



OPEN ACCESS

EDITED BY

Lia Morra,
Polytechnic University of Turin, Italy

REVIEWED BY

Satya Ranjan Dash,
KIIT University, India
Massimo La Rosa,
National Research Council (CNR), Italy

*CORRESPONDENCE

João Carlos Sousa,
✉ jcsousa@med.uminho.pt

RECEIVED 11 January 2024

ACCEPTED 05 March 2024

PUBLISHED 25 March 2024

CITATION

Santos-Silva MA, Sousa N and Sousa JC (2024),
Artificial intelligence in routine blood tests.
Front. Med. Eng. 2:1369265.
doi: 10.3389/fmede.2024.1369265

COPYRIGHT

© 2024 Santos-Silva, Sousa and Sousa. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Artificial intelligence in routine blood tests

Miguel A. Santos-Silva^{1,2}, Nuno Sousa^{1,2,3,4} and João Carlos Sousa^{1,2*}

¹Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho, Braga, Portugal, ²ICVS/3B's—PT Government Associate Laboratory, Braga, Portugal, ³Clinical Academic Center-Braga (2CA), Braga, Portugal, ⁴Association P5 Digital Medical Center (ACMP5), Braga, Portugal

Routine blood tests drive diagnosis, prognosis, and monitoring in traditional clinical decision support systems. As a routine diagnostic tool with standardized laboratory workflows, clinical blood analysis offers superior accessibility to a comprehensive assessment of physiological parameters. These parameters can be integrated and automated at scale, allowing for in-depth clinical inference and cost-effectiveness compared to other modalities such as imaging, genetic testing, or histopathology. Herein, we extensively review the analytical value of routine blood tests leveraged by artificial intelligence (AI), using the ICD-10 classification as a reference. A significant gap exists between standard disease-associated features and those selected by machine learning models. This suggests an amount of non-perceived information in traditional decision support systems that AI could leverage with improved performance metrics. Nonetheless, AI-derived support for clinical decisions must still be harmonized regarding external validation studies, regulatory approvals, and clinical deployment strategies. Still, as we discuss, the path is drawn for the future application of scalable artificial intelligence (AI) to enhance, extract, and classify patterns potentially correlated with pathological states with restricted limitations in terms of bias and representativeness.

KEYWORDS

blood analyses, blood, artificial intelligence (AI), machine learning (ML), diagnosis

Introduction

Artificial intelligence (AI) stems from the data generated mainly since the beginning of the fourth industrial revolution, which has progressively changed how people live, interact, and work (Sarker, 2021). Automated systems, meant to emulate human cognitive capabilities, deploy supervised applications to perform repetitive tasks more accurately and efficiently, saving time and effort for high-volume workloads. In medicine, AI has become a valuable tool for improving patient outcomes, particularly in diagnostics, where image and text-based systems supported by machine learning (ML) and deep learning (DL) technologies are reaching remarkable clinical results (Reardon, 2019). The COVID-19 pandemic is the paramount example of how AI applications enable new screening tools and achieve early diagnosis by measuring disease severity (Luo et al., 2021), progression (Demichev et al., 2021), and mortality prediction (Lin et al., 2021) through the interpretation of routine blood tests. For instance, a recent meta-analysis from Li et al. demonstrated that computational methods based on multi-center clinical datasets could generate more accurate COVID-19 diagnosis, stratify patients into clusters of severity and discriminate them from Influenza with 97.9% specificity (Li et al., 2020). Applications such as the previous example become even more relevant when applied to low-income

underdeveloped countries where access to diagnostic workflow is limited and the need for real-time point-of-care systems for disease screening is imperative.

John McCarthy first outlined the concept of AI in 1956 during the Dartmouth conference, on which several scientists discussed the concept of “thinking machines” in different areas such as abstraction, creativity, computational theory, natural language processing, and neural networks (Kline, 2011). Since then, progress slowed and remained stationary until 2012, when an ImageNet-DL-Algorithm triggered significant attention for the technology, with high-accuracy performance classification metrics that disrupted the current state-of-the-art (Krizhevsky et al., 2012). AI is defined as a computer science subdivision that aims to automatically understand and create intelligent systems based on high amounts of data (Shukla Shubhendu and Vijay, 2013). In medicine, the inequities and deficiencies that arose from the global COVID-19 pandemic catalyzed a boost in AI applications. Therefore, it aims to deliver effective, high-quality care, leveraging increasing clinical world data to democratize and decentralize health into patient care. The transformation of a patient’s blood analysis into a probability state to epitomize a likely diagnosis is already a reality (Gunčar et al., 2018).

The purpose of improving population health and patient care with parallel reduction of healthcare costs supports the implementation of AI strategies in the medium and long-term periods. Bajwa et al., 2021 Concepts such as precision medicine, ranging from diagnostics to prognostics and therapeutics with connected care, are under development (The Medical Futurist, 2022). In parallel, AI strategies disrupt the classical paradigm of scientific knowledge construction. Instead of collecting small datasets that try to answer sequential questions (classical approach), the new paradigm settles on collecting high amounts of data where scientists try to find multiple answers directly (Ahmad et al., 2021). However, significant challenges arise during this new paradigm: the black-box nature of AI algorithms endorses the need to generate explanatory, comprehensive systems able to dialogue with the physician to justify each clinical prediction or outcome (Bruckert et al., 2020). Also, legal and regulatory matters are under development, which will be crucial to regulate how AI algorithms are built and how continuous learning is evaluated.

Here, we focus on routine blood analysis as a proxy for determining pathological states supported by AI algorithms. We offer a comprehensive description of the ML pipeline with contextualization on the learning strategies (machine, reinforcement, deep, and federated learning), model development (application, preprocessing, modelling, and validation), and clinic deployment. We summarize the pathologies based on general health parameters (summarized according to their function and associated causes of variation), their inherent classification performance, and principal findings associated with model development and selected blood parameters. Finally, we discuss challenges related to clinic deployment and suggest future research directions for the development of models.

Overall, this review provides guidance for future research by summarizing reports combining AI and routine blood tests to diagnose disease or prognosis. Finally, it describes the methodologies used and contributes to the continued use of this technique in providing deeper insights into the potential of non-

appraised blood metabolites in traditional clinical decision support systems.

