (HB) showed an accuracy of 100% on the test dataset, as shown in Fig. 13. The Neural Analytics Module provides a comprehensive interface for creating, training, evaluating, and deploying neural network models designed for anemia classification. One of the key capabilities of the module is its support for experimentation with different feature sets and architectures. As part of the evaluation process, a simplified model was built using only one input parameter HGB to test the predictive potential of individual features. Remarkably, this neural network achieved 100% accuracy on the test set, with identical performance on the training data. The loss values for the training and test datasets were also low (0.532 and 0.500, respectively), indicating minimal overfitting and high generalization capacity. This result suggests that in specific cases, especially with clearly defined thresholds, even a single hematological marker can be sufficient for a confident diagnosis.

In addition to this minimal model, several other neural architectures were evaluated using expanded feature sets including MCH, MCHC, MCV, HCT, RDW, and iron-related biomarkers (Fer, B12). These networks showed varying degrees of accuracy and loss values, reflecting the impact of feature combinations and network complexity on classification performance. The module interface allows users to select networks, inspect their training metrics, and match them with datasets for inference, thus providing full lifecycle support for machine learning in clinical diagnostics.

The key difference between a neural network and ensemble methods is the form of the output data: instead of a binary label (anemia/no anemia), the model gives a probabilistic estimate for each of the possible states - the presence of anemia, microcytic, macrocytic and normocytic forms. This provides flexibility in interpretation and can be used to calculate clinical risk in diagnostically non-obvious cases.

Ensemble models (especially Boosting) provide high accuracy, resistance to overfitting and interpretability of decisions, which makes them an effective tool for

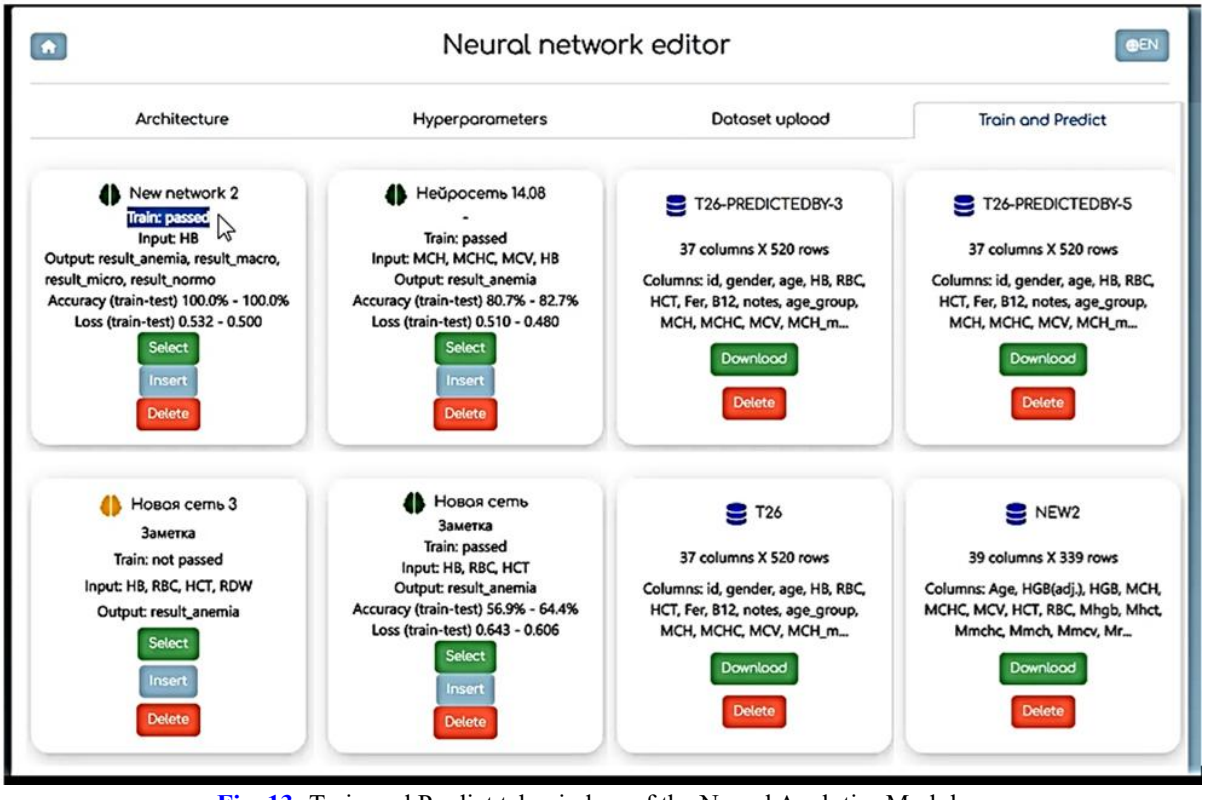

Fig. 13: Train and Predict tab window of the Neural Analytics Module.