How AI learns

Currently, AI drives innovation processes involving analytical (data-driven decision-making), functional (operating according to analytical AI), interactive (communication), textual (nature language processing), and visual (augmented reality) technologies. AI enables the development of models to solve real-world problems based on different learning strategies, such as machine learning, deep learning, data mining, rule-based modelling, fuzzy logic, knowledge representation, case-based reasoning, text mining, visual analytics, and optimization, among others (Sarker, 2022). Next, we will briefly explain these learning strategies.

ML is a pattern recognition method that automatically detects regularities in large amounts of data. Based on statistical methods, this process evaluates interactions between variables and finds the most effective way of using them to reach a predetermined goal without requiring human intervention to define a strict set of rules or programming hypotheses (Kerr et al., 2012). ML has become the preferred framework for deploying AI applications, supported and leveraged by the continuous increase of data availability (big data). Although these concepts are similar and closely related, they are distinct: pattern recognition is one possible approach to artificial intelligence, and machine learning is a way to pattern recognition (Alsuliman et al., 2020). Data is considered mandatory for the model’s development, and it is commonly available in different forms such as structured (highly organized on relational databases), unstructured (without pre-defined format), semi-structured (organized but not on relational databases), and metadata (data properties about data) (Sarker, 2021).

ML algorithms such as gaussian naïve Bayes (GNB), k-nearest neighbors (KNN), support vector machines (SVM), decision trees (DT), linear regression (LR), or (Box 1) are the most common techniques generally applied for supervised learning strategies (Table 1 provides a comprehensive list). These algorithms use sample inputs for model development and

BOX 1 Glossary of key terms.

AUROC	The area under the receiving operating characteristic curve computed with the true positive <i>versus</i> false positive rates. It provides an aggregate measure of performance across all possible classification thresholds
Bootstrap	A statistical technique for sample extraction with replacement, allowing repeated training and fixed test
Cross-validation	The re-sampling method used to test and train different portions of data in several iterations of the model development
Ensemble	Combination of base estimators’ predictions to improve robustness and generalizability over a single estimator
Feature	Information input into the model during training and evaluation
Kernel	The function applied to the original non-linear data to create higher-dimensional spaces in which data will become separable
Overfitting	Process in which the statistical model adapts perfectly to the training data but does not generalize well on new data
Training set	The subset of data used for the model’s learning and optimization

TABLE 1 A comprehensive list of supervised ML algorithms for classification according to the desired learning strategy (Pedregosa et al., 2011).

Learning	Algorithm	Description
Linear	Linear regression (LR)	The target value is a linear combination of the features
	Lasso	Uses variable selection and regularization to estimate sparse coefficients, improving the accuracy and interpretability of the model
	Linear discriminant analysis (LDA)	Transforms input data into a linear subspace that maximizes separation between classes and predicts targets in closed-form solutions. Can also be used to provide dimensionality reduction in high-dimensional data sets
	Generalized linear models (GLM)	Statistical model that could be used to model dependent variables that represent non-normally distributed data, i.e., glmnet.
	linear regression (LR)	Statistical approach that seeks to generate a model from a collection of data that allows the prediction of values taken by a categorical variable (usually binary), using a series of continuous explanatory variables and/or binaries
	Multivariate logistic regression (MLR)	Extension of logistic regression to problems with more than two discrete outcomes
Bayesian	Naïve Bayes (NB)	This algorithm uses the Bayes' theorem with conditional independence between every pair of features to the value of the class variable
	Bayesian networks (BN)	These networks are adaptable since they fit joint probability distributions and allow knowledge extraction, reflecting how the occurrence of one variable is affected by the state of another
	Average two dependence estimators (A2DE)	A2DE achieves high accuracy by averaging among a small number of plausible Naïve-Bayes-like models that have fewer (and hence less detrimental) independence assumptions than Naïve Bayes
Nearest neighbors	K-Nearest Neighbors (KNN)	Uses the location of training samples that are closest in distance to the new point to estimate its label, based on the initial k data points
Support vector machines	Support Vector Machines (SVM)	Effective in smaller, but high-dimensional data sets, that assigns training examples to points in space in order to maximize the distance between the two categories. Different kernels could be used to evaluate new instances, which are mapped into that same space and classified according to which side of the gap they fall
Decision trees	Chi-square automatic interaction detection (CHAID)	A type of decision trees predicts the target value by learning simple decision rules inferred from data features. Specifically, it selects the most important feature using a chi-square measurement and iterates the procedure until all sub-informational data have a single choice
	classification and regression trees (CART)	CART builds binary trees by selecting the feature and threshold that provides the most information gain at each node
Neural Networks	Artificial Neural Network (ANN)	This algorithm uses non-linear functions made from one or two hidden layers between the input and the output dimensions
	Deep Neural Network (DNN)	DNN is an ANN with multiple layers between the input and output dimensions, and it is designed to emulate the principles and structure of a human neural network
	Multiple Layer Perceptron (MLP)	Non-linear function different from logistic regression since it can employ one or more non-linear layers between the input and output dimensions
	Shallow neural network	Employs a linear function in the second hidden layer of the two-layer network
	Recurrent Neural Networks (RNN)	It is cyclic DNN that loops outputs from specific nodes to affect the subsequent ones, and it is mainly applied to text-recognition and natural language processing
	Long short-term memory (LSTM)	It refers to a type of recurrent neural networks (RNNs) that learn long-term dependencies with feedback connections, and it is especially used in time-series data
Ensemble	Random forests	Random forests employ averaging to increase prediction accuracy and control overfitting by merging various trees, using different sub-samples of the training set, which decreases variances and results in a superior overall model
	Adaboost	Adaboost fits weak learners (models that are just slightly better than random guessing) in sub-samples of the training set and gets a final forecast by merging the guesses by majority vote
	Extra-trees	Extra-trees differ from traditional decision trees since it applies random splits across randomly selected features, picking the best split for creating the tree. In ensemble, it averages a meta estimator that fits randomized extra trees to improve accuracy and prevent overfitting
	Dynamic ensemble selection	Uses the most locally accurate decision classifier by calculating the accuracy of each individual classifier in specific local parts of the feature space surrounding a test sample

(Continued on following page)

TABLE 1 (Continued) A comprehensive list of supervised ML algorithms for classification according to the desired learning strategy (Pedregosa et al., 2011).

Learning	Algorithm	Description
	Gradient boosting (GBM)	Gradient boosting trains a large number of weak learners in a gradient descent function, where each learner minimizes the loss function of the previous model, resulting in an ensemble of learners that is improved incrementally until a stopping condition is achieved
	Extreme gradient boosting (XGBoost)	XGBoost is a regularized variation of GBM that controls overfitting and enhances performance by using linear and tree-based models that improve its capacity to execute parallel computation on a single computer, making it faster and more efficient
	Light gradient boosting (LightGBM)	LightGBM learns from data more effectively than standard GBMs because it employs histogram-based binning, which converts continuous feature values into discrete bins, reducing training time and memory consumption
	Category boosting (Catboost)	In opposition to GBM, lightGBM or XGBoost, Catboost uses symmetric and balanced trees, keeping the decision criteria consistent across all nodes, which makes this algorithm less prone to overfitting

subsequent data for model prediction. Apart from predicting specific diseases, other methods such as K-means, principal component analysis (PCA), or Pearson correlation (ρ) allow data exploration for clustering and dimensionality reduction through maximizing variance between samples. Thus, they enable an in-depth exploration of biomedical data with significant importance in medical diagnosis.

Supervised ML algorithms differ from unsupervised ones because they comprise training data wherein the real state of the data is known, for instance, which subjects have anemia and which are healthy. Based on training data, the algorithm generates a model that is applied to predict the state of a set of subjects for which the true state is unknown. These predictions settle in the form of a classification problem that identifies discrete states, such as different stages of anemia. Alternatively, they are established as a regression problem that evaluates continuous variables and predicts, for example, the numeric value of hemoglobin (Svensson et al., 2015). However, suppose the true state of the data is unknown. In that case, learning might be conducted unsupervised, where algorithms infer underlying patterns in unlabeled data to find sub-clusters of the original data, identify outliers, or produce low-dimensional representations. This way, it could be possible to recognize new associations that were not perceptible. In the example above, the algorithm could separate clusters of patients between anemic and non-anemic, even without knowing the true value of hemoglobin.

Reinforcement learning (RL) is a feedback-based approach where algorithms learn through trial and error by balancing the management of input knowledge with exploring unknown data. The model executes the task by understanding some basic rules and learns by weighting certain variables to find the correct solution. The supervisor should only indicate whether the algorithm's answer is correct; it is like supervised learning but delegates decision-making (weighting) to the algorithm's trial-error. An up-to-date application is the continuous management of oxygen flow rate for critically ill COVID-19 patients (Zheng et al., 2021), where the algorithm learned the appropriate flow rate for each patient, reducing the mortality rate and increasing the savings of oxygen-scarce resources in the pandemic. However, these algorithms are known as 'data-hungry' since they need large amounts of data to train different paths to

achieve sustainable performance, which is a limitation when applied to non-structured clinical information.

Deep learning (DL) is another class of machine learning, conceptually similar to the human brain since it mimics the inner mechanisms of brain neurons to transport and process data, create patterns, and enlighten decision-making. These algorithms extract high-level interactions between hidden layers of features from the input and learn complex interactions to develop accurate models from raw data. In medical diagnosis, this method has an application in image analysis, namely, on X-ray risk fracture diagnosis, breast density mammography analysis, or cardiovascular and pulmonary image reconstruction, all with 510(k) premarket notification clearance from the FDA (Benjamins et al., 2020). Deep learning requires minimal human intervention (except for sample labelling) but large amounts of curated data sets. Additionally, computational power is also relevant to conducting these tasks. Some researchers estimated that a life cycle to train several large AI models could emit nearly five times the CO₂ of an average American car (Hao, 2023). These challenges conducted the study and development of a new learning approach based on sparse modeling technology. The key differences compared to DL are the ability to provide comparable or even better accuracy results, working with small datasets, and performing feature extraction with much less computational power (1% of the energy required for DL). It also provides an explainable 'white box' the user can perceive (Fujiwara, 2021). Although this technology opens new routes in medical AI, the applications are still at a proof-of-principle and feasibility stage for cerebral infarction diagnosis assistance, liver cancer classification with a diagnosis support system, or anomaly detection in ECG signals. Therefore, it is still far from being approved for clinical deployment.

Federated learning (FL) is a new learning paradigm aiming to correct limitations in the current state-of-the-art model development for data governance, privacy, update, and sharing. FL moves the model to the data instead of input data for modeling. This approach enables training common AI models from multiple independent data sources (with proprietary data governance, privacy, and access policies) to deploy unbiased, generalizable, and appropriate-fitting models. The most established workflow of FL was proposed by Brendan McMahan et al. (2017).

This workflow includes the distribution of the global model on independent ‘clients’ that train the model in its data and send the adjusted local model to the global server to perform the trained models’ aggregation; this cycle repeats until the global model converges. [Dayan et al. \(2021\)](#) implemented an FL approach for the COVID-19 prognosis of oxygen supplies on symptomatic patients using the inputs of vital signs, laboratory data, and chest X-rays from 20 institutes. The federated model impacted prediction metrics across all participating sites (trained locally) by an average increase of 16% and 38% in AUC and generalizability, respectively.

While learning remains a matter of study with newer approaches being developed ([Kairouz et al., 2021](#)), the pipeline of AI-based prediction models is still under standardization. The increasing number of reports in this field compels the establishment of guidelines not to gauge the quality of the prediction models but to provide indications for transparent and unified reporting of this matter.

Since this review focuses on AI application to blood parameters to extract clinical value, we next provide a brief overview of the most common routine blood tests before going deep into how AI has been used to extract clinical value from them.