Table 4: Machine Learning Model Accuracy Comparison Results.

Model	Accuracy	Model Output	Features
XGBoost (Boosting)	98.1%	Binary (0/1)	High interpretability, stability
Random Forest (Bagging)	96.4%	Binary (0/1)	Noise resistance
Voting (Soft)	95.8%	Binary (0/1)	Moderately accurate, simple aggregation
Voting (Hard)	94.6%	Binary (0/1)	Less sensitivity, simplicity
Stacking Ensemble	96.9%	Binary (0/1)	Flexibility through meta-learning
Neural Network	100.0%	Probabilistic (0-1)	Adaptability, sensitivity to patterns

automating diagnostics in conditions of a limited number of features. At the same time, neural networks have greater adaptability and sensitivity to weakly expressed patterns, allowing them to.form probabilistic forecasts, which is valuable in building predictive decision support systems. The results of comparing the accuracy of the models are presented in Table 4.

Thus, the combined use of ensemble and neural network approaches provides both classification reliability and clinical flexibility, opening up opportunities for personalized diagnostic support for the patient.

3.6 Case study

To demonstrate the practical applicability of the developed models, a detailed analysis of the diagnostic result was carried out on a specific example of a patient. Visualization of the initial and predicted dataset allows to trace the full path of data processing - from "raw" laboratory values to complex probabilistic interpretation.

A table with laboratory indicators is fed to the system, including the values of HGB, HCT, RBC and red blood cell indices (MCV, MCH, MCHC). Fig. 14a shows the classification results based on the Ensembling module, for which HB, RBC and HCT were selected as input, and presence of anemia as output. Morphological Classification provided the results shown in Fig. 14b, also determining the type and character of anemic cases.

Based on the output data of the trained neural network model, the following probabilistic estimates were obtained:

- Presence of anemia - 0.80 (80%);
- Microcytic form - 0.70 (70%);
- Macrocytic form - 0.00 (0%);
- Normocytic form - 0.20 (20%);
- Iron deficiency anemia (IDA) - 0.65 (65%);
- B12-deficiency anemia - 0.00 (0%);
- Anemia of chronic diseases - 0.10 (10%).

Thus, the model confidently classified the condition as anemia with a predominance of the microcytic morphological type and a high probability of iron deficiency nature. These results are fully consistent with clinical expectations with low MCH and MCV values and reduced ferritin values.[46,47]

Through Fig. S6, Fig. S8 and Fig. 14, it can be seen that the structure of the original dataset is supplemented with columns displaying aggregated, processed data, risk indicators and classification results. For Machine Learning modules, the data is shortened to include only selected input and output variables which are used by the models. All of this facilitates clinical interpretation and minimizes the need for manual analysis.

3.7 Interpretation of the results

The results obtained in the course of experimental testing demonstrate high accuracy, stability and reproducibility of the proposed software package in the classification of clinical and hematological syndromes, in particular, anemias of various types. Models implemented on the basis of XGBoost, Random Forest, voting ensembles and neural networks showed accuracy metrics from 94.6% to 100%, which exceeds the standard values achieved in manual laboratory interpretation.

High accuracy and F1-score values reflect the ability of the system not only to record the presence of anemic syndrome, but also to confidently differentiate between morphological types (micro-, macro- and normocytic forms). This is especially important in the context of practical medicine, where misinterpretation of MCH/MCV leads to errors in the prescription of therapy (for example, inappropriate administration of iron supplements for B12-deficiency anemia). The examples presented in Section 3.6 (Case Study) confirm the consistency of automatic inference with clinical logic.

To better understand the relationship between the technical metrics obtained during model validation and their clinical relevance in real-world diagnostic conditions, Table 5 presents the correspondence between the main quantitative indicators and their interpretation from the practitioner's perspective. This makes it possible to justify the choice of specific models not only on the basis of formal accuracy, but also taking into account their practical usefulness in the clinical and diagnostic context.

Classification algorithms have demonstrated resilience to data variability, including cases with borderline values. This confirms the correctness of the built rules and fuzzified mechanisms for calculating the degree of belonging to diagnostic classes. In particular, the ability of models to handle "mixed forms" of anemia - when there are both high levels of Mmicro and Mnormo - is an indicator of the diagnostic logic preparedness implemented in the system.