Blood and routine blood tests

Blood is the only fluid tissue present in the human body. Typically, an average adult has 6–7 L of blood in total. Cell elements compose approximately 45% of the blood; the remaining 55% is the fluid portion, designated plasma or serum. Many diseases cause changes in blood composition; therefore, blood analysis is important in clinical diagnosis ([Badrick, 2013](#)). Routine blood tests (RBT) typically merge the hematology and biochemistry analysis to explore changes in the cellular and molecular parts of the blood ([Table 2](#)). Depending on the type of blood analysis, laboratory workflows require at least two independent blood-collecting tubes for the separate study of hematology and biochemistry, which often forces the patient to provide 4–10 mL of venous blood.

In hematology, cell blood count (CBC) is the most performed exam. It includes not only the analysis of the three most important types of cells, erythrocytes (red blood cell, RBC), leukocytes (white blood cell, WBC), and thrombocytes (platelet) but also comprises differential information on WBCs subgroups (lymphocytes, segments, monocytes, eosinophils, basophils). Hematology also enlightens on hemoglobin concentration (Hb), hematocrit percentage (HTC), mean corpuscular volume (MCV), mean cell hemoglobin concentration (MCHC), and the red cell distribution width (RDW) ([Celkan, 2020](#)).

Biochemistry panels examine other chemical substances, such as electrolytes, hormones, and proteins. The portion of the blood that remains after all blood cells removal is composed mainly of water (90%), proteins (9%) that regulate plasma osmotic pressure and are important in the transport of fatty acids, thyroid and steroid hormones, and other chemical substances (1%) such as gases, nutrients, and vitamins ([Marieb and Hoehn, 2012](#)). The serum refers to plasma without clotting factors, i.e., fibrinogen, and is commonly used for chemistry testing and coagulation studies ([Chatburn and Hematology, 2010](#)). A general health blood

parameters panel typically includes CBC with differential, comprehensive metabolic and lipid panels, uric acid, GGT, and TSH ([Richard et al., 2011](#)).

Methods

We conducted a literature review of studies published between 2012 and 2022 that used artificial intelligence methodologies, namely, machine learning algorithms, to extrapolate clinical outcomes from routine blood tests. Using the query ‘artificial intelligence OR machine learning AND routine blood tests’ in the PubMed[®] electronic database, we found 164 articles that proceeded to the screening stage. Rayyan Management Software was used to import discovered reports, conduct study selection, and apply eligibility criteria.

Original English-language studies that reported diagnosis or prognosis of ICD-10 diseases based only on predefined blood parameters, namely, RBC, Hemoglobin, Hematocrit, MCV, RDW, WBC, Neutrophils, Eosinophils, Basophils, Lymphocytes, Platelets, Glucose, Urea, Creatinine, Potassium, Sodium, Chloride, Albumin, ALP, ALT, AST, Calcium, Triglycerides, HDL-cholesterol, Total cholesterol, LDL-cholesterol, Uric acid, GGT, TSH and C-reactive protein, were eligible for inclusion. The analysis did not include studies that included other biofluids parameters or reviews, systematic reviews, meta-analyses, protocols, commentaries, or book chapters.

The International Statistical Classification of Diseases and Related Health Problems 10th revision (ICD-10) was used to categorize the 54 studies, which were divided into 10 disease classes: infections (or parasitic diseases) (9), neoplasms (6), blood (3), endocrine (nutritional or metabolic) (5), mental (behavioral or neurodevelopmental) (2), circulatory (3), respiratory (2), digestive (5), genitourinary (1), and particular diseases (COVID-19) (18).

An overview of machine-learning studies based on routine blood tests for diagnosis or prognosis of ICD-10 pathologies

Most of the studies covered in this review fall into the diagnosis category; we identify the exceptions in the outcome column of each study (see tables below). We next describe how machine learning has been applied to extract clinical value from routine blood tests for specific diseases (using the ICD-10 classification as reference); [Table 3](#) to 13 summarize information (outcome, sample, selected features, methods, and most relevant findings) for each study.

Infectious or parasitic diseases (ICD-10 class I)

The infections or parasitic diseases studied include the human T-lymphotropic virus, bacteremia, bloodstream infection, general bacterial infection (in the surgical room and at hospital admission), malaria, and dengue. The studies used the traditional blood-based indicators of infection stated in [Table 3](#), namely, white blood cells,

TABLE 2 General health panel required in routine blood examinations (Matthew and Pincus, 2011).

	Parameter	Description	Below normality, related causes	Above normality, related causes
CBC with differential	RBC	Hemoglobin transport	Anemia; Blood loss; Bone marrow disorders; Cancer (certain types)	Low oxygen related to heart disease, pulmonary fibrosis, smoking, or high-altitude living; High consumption of anabolic steroids such as erythropoietin; Myeloproliferative diseases such as polycythemia vera; kidney diseases
	Hemoglobin	Oxygen and carbon dioxide exchange from the lungs to the tissues	Anemia; Blood loss; Thalassemia	Same as RBC causes
	Hematocrit	The proportion of red blood cells in the whole blood	Anemia; Blood loss; Cancer (certain types)	Dehydration, smoking or high-altitude living; Heart, lung, or kidney diseases; Polycythemia vera
	MCV	The average volume of red blood cells in the whole blood	Iron-deficiency anemia; Thalassemia; Lead-poisoning; Chronic disease	Folic acid or B12 deficiency; Preleukemia; Immune hemolytic anemia; Liver disease
	RDW	Size (anisocytosis) of red blood cells	-	Heart, kidney, or liver disease; Diabetes; Cancer
	WBC	Immunity	Chemotherapy; Myelodysplastic syndrome; Autoimmune disorders; Leukemia; HIV	Viral or bacterial infection; Inflammation; Rheumatoid arthritis; Pregnancy; Allergies, smoking, or stress
	Neutrophils		Viral infection; Hepatitis; Aplastic anemia; Lupus	Bacterial infection
	Eosinophils		Bacterial infection	Allergies; Parasitic infection
	Basophils		Hyperthyroidism; Allergies; Infections	Chronic inflammation; Hypothyroidism; Myeloproliferative disorders
	Lymphocytes		Infections; Tuberculosis; Drug reactions; Stress	Viral infections (i.e., Epstein-Barr virus)
Monocytes		Bone marrow disorders; infections; Systemic lupus erythematosus	Infections; rheumatoid arthritis; chronic myelomonocytic leukemia	
Platelets	Blood coagulation	Cancer (leukemia, lymphoma); Viral infections; Anemia (certain types); Chemo and radiotherapy	Genes mutation (essential thrombocythemia); Infection; Cancer; Inflammation; Iron deficiency	
Metabolic	Glucose	Energy regulation	Diabetes treatment; Drug reactions	Diabetes mellitus; Infection (severe)
	Urea	A waste product of protein digestion	A low-protein diet (malnutrition); Severe liver damage	Dehydration; Urinary tract obstruction; Congestive heart failure or recent heart attack; Kidney malfunction
	Creatinine	A waste product of muscles	Muscle diseases; Excess water loss; Liver diseases	Dehydration; High-intensity exercise; Kidney malfunction (stones, infection, failure)
	Potassium	Electrolyte on body fluid regulation and nerve function	Vomiting or diarrhea; Kidney damage	Diabetes mellitus; Advanced renal failure; Alcohol, burns
	Sodium	Electrolyte on body fluid regulation and nerve function	Vomiting, diarrhea, or burns; Nephritis or diabetic acidosis; Kidney or heart failure	Severe vomiting, diarrhea, or burns; Dehydration, excessive sweating, or adrenal glands disorders
	Chloride	Electrolyte on blood volume and osmotic pressure regulation	Severe vomiting, diarrhea, or excessive sweating Congestive heart failure, lung disease	Dehydration Kidney disease or Cushing's syndrome
	Albumin	A protein carrier for hormones, vitamins, and enzymes and prevents leaking on blood vessel	Kidney, liver, digestive, or thyroid diseases; Malnutrition or infection	Dehydration, severe diarrhea; Steroids, insulin, and hormones intake
	ALP	An enzyme that removes the phosphate group of several proteins	Malnutrition, vitamin deficiency; Hypothyroidism	Liver or bone disorders
	ALT	An enzyme that converts alanine for energy production	Chronic kidney disease; B6 vitamin deficiency	Liver disease; Hemochromatosis; Mononucleosis
	AST	An enzyme that catalyzes aspartate conversion	Kidney, liver, or cancer disease; B6 vitamin deficiency; Autoimmune or genetic conditions	Bruising, trauma, necrosis; Infection; Neoplasia of liver or muscle

(Continued on following page)

TABLE 2 (Continued) General health panel required in routine blood examinations (Matthew and Pincus, 2011).

	Parameter	Description	Below normality, related causes	Above normality, related causes
	Calcium	Mineral with a vital role in muscle tone and excitability	Acute pancreatitis; renal disease; D vitamin deficiency	Excess secretion of PTH; Cancer
Lipidic	Triglycerides	Lipid	Low-fat diet; Hyperthyroidism; Malabsorption syndrome	Liver, kidney, or thyroid disease; Alcohol, obesity, smoking; Uncontrolled diabetes
	HDL-cholesterol, direct	High-density lipoprotein	Unhealthy lifestyle; Smoking	Unhealthy diet; Genetics; Hypothyroidism
	Total cholesterol	Lipoprotein	Malnutrition or malabsorption; Anemia; Thyroid or liver disease	Unhealthy diet and lifestyle; Obesity
	LDL-cholesterol, calculated	Low-density lipoprotein	Hemorrhagic stroke; Cancer; Anxiety or depression	Unhealthy lifestyle; Genetics; Age
Others	Uric acid	A waste product of purines metabolism	Wilson's disease; Fanconi syndrome; Alcoholism	Unhealthy lifestyle; Diabetes mellitus; Alcoholism
	GGT	Enzyme	Unhealthy diet; B6 or magnesium deficiency	Liver or bile ducts disease
	TSH	Hormone	Hyperthyroidism	Hypothyroidism
	C-reactive protein	Acute phase reactant protein	-	Inflammation; Bacterial or viral infections; Autoimmune disorders; Heart attack; Sepsis

RBC, red blood cells; MCV, mean corpuscular volume; RDW, red cell distribution width; WBC, white blood cells; ALP, alkaline phosphatase; AST, aspartate transferase; ALT, alanine transaminase; GGT, Gama-glutamyl Transferase; TSH, thyroid stimulating hormone; HIV, human immunodeficiency virus; PTH, parathyroid hormone.

platelets, glucose, creatinine, albumin, AST, and CRP. However, 'exceptions' associated with feature selection were verified in surgical-site infection (ALP and sodium), malaria (RBC, MPV, MCV), bacteremia (MPV), and bloodstream infection (RDW). The studies on surgical-site infection were concerned with the relationship between time of blood analysis and the prediction of a diagnosis. Significant progression was accomplished by Kocbek et al., with the prediction of the full lasso model (AUC=0.95) at different timeframes (60, 30, and 15 days), benefiting from the findings of Soguero-Ruiz et al. (AUC=0.87) stating the importance to adjust the temporal structure of blood analysis to increase classification performance (Soguero-Ruiz et al., 2015; Kocbek et al., 2019). While ALP was selected for post-operative surgical infection, sodium was featured for surgical site infection. Malaria was studied for supervised classification with discrimination between uncomplicated malaria (UM), severe malaria (SM), and non-malaria infections (nMI). Distinctions were modeled by an artificial neural network with three different layers using distinctive features for each discrimination. Interestingly, SM separated from nMI based on the unique combination between MPV and MCV as classifiers of SM (AUC=0.98). UM and nMI were also distinguished based on PLT, RBC, and LY (AUC=0.86). The approach to Dengue differed from the approach to malaria. Ho et al. evaluated the probability of the condition in a predefined timespan of 35 weeks (Ho et al., 2020). Clinical data was fed into a deep neural network in competition with other learners, reaching an internal validation that surpassed sensitivities of 90% in a 3% prevalence cohort. Surprisingly, the weak learners performed similarly, indicating the clinical data value based only on age, temperature, and WBC. The approach from Sarbaz et al. outlined an infection by the human t-lymphotropic virus type I, a retrovirus known to be asymptomatic in most cases and evolve to malignancy and neural diseases in a few patients (Sarbaz et al., 2013). Even so, the dataset used has a relative