In addition, the use of neural network models with probabilistic output made it possible not only to record the presence of the syndrome, but also to form a confidence scale, which is especially valuable when integrated into decision support systems (CDSS). This reduces the diagnostic burde

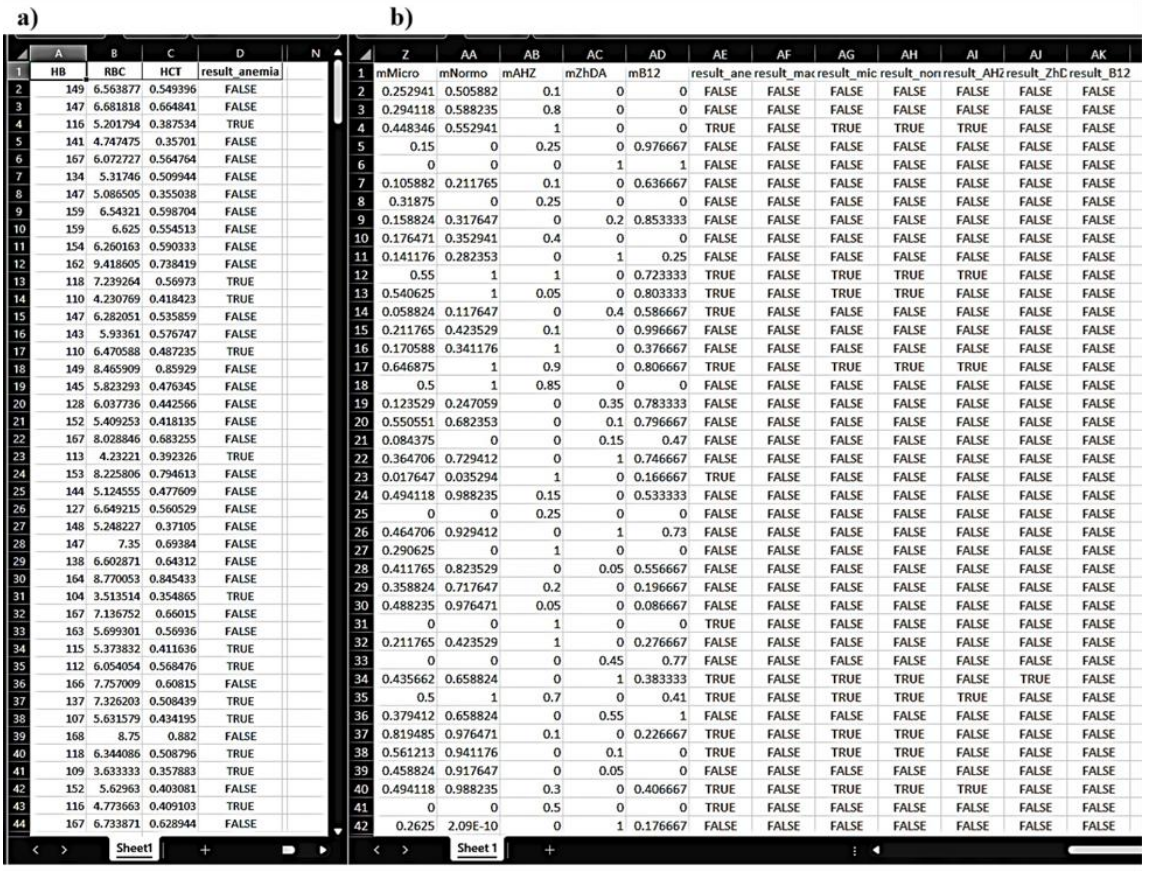

Fig. 14: Ensembling (a) and Morphological Classification (b) predictions.

Table 5: Compliance of quality metrics with classification and their clinical interpretation.

Metrics	Interpretation
Accuracy	Overall model accuracy: proportion of correctly classified patients. It is important during screening.
Precision	The proportion of true positive values among all predicted positive ones. Reduces the risk of overdiagnosis.
Recall	The proportion of true positive values among all real positive ones. It is critical to exclude the skipping of anemia.
F1-score	Harmonic mean of Precision and Recall. It is stable in unbalanced classes.
Probability Output (NN)	Allows the doctor to assess the degree of confidence of the model in the diagnosis and adjust the clinical decision.
Feature Importance (XGBoost)	Reflects the contribution of each attribute to the final decision. Increases transparency and trust in the system.

on the doctor, reduces the risk of subjective error and allows the system to be used in conditions of limited time, for example, in emergency departments or at the primary care level.

3.8 Limitations

Despite the high indicators of accuracy and stability demonstrated during testing, the implementation of the diagnostic complex "Health Passport" has a number of limitations, both at the level of the models used and in relation to external factors affecting the quality of classification. Some algorithms, especially the basic classifiers (logistic regression, KNN), show a limited ability to capture complex nonlinear relationships between traits, which can reduce effectiveness in cases of borderline and atypical forms of anemia. Ensemble methods (in particular, Voting and Bagging) provide increased stability, but are inferior in adaptability to neural networks.[48] On the other hand, neural networks, although they have shown 100% accuracy in test scenarios, require longer training, regular validation, and may suffer from a lack of interpretability in a clinical context.

The quality of the diagnostic conclusion directly depends on the correctness of the initial laboratory data. Analyzer calibration failures, sampling errors, and non-uniformity of units (*e.g.*, differences in MCHC formats: g/L vs. g/dL) can lead to a decrease in classification accuracy. The Data Preprocessing and Mathematical Evaluation modules partially compensate for these risks through filtering and normalization, but complete protection against systematic errors is possible only if laboratory diagnostic standards are met. A separate task remains the interpretation of mixed forms of anemia, in which signs of two or more morphological types are observed at the same time.[49] In such cases, classical logical rules (*e.g.*, MCV/MCH thresholds) are insufficient. Although the system

supports a fuzzified assessment of the degree of belonging to several types, its conclusion requires careful clinical interpretation. There are also difficult cases with "smeared" indicators, where the values of the indices are close to the limits of the norm, especially in combination with dubious biochemical markers.

3.9 Practical implications

The developed software and hardware complex has a high degree of practical applicability in modern digital healthcare systems, especially in conditions of limited resources and the growing need for automation of clinical and diagnostic processes. The system can be easily adapted for use within telemedicine platforms, providing automatic collection and interpretation of laboratory data without the need for face-to-face contact with a doctor. The ability to integrate into web interfaces and mobile solutions allows the complex to be used in remote settlements, rural medical centers, as well as within the framework of on-site diagnostics. This is especially true in countries with geographically dispersed medical infrastructure and a limited number of hematologists.

Thanks to the standardized decision-making logic and transparency of the criteria used, the system can be implemented as part of clinical guidelines or local protocols for diagnosing anemia. The structured presentation of the output data (morphological type, M index, probability of pathology, comments on the type of anemia) complies with the principles of evidence-based medicine and can be used as part of electronic medical records or as a module within clinical diagnostic complexes. The system performs a preliminary interpretation of laboratory data, allowing the clinician to focus on the clinical decision and eliminating the need for manual analysis of indicators. This is especially important in conditions of overload of outpatient links, where the doctor must analyze dozens or hundreds of laboratory test protocols daily. Automation of processes increases the speed of diagnosis, reduces the likelihood of missing pathologies and contributes to a timelier prescription of therapy.

3.10 Future directions

The developed software and hardware complex demonstrates high potential as a new generation diagnostic system, but its sustainable evolution and scalability require the implementation of a number of scientific, technological and clinical areas. The current version of the system was trained and tested on samples predominantly from a single center or source. To increase the generalizability of models, it is necessary to integrate multicenter data, including patients of different gender, age, ethnicity, and clinical background. This will ensure that the classifiers are resilient to regional variations in laboratory standards and increase reliability at the population scale.

One of the priorities is to expand the volume of input data by including biochemical markers (*e.g.*, transferrin, soluble ferritin receptor), enzyme-linked immunoassays, and genetic predictors of anemia. This will not only increase the sensitivity of diagnosis, but also move from morphological classification to pathogenetic classification, including the detection of hereditary anemias, inflammatory conditions and malabsorption syndromes. Although ensemble methods (*e.g.*, XGBoost) are highly interpretable due to the importance of features, neural network models remain "black boxes". The development of the explainable AI (XAI) component using modern interpretation methods - such as SHAP, LIME, Integrated Gradients - will reveal the logic of even deep models and make their results clinically transparent and valid.

In the future, it is planned to introduce an online learning mechanism, in which the models will be able to gradually adapt to the incoming clinical data in the field. This approach will allow for population shifts, changing testing conditions, and new laboratory standards without the need for a complete retraining of the entire system. The future development of the platform involves close integration with electronic medical documentation systems, telemedicine gateways, as well as participation in the construction of national blood disease registries. In the long term, personalized patient paths with dynamic AI-based analytics can be implemented.

4. Conclusion

The developed intelligent software package of the health passport is a modern solution for automated morphological classification and differential diagnosis of clinical and hematological syndromes, primarily anemias of various etiologies. The system combines traditional medical formulas, inference algorithms, ensemble methods of machine learning and neural network models in a single technologically and methodologically consistent circuit. The modular architecture of the platform covers all stages of the diagnostic process - from data collection and preprocessing to mathematical evaluation, classification and forecasting. The results of experimental testing demonstrated high accuracy of classification (up to 100% on test samples), stability of calculations, reproducibility of conclusions, and clinical interpretability.

The development is focused on integration into practical healthcare: the system can be used both in stationary and remote (telemedicine) scenarios, including primary care, laboratory diagnostic units, screening programs and digital doctor's offices. Automation of calculations and interpretation makes it possible to reduce the burden on medical personnel, reduce subjectivity in decision-making and increase the standardization of anemia diagnosis. The proposed solution demonstrates an example of the transition from manual laboratory interpretation to intelligently controlled diagnostic processes and can be expanded to cover a wider range of hematological and biochemical pathologies. The results obtained serve as a justification for further implementation of the system in real clinical processes and its scaling within the framework of a new generation of digital healthcare.