balance between three outcomes: normal (n=101), leukemia (n=94), and HTLV-I (n=107). The supervised classification model is based on a decision-trees algorithm-CHAID, which is the chi-squared automatic interaction detection, evaluating the association between input features exploring the levels of the three to maximize the classification performance. The internal validation achieved excellent performance (AUC>0.90) with a sensitivity of 95.8% in recognition of patients based on leukocytes, platelets, and percentage of eosinophils information. Bacteremia was initially studied in 2014 by Ratzinger et al. in a cross-sectional study with the largest cohort associated with infectious diseases (n=15,985) with 1,286 presenting a positive blood culture result: *E. coli* (n=406), *S. aureus* (n=297), *K. pneumoniae* (n=83) and others (n=500) (Ratzinger et al., 2014). The dataset split kept the 8% prevalence of bacteremia in training and validation sets, and the statistical analysis identified NE/WBC as the most important individual predictor (AUC=0.694). The A2DE algorithm (naïve-Bayes-based) produced two models with similar performance: model 1 (20 variables, NPV=0.966) and model 2 (10 variables, NPV=0.966). Results of internal validation kept the classification performance constant and selected age, creatinine, CRP, eosinophil, bilirubin, lymphocytes, monocytes, monocytes (%), neutrophils (%), and sodium as important predictors of bacteremia. Mooney et al. focused on a pregnant or post-partum cohort, where the bacteremia prevalence was lower (nearly 3% in 255 patients) (Mooney et al., 2021). The random forests classifier achieved an NPV of 97.4%, supported by the NLR, MPV, and BA indexes. Finally, Zoabi et al. evaluated bloodstream infection with a gradient-boosting decision tree and compared the results of the full (AUC=0.83) and compact (AUC=0.81) models with the standard conventional scores (AUC=0.62) (Zoabi et al., 2021). The evaluation of the model was made available, differentiating this study from the previous and enabling a prospective assessment of the method.

TABLE 3 Machine-learning-based routine blood tests for the diagnosis (or prognosis) of certain infections or parasitic diseases.

Study	Outcome	Sample	Selected features	Methods (AUC)	Findings
Sarbaz et al. (2013)	HTLV-I	101 (normal) 94 (leukemia) 107 (HTLV-I)	WBC, PLT, EO%	L: supervised, classification; FS: Pearson correlation; C: CHAID (accuracy 91%); V: NA	HTLV-I distinguished from leukemia or normal patients with clinical data based on differential CBC. External validation: NA Clinical deployment: NA
Ratzinger et al. (2014)	Bacteremia	15,985 (1,286)	NE/WBC	L: supervised, classification; FS: wrapper approach; C: A2DE-20 variables (0.76) and A2DE-10 variables (0.75); V: 0.80 and 0.78	Low-risk group: NPV >98.8%. External validation: NA Clinical deployment: NA
Soguero-Ruiz et al. (2015)	Surgical-site infection (post-operative)	1,005 (101)	Thrombocytes, ALP, CRP, Albumin, Creatinine, WBC	L: supervised, classification FS: RBF-RFE C: non-linear SVM (0.87) V: leave-one-out cv	Adjusting the temporal structure of blood tests improves the system's accuracy. External validation: NA Clinical deployment: NA
Rawson et al. (2019)	Bacterial infection (hospital admission)	104 (35%)	CRP, WBC, Creatinine, ALT, Bilirubin, ALP	L: supervised, classification FS: NA C: SVM (0.84) V: 10-fold cv	Infection predicted in a timeframe of 72 h after admission. External validation: NA Clinical deployment: NA
Kocbek et al. (2019)	Surgical site infection	1,137 (233)	CRP, WBC, Sodium, Hb, Thrombocytes, Albumin	L: supervised, classification FS: NA C: Full Lasso Model (0.95) V: repeated hold-out cv	Infection was predicted based on three timeframes of 60, 30, and 15 days before surgery. External validation: NA Clinical deployment: NA
Moranga et al. (2020)	Malaria	2,207 (UM=703) (SM=526) (nMI=978)	UM ≠ nMI: PLT, RBC, LY; SM ≠ nMI: MPV, MCV; SM: RBC, PLT	L: supervised, classification FS: NA C: ANN [UM ≠ nMI (0.866), SM ≠ nMI (0.983)], V: 10-fold cv	Models are classified based on the combination of PLT, RBC, LY, LY%, and MPV. External validation: NA Clinical deployment: NA
Ho et al. (2020)	Dengue	4,894 (2,942)	Age, Temperature, WBC	L: supervised, classification FS: NA C: DNN (0.858) V: 10-fold cv	For all three models, pre-peak sensitivities (<35 weeks) were higher than 90%. External validation: NA Clinical deployment: NA
Mooney et al. (2021)	Bacteremia (pregnant and post-partum)	255 (129)	NLR, MPV, BA	L: supervised, classification FS: NA C: RF (0.98) V: 10-fold cv	NLR >20 achieved a negative predictive value of 97.4% for a 3% prevalence cohort. External validation: NA Clinical deployment: NA
Zoabi et al. (2021)	Bloodstream infection	7,889 (2,590)	Albumin, RDW, Creatinine	L: supervised, classification FS: NA C: Decision-Tree (gradient boosting): inclusive (0.82), compact (0.81) V: cross-validation	ML showed substantial improvement in the AUC score compared to traditional methods (0.83vs0.62 on the inclusive model) and (0.81vs0.62 on the compact model). External validation (proxy): available at github.com/nshomron/infecpred Clinical deployment: NA

AUC, Area under the ROC (receiver-operating characteristic) curve, HTLV-I, Human T-lymphotropic virus type-I, UM, uncomplicated malaria; nMI, Non-Malarial Infections; SM, severe malaria; WBC, white blood cells; PLT, platelets, EO% eosinophils count, NE, neutrophils; ALP, alkaline phosphatase, CRP C-Reactive Protein, Hb Hemoglobin, LY, lymphocytes, LY% lymphocytes count, MPV, mean platelet volume; MCV, mean corpuscular volume; RBC, red blood cells; NLR, Neutrophil-to-Lymphocyte ratio, BA, basophils; RDW, red cell distribution width, L Learning, FS, feature selection, C Classification, V Validation, CHAID, Chi-Squared Automatic Interaction Detection; NA, not available, A2DE, Averaged 2-Dependence Estimator; RBF-RFE, Radial Basis Function - Recursive Feature Elimination; SVM, support vector machines; CV, Cross-Validation; ANN, artificial neural networks; DNN, deep neural networks; RF, Random-Forests; CBC, cell blood count; NPV, negative predictive value; ML, machine learning.