Acknowledgments
This research was carried out with the support of grant funding for scientific and (or) scientific and technical projects for 2023-2025 of the Ministry of Science and Higher Education of the Republic of Kazakhstan (grant №AP19679525). The authors would like to express their sincere gratitude to the staff of LLP «A-Medical» (Almaty, Kazakhstan) who participated in this study as experts.

Conflict of Interest
The authors declare no conflicts of interest.

Supporting Information
Applicable.

Abbreviations

Abbreviation	Definition
ABS LYMP	Absolute Lymphocyte Count
ACD	Anemia of Chronic Disease
ANN	Artificial Neural Networks
B12	Vitamin B12
CBC	Complete Blood Count
CNN	Convolutional Neural Network
DNN	Deep Neural Network
DSS	Decision-Support Systems
DT	Decision Tree
EDA	Exploratory Data Analysis
ELM	Extreme Learning Machine
HCT	Hematocrit
HGB	Hemoglobin
IDA	Iron Deficiency Anemia
kNN	K-Nearest Neighbors
LightGBM	Light Gradient Boosting Machine
LIME	Local Interpretable Model-agnostic Explanations
LR	Logistic Regression
MCH	Mean Corpuscular Hemoglobin
MCHC	Mean Corpuscular Hemoglobin Concentration
MCV	Mean Corpuscular Volume
MIMIC	Medical Information Mart for Intensive Care
ML	Machine Learning
PCA	Principal Component Method
PCT	Plateletcrit
PDW	Platelet Distribution Width
PLT	Platelet Count
RBC	Red Blood Cell Count
RDW	Red Cell Distribution Width
RNN	Recurrent Neural Network
SE	Error of the Mean
SHAP	SHapley Additive Explanations
STD	Standard Deviation
SVM	Support Vector Machine
TLC	Total Leukocyte Count
WBC	White Blood Cell Count
XGBoost	Extreme Gradient Boosting

CRediT Statement
Indira Uvaliyeva: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing, Visualization, Supervision. **Zhanat Idrisheva:** Methodology, Validation, Formal analysis, Investigation, Writing, Visualization, Supervision. **David Borozenets:** Software, Validation, Investigation, Resources, Data curation. **Shynar Tezekpayeva:** Validation, Formal analysis, Investigation, Resources, Data curation. **Zarına Khassenova:** Conceptualization, Methodology, Formal analysis, Writing, Visualization; **Zhanar Beldeubayeva:** Conceptualization, Methodology, Validation, Formal analysis, Writing.

References
[1] A. Baldi, S.-R. Pasricha, Anaemia: worldwide prevalence and progress in reduction, *Nutritional Anemia*, 2022, 3-17, doi: 10.1007/978-3-031-14521-6_1.

[2] G. A. Stevens, C. J. Paciorek, M. C. Flores-Urrutia, E. Borghi, S. Namaste, J. P. Wirth, P. S. Suchdev, M. Ezzati, F. Rohner, S. R. Flaxman, L. M. Rogers, National, regional, and global estimates of anaemia by severity in women and children for 2000–19: a pooled analysis of population-representative data, *The Lancet Global Health*, 2022, **10**, e627-e639, doi: 10.1016/s2214-109x(22)00084-5.

[3] O. O. Awe, D. M. Dogbey, R. Sewpaul, D. Sekgala, N. Dukhi, Anaemia in children and adolescents: a bibliometric analysis of BRICS countries (1990–2020), *International Journal of Environmental Research and Public Health*, 2021, **18**, 5756, doi: 10.3390/ijerph18115756.

[4] A. J. Read, A. K. Waljee, J. B. Sussman, H. Singh, G. Y. Chen, S. Vijan, S. D. Saini, Testing practices, interpretation, and diagnostic evaluation of iron deficiency anemia by US primary care physicians, *JAMA Network Open*, 2021, **4**, e2127827, doi: 10.1001/jamanetworkopen.2021.27827.

[5] N. Milman, Anemia: still a major health problem in many parts of the world!, *Annals of Hematology*, 2011, **90**, 369-377, doi: 10.1007/s00277-010-1144-5.

[6] A. Owais, C. Merritt, C. Lee, Z. A. Bhutta, Anemia among women of reproductive age: an overview of global burden, trends, determinants, and drivers of progress in low- and middle-income countries, *Nutrients*, 2021, **13**, 2745, doi: 10.3390/nu13082745.

[7] S.-R. Pasricha, L. Rogers, F. Branca, M.-N. Garcia-Casal, Measuring haemoglobin concentration to define anaemia: WHO guidelines, *The Lancet*, 2024, **403**, 1963-1966, doi: 10.1016/s0140-6736(24)00502-6.