Neoplasms (ICD-10 class II)

The application of AI and blood parameters for neoplasia herein reviewed include colorectal, leukemia (pediatric acute lymphoblastic, leukemias differentiation and lymphocytic prognosis), and non-specified cancer diagnosis in a primary care center. Erythrocytes, hemoglobin, hematocrit, RDW, leukocytes,

platelets, aspartate aminotransferase (AST), calcium, and LDL-cholesterol were the Cancer-related blood parameters identified (Table 2). Studies related to colorectal cancer were Table 4 highly consistent since they built on a high dimensional sample size (>10k patients), employed a supervised decision tree classification algorithm with similar internal validation (AUC=0.82 and AUC=0.81), and externally validated with equal or higher

TABLE 4 Machine-learning-based routine blood tests for the diagnosis (or prognosis) of neoplasms.

Study	Outcome	Sample	Selected features	Methods (AUC)	Findings
Kinar et al. (2016)	CRC	Israel: 606,403 (3,135) UK: 25,613 (5,061)	Hb, MCH, MCHC, HTC, MCV, RDW	L: supervised, classification FS: NA C: Decision trees (0.82) V: cross-validation	The model's performance on a 10–12-month time window achieved AUC=0.79
					Sensitivities at a 6-month time window were 10% higher compared to anemia guidelines
					External validation: UK (0.81)
					Clinical deployment: NA
Hornbrook et al. (2017)	CRC	17,095 (900)	Gender, Birth year, CBC	L: supervised, classification FS: NA C: Decision trees (0.81) V: cross-validation	The CRC detection model performed best in detecting cecum and ascending colon tumors rather than in transverse and sigmoid colon and rectum
					External validation: MHS (Israel) (0.87), NHS (0.85)
					Clinical deployment: CRC program in Israel
Mahmood et al. (2020)	Pediatric ALL	94 (50)	PLT, Hb, WBC, Gender	L: supervised, classification FS: CART C: CART (0.87) V: 10-fold cv	Platelet abnormality significant predictor in pediatric ALL.
					External validation: NA
					Clinical deployment: NA
Soerensen et al. (2022)	Cancer diagnosis within 90 days on primary care	Cohort I: 5,224 (1,042) Cohort II: 1,712 (1,368)	ALB, PLT	L: supervised, classification FS: NA C: LR (0.80), ANN (0.91) on cohort I and LR (0.79), ANN (0.79) on cohort II V: NA	Reduced albumin and increased platelet levels increase cancer risk in a concentration-dependent way
					External validation: NA
					Clinical deployment: NA
Haider et al. (2022)	Leukemias differentiation: AML, APML, CML, ALL, CLL, Other's	1,577: (354), (96), (213), (272), (153), (489)	CBC	L: supervised, classification FS: NA C: ANN (0.83) V: NA	CBC not only differentiates from six lineages of leukemia but also remains predictive for the type (acute, chronic, or other)
					External validation: NA
					Clinical deployment: NA
Meiseles et al. (2022)	Prognosis of Lymphocytic leukemia treatment within 2 years	109	Hb, Time from diagnosis, RDW, NLR	L: supervised, classification FS: NA C: GBM using inexpensive features (0.86); decision trees (0.74) V: 10-fold cv	Low NLR and high values of RDW are relevant predictors for treatment need
					External validation: NA
					Clinical deployment: NA

AUC, Area under the ROC (receiver-operating characteristic) curve; CRC, colorectal cancer; ALL, acute lymphoblastic leukemia; AML, acute myeloid leukemia; APML, acute promyelocytic leukemia; CML, chronic myeloid leukemia; CLL, chronic lymphoid leukemia; Hb Hemoglobin; MCH, mean corpuscular hemoglobin; MCHC, mean corpuscular hemoglobin concentration; HTC, hematocrit; MCV, mean corpuscular volume; RDW, red cell distribution width; CBC, cell blood count; PLT, platelets; WBC, white blood cells; ALB, albumin; NLR, Neutrophil-to-Lymphocyte ratio; L Learning, FS, feature selection, C Classification, V Validation, CART, classification and regression trees algorithm, NA not available; GBM, gradient boosting model; CV, Cross-Validation; UK united kingdom; MHS, maccabi healthcare services; NHS, national health service.

performance (AUC=0.81, AUC=0.87, and AUC=0.85). This consistency endorses the relevance of its findings. Noteworthy, different outcomes were verified: while Kinar et al. demonstrated sensitivity stability between 480 and 240 days before diagnosis (AUC≈0.76) with a posterior increase in the last 240 days (AUC>0.80) (Kinar et al., 2016), Hornbrook et al. identified sub-regions of colorectal cancer that were better diagnosed, namely, the cecum and the ascending colon (Hornbrook et al., 2017).