[8] S. Pach, E. L. Webb, A. Edielu, R. Nagawa, V. Anguajibi, S.

Mpooya, H. Wu, S. Colt, P. Mawa, J. Richter, J. F. Friedman, A. L. Bustinduy, Baseline liver ultrasound findings in preschool children from the praziquantel in preschoolers trial in lake albert, Uganda, *Pediatric Infectious Disease Journal*, 2024, **43**, 14-20, doi: 10.1097/inf.0000000000004119.

[9] R. K. Benedict, T. W. Pullum, S. Riese, E. Milner, Is child anemia associated with early childhood development? A cross-sectional analysis of nine Demographic and Health Surveys, *PLoS One*, 2024, **19**, e0298967, doi: 10.1371/journal.pone.0298967.

[10] G. A. Stevens, T. Beal, M. N. N. Mbuya, H. Luo, L. M. Neufeld, O. Y. Addo, S. Adu-Afarwuah, S. Alayón, Z. Bhutta, K. H. Brown, M. E. Jefferds, R. Engle-Stone, W. Fawzi, S. Y. Hess, R. Johnston, J. Katz, J. Krasevec, C. M. McDonald, Z. Mei, S. Osendarp, C. J. Paciorek, N. Petry, C. M. Pfeiffer, M. J. Ramirez-Luzuriaga, L. M. Rogers, F. Rohner, V. Sethi, P. S. Suchdev, M. Tessema, S. Villapando, F. T. Wieringa, A. M. Williams, M. Woldeyahannes, M. F. Young, Micronutrient deficiencies among preschool-aged children and women of reproductive age worldwide: a pooled analysis of individual-level data from population-representative surveys, *The Lancet Global Health*, 2022, **10**, e1590-e1599, doi: 10.1016/s2214-109x(22)00367-9.

[11] D. Kinyoki, A. E. Osgood-Zimmerman, N. V. Bhattacharjee, N. J. Kassebaum, S. I. Hay, Anemia prevalence in women of reproductive age in low-and middle-income countries between 2000 and 2018, *Nature medicine*, 2021, **27**(10), 1761-1782, doi: 10.1038/s41591-021-01498-0.

[12] M. W. Wlodarski, A. Vlachos, J. E. Farrar, L. M. Da Costa, A. Kattamis, I. Dianzani, C. Belendez, S. Unal, H. Tamary, R. Pasauliene, D. Pospisilova, J. de la Fuente, D. Iskander, L. Wolfe, J. M. Liu, A. Shimamura, K. Albrecht, B. Lausen, A. G. Bechensteen, U. Tedgard, A. Puzik, P. Quarello, U. Ramenghi, M. Bartels, H. Hengartner, R. A. Farah, M. Al Saleh, A. Ali Hamidieh, W. Yang, E. Ito, H. Kook, G. Ovsyannikova, L. Kager, P.-E. Gleizes, J.-H. Dalle, B. Strahm, C. M. Niemeyer, J. M. Lipton, T. M. Leblanc, Diagnosis, treatment, and surveillance of Diamond-Blackfan anaemia syndrome: international consensus statement, *The Lancet Haematology*, 2024, **11**, e368-e382, doi: 10.1016/s2352-3026(24)00063-2.

[13] A. Wijerathna-Yapa, R. Pathirana, Sustainable agro-food systems for addressing climate change and food security, *Agriculture*, 2022, **12**, 1554, doi: 10.3390/agriculture12101554.

[14] S. Shinde, C. A. Yelverton, M. Yussuf, L. Nurhussien, D. Wang, W. W. Fawzi, Effects of vitamin and multiple micronutrient supplementation for pregnant and/or lactating women on maternal and infant nutritional status in low- and middle-income countries: a systematic review and meta-analysis, *Advances in Nutrition*, 2025, 100487, doi: 10.1016/j.advnut.2025.100487.

[15] J. Daru, Sustainable development goals for anaemia: 20 years later, where are we now?, *The Lancet Global Health*, 2022, **10**, e586-e587, doi: 10.1016/s2214-109x(22)00127-9.

[16] M. M. Hasan, R. J. S. Magalhaes, S. Ahmed, S. Pervin, M. Tariqujjaman, Y. Fatima, A. A. Mamun, Geographical variation and temporal trend in anemia among children aged 6–59 months in low- and middle-income countries during 2000–2018: forecasting the 2030 SDG target, *Public Health Nutrition*, 2021, **24**, 6236-6246, doi: 10.1017/s1368980021002482.

[17] I. N. Muhsen, D. Shyr, A. D. Sung, S. K. Hashmi, Machine learning applications in the diagnosis of benign and malignant hematological diseases, *Clinical Hematology International*, 2021, **3**, 13, doi: 10.2991/chi.k.201130.001.

[18] H. Tvedten, Classification and laboratory evaluation of anemia, *Schalm's veterinary hematology*, 2022, 198-208, doi: 10.1002/9781119500537.ch25.

[19] D. D. Patel, P. N. Desai, J. A. Prajapati, A. B. Patel, A. D. Patel, Red blood cell histogram in different morphological types of anemia in comparison with peripheral blood smears: a comparative analysis, *Journal of Clinical and Basic Research*, 2024, **8**, 5-9, doi: 10.61186/jcbr.8.2.5.

[20] F. Farine, A. M. Rapisarda, C. Roani, C. Giuli, C. Comisi, A. Mascio, T. Greco, G. Maccauro, C. Perisano, Predictive factors of amputation in diabetic foot, *Biomedicines*, 2024, **12**, 2775, doi: 10.3390/biomedicines12122775.

[21] D. C. E. Saputra, K. Sunat, T. Ratnaningsih, A new artificial intelligence approach using extreme learning machine as the potentially effective model to predict and analyze the diagnosis of anemia, *Healthcare*, 2023, **11**, 697, doi: 10.3390/healthcare11050697.

[22] R. Vohra, A. Hussain, A. K. Dudyala, J. Pahareeya, W. Khan, Multi-class classification algorithms for the diagnosis of anemia in an outpatient clinical setting, *PLoS One*, 2022, **17**, e0269685, doi: 10.1371/journal.pone.0269685.

[23] A. E. Obstfeld, Hematology and machine learning, *The Journal of Applied Laboratory Medicine*, 2023, **8**, 129-144, doi: 10.1093/jalm/jfac108.

[24] M. Göl, C. Aktürk, T. Talan, M. S. Vural, İ. H. Türkbeyler, Predicting malnutrition-based anemia in geriatric patients using machine learning methods, *Journal of Evaluation in Clinical Practice*, 2025, **31**, e14142, doi: 10.1111/jep.14142.

[25] P. Appiahene, S. S. D. Dogbe, E. E. Y. Kobina, P. S. Dartey, S. Afrifa, E. T. Donkoh, J. W. Asare, Application of ensemble models approach in anemia detection using images of the palpable palm, *Medicine in Novel Technology and Devices*, 2023, **20**, 100269, doi: 10.1016/j.medntd.2023.100269.

[26] S. Chakraborty, K. Kansara, R. Dinesh Kumar, D. Swaminathan, K. Aatre, S. Acharya, Non-invasive estimation of clinical severity of anemia using hierarchical ensemble classifiers,

Journal of Medical and Biological Engineering, 2022, **42**, 828-838, doi: 10.1007/s40846-022-00750-3.

[27] B. E. Dejene, T. M. Abuhay, D. S. Bogale, Predicting the level of anemia among Ethiopian pregnant women using homogeneous ensemble machine learning algorithm, *BMC Medical Informatics and Decision Making*, 2022, **22**, 247, doi: 10.1186/s12911-022-01992-6.

[28] S. Yeruva, M. S. Varalakshmi, B. P. Gowtham, Y. H. Chandana, P. K. Prasad, Identification of sickle cell anemia using deep neural networks, *Emerging Science Journal*, 2021, **5**, 200-210, doi: 10.28991/esj-2021-01270.

[29] M. Shahzad, A. I. Umar, S. H. Shirazi, Z. Khan, A. Khan, M. Assam, A. Mohamed, E.-A. Attia, Identification of anemia and its severity level in a peripheral blood smear using 3-tier deep neural network, *Applied Sciences*, 2022, **12**, 5030, doi: 10.3390/app12105030.

[30] N. E. Garduno-Rapp, Y. S. Ng, J. L. Weon, S. N. Saleh, C. U. Lehmann, C. Tian, A. Quinn, Early identification of patients at risk for iron-deficiency anemia using deep learning techniques, *American Journal of Clinical Pathology*, 2024, **162**, 243-251, doi: 10.1093/ajcp/aqae031.

[31] G. Prashanthi, S. P. Singh, Identification of sickle cell anemia by employing hybrid optimization and recurrent neural network, *International Conference on Pervasive Computing and Social Networking (ICPCSN)*, Salem, India, June 19-20, 2023, 853-858, doi: 10.1109/ICPCSN58827.2023.00146.

[32] T. Berghout, Iron deficiency anemia diagnosis for young children: image-driven comparative analysis of recurrent and projected neural networks, *International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE)*, Mexico, October 23-25, 2024, 1-6, doi: 10.1109/CCE62852.2024.10770955.

[33] J. W. Asare, P. Appiahene, E. T. Donkoh, G. Dimauro, Iron deficiency anemia detection using machine learning models: a comparative study of fingernails, palm and conjunctiva of the eye images, *Engineering Reports*, 2023, **5**, e12667, doi: 10.1002/eng2.12667.

[34] A. Kumar, C. Y, A. Kodipalli, T. Rao, Anemia detection and severity prediction using classification algorithms with optimized hyperparameters, boosting techniques and XAI, *International Conference for Emerging Technology (INCET)*, Belgaum, India, May 24-26, 2024, 1-5, doi: 10.1109/INCET61516.2024.10593447.

[35] O. O. Awe, J. M. Adepoju, E. Boniface, O. D. Awe, Comparative analysis of random forest and neural networks for anemia prediction in female adolescents: a LIME-based explainability approach, *Practical Statistical Learning and Data Science Methods*, 2024, 555-573, doi: 10.1007/978-3-031-72215-8_23.

[36] M. F. McGhee, C. A. Saxelby, N. McKay, A guide to laboratory investigations, *CRC Press*, 2021, **7**, 153, doi: 10.1201/9781003049685.

[37] G. S. Tegenaw, D. Amenu, G. Ketema, F. Verbeke, J. Cornelis, B. Jansen, Evaluating a clinical decision support point of care instrument in low resource setting, *BMC Medical Informatics and Decision Making*, 2023, **23**, 51, doi: 10.1186/s12911-023-02144-0.

[38] A. S. A. Alsaedi, N. A. H. Alharbi, T. Dhaifallah, A.M. Alawamer, A. T. Al Jabri, A. Homeda, F.A.H. Alsarrani, A. F. Alsharef, The Role of Predictive Analytics in Theoretical Modeling of Clinical Laboratory Workflow Optimization, *Journal of International Crisis and Risk Communication Research*, 2024, **7**, 1864-1880, doi: 10.63278/jicrcr.vi.2383.

[39] S. Roy, T. Meena, S.-J. Lim, Demystifying supervised learning in healthcare 4.0: a new reality of transforming diagnostic medicine, *Diagnostics*, 2022, **12**, 2549, doi: 10.3390/diagnostics12102549.

[40] M. L. Doherty, K. E. Youens, Laboratory Information Systems, *Clinical Laboratory Management*, 2024, 285-297, doi: 10.1002/9781683673941.ch21.

[41] L. O. Schwen, T.-R. Kiehl, R. Carvalho, N. Zerbe, A. Homeyer, Digitization of pathology labs: a review of lessons learned, *Laboratory Investigation*, 2023, **103**, 100244, doi: 10.1016/j.labinv.2023.100244.

[42] M. Romanchikova, S. A. Thomas, A. Dexter, M. Shaw, I. Partarrieau, N. Smith, J. Venton, M. Adeogun, D. Brettle, R. J. Turpin, The need for measurement science in digital pathology, *Journal of Pathology Informatics*, 2022, **13**, 100157, doi: 10.1016/j.jpi.2022.100157.

[43] M. N. Garcia-Casal, O. Dary, M. E. Jefferds, S.-R. Pasricha, Diagnosing anemia: Challenges selecting methods, addressing underlying causes, and implementing actions at the public health level, *Annals of the New York Academy of Sciences*, 2023, **1524**, 37-50, doi: 10.1111/nyas.14996.

[44] Z. Hevessy, G. Toth, P. Antal-Szalmas, M. Tokes-Fuzesi, J. Kappelmayer, B. Karai, E. Ajzner, Algorithm of differential diagnosis of anemia involving laboratory medicine specialists to advance diagnostic excellence, *Clinical Chemistry and Laboratory Medicine (CCLM)*, 2024, **62**, 410-420, doi: 10.1515/cclm-2023-0807.

[45] S. Yashwanth, V. Kulkarni, H. T. Chethana, N. C. Chaitra, Smart Hospitals: Integrating Connectivity and Intelligence, *Smart Hospitals: 5G, 6G and Moving Beyond Connectivity*, 2024, 1-22, doi: 10.1002/9781394275472.ch1.

[46] F. Khaled, A comparative study of hematological parameters and serum ferritin levels between males and females in el-Beida city, Libya, *AlQalam Journal of Medical and Applied Sciences*, 2025, 843-845, doi: 10.54361/ajmas.258243.

[47] N. Ding, Y.-H. Ma, P. Guo, T.-K. Wang, L. Liu, J.-B. Wang, P.-P. Jin, Reticulocyte hemoglobin content associated with the risk of iron deficiency anemia, *Heliyon*, 2024, **10**, e25409, doi: 10.1016/j.heliyon.2024.e25409.
[48] K. Zhuang, C. Zhang, Z. Chen, T. She, M. Wang, Integrating convolutional neural networks with ensemble methods for enhanced diabetes diagnosis: a multi-dataset evaluation, *Frontiers in Medicine*, 2025, **12**, 1657889, doi: 10.3389/fmed.2025.1657889.
[49] G. C. Guidi, Hematological diagnostics, *Clinical and Laboratory Medicine Textbook*, 2023, 163-193, doi: 10.1007/978-3-031-24958-7_15.