Regarding leukemia, Mahmood et al. began to evaluate the ability to diagnose pediatric acute lymphoblastic leukemia (ALL) in a small cohort (n=94), where fifty patients had the disease

(Mahmood et al., 2020). The study findings were achieved through the comparison of four classifiers wherein the classification and regression trees (CART) performed better (accuracy=0.87) with a decision tree that included low platelet (43%) and hemoglobin (24%) levels and high levels of white blood cells (4%). The disease was furtherly distinguished by Haider et al. on a set of other pathologic conditions: acute myeloid leukemia (AML, n=354), acute promyelocytic leukemia (APML, n=96), chronic myeloid leukemia (CML, n=213), and chronic lymphoid leukemia (CLL, n=153) (Haider et al., 2022). The authors based the approach on a conventional cell blood count analysis and developed an artificial

neural network to classify the six lineages of the disease: AML (AUC=0.905), APML (AUC=0.805), CML (AUC=0.937), CLL (AUC=0.870) and ALL (AUC=0.829). Internal validation sets increased overall accuracy from 83.1% to 84.7%, which denotes a non-overfitted model. Meiseles et al. evaluated the prognosis of treatment needed in 2 years for patients with lymphocytic leukemia with a dataset of 109 patients (Meiseles et al., 2022). The outcome was predicted with a gradient boosting model (GBM, AUC=0.768) and compared with a general linear model (GLM, AUC=0.753), both with higher performances when related to the current scoring system for prognostic evaluation of patients with CLL (CLL-IPI, AUC=0.52). Despite predicting the general progress of the disease, the CLL-IPI does not evaluate the necessity of the treatment, and even a simple decision tree based on inexpensive features—Hb, time since diagnosis, NLR, and RDW—achieved higher performance (AUC=0.74).

Finally, a non-specified cancer diagnosis was approached by Soerensen et al., through the modeling (training and internal validation) on cohort I (n=6,266 from 2011 to 2018) and the evaluation on cohort II (n=3,080 from 2019 to 2020). The primary outcome was “cancer within 90 days,” and the proposed methodology included an artificial neural networks *versus* logistic regression approaches (Soerensen et al., 2022). The results were slightly different since ANN predicted better in the modeling cohort (AUC=0.91) but decreased its performance in the evaluation cohort (AUC=0.79); in opposition, LR demonstrated higher stability in both cohorts (n1, AUC=0.80, and n2, AUC=0.79). The concentration decrease in albumin with a dependent increase in platelet levels was related to an increased risk of cancer, even for patients whose metabolite relationship was verified in ‘normal’ ranges.

Diseases of the blood or blood-forming organs (ICD-10 class III)

The approach to blood diseases fundamentally combined standard CBC parameters with artificial neural networks. The outcomes included the diagnosis of iron deficiency anemia in women, thalassemia minor (TM) in the general adult population, and the distinction between iron-deficiency anemia and β -thalassemia in three scenarios (males, females, and both). The approach from Yilmaz et al. involved studying several ANN strategies to verify which one had the best accuracy without performing feature selection in the dataset (Yilmaz and Bozkurt, 2012). The accuracy results obtained were highly similar between the studied strategies (accuracies \geq 0.98), and comparison with previous studies (Azarkhish et al., 2012) showed a slight increase in sensitivity from 0.968 to 0.976, conferring an excellent opportunity to perform an external validation of the model with consequent valid clinical deployment. In thalassemia minor, Magen et al., studied a cohort of 185 verified alpha and beta TM patients with a control group that included IDA, myelodysplastic (MDS), and healthy subjects (Barnhart-Magen et al., 2013). Despite feeding the ANN with six CBC metabolites, only RBC, RDW, and MCV values achieved higher metrics (Table 5). However, the specificity of 1.00 (TM vs. healthy and MDS) decreased to 0.90 (TM vs. healthy,

MDS, and IDA). Çil et al. reported improvement with the distinction of β -thalassemia and IDA studied in gender groups with different algorithms (weak learners and neural networks) (Çil et al., 2020). The principal findings include a common gender RELM algorithm (specificity=0.966), an ELM, RELM for females (specificity=0.952), and an SVM model for males (specificity=0.938). While these scores surpassed the previous studies, the sample size was small, limiting the study findings.

Endocrine, nutritional, or metabolic diseases (ICD-10 class IV)

This review reports endocrine, nutritional, or metabolic diseases mainly applied to the diagnosis of type 2 diabetes mellitus (T2DM) and the prediction of blood glucose on type 1 diabetes mellitus (T1DM) and ensuing disease complications (insulin resistance, polyneuropathy, and iatrogenic hypoglycemia). (Table 6) Regarding diagnosis, Kopitar et al. approached early T2DM on a cohort of 3,723 individuals employing different ML algorithms without significant improvements related to diagnosis accuracy or newly relevant features (Kopitar et al., 2020). Indeed, the authors concluded that the model’s stability in linear regression was preferred against other learning algorithms, and the increased data available in electronic health records was useful to update prediction models and stabilize important features: hyperglycemia, age HDL-cholesterol, and triglycerides.

Kushner et al. studied T1DM blood glucose prediction using a shallow neural network based on historical continuous blood glucose monitoring (Kushner et al., 2020). The results improved the current condition through a more extended prediction (t=240min vs. 120min) with lower error (RMSE, 60min=28 mg/dL vs. 43 mg/dL). Bernardini et al. initially featured disease complications with evaluating clinical factors associated with insulin resistance (Bernardini et al., 2019). The ensemble regression forest allowed the identification of non-glycemic blood parameters (HDL and total cholesterol, age, uricemia, WBC, and GGT) as clinical factors that could provide early detection of glucose deterioration. These findings agree with previous literature that individually associated uricemia and WBC to insulin-resistant conditions and GGT in high-risk T2DM individuals. Higher sample studies employed supervised classification algorithms for risk prediction of polyneuropathy (n=5,846) and iatrogenic hypoglycemia. Regarding polyneuropathy, the authors found that different ML models produced different features selection and consequent classification metrics, relating co-morbidities (nephropathy or retinopathy) to a rise in ANN (AUC=0.892), increased neutrophil levels in random forests boosting (AUC=0.898) or blood glucose levels in linear regression (AUC=0.892). Of notice, the principal finding suggests that the choice of the ML algorithm should consider not only the performance metrics but also the kind of clinical information to assess: the identification of early (i.e., ANN) or late biomarkers (i.e., linear regression) of polyneuropathy, or the identification of pathophysiological mechanisms (i.e., decision trees). The risk of developing iatrogenic hypoglycemia (glucose \leq 70 mg/dL) was approached by Mathioudakis et al. using a stochastic gradient boosting ML model in an extensive data study (n=1 612,425)

TABLE 5 Machine-learning-based routine blood tests for the diagnosis (or prognosis) of diseases of the blood or blood-forming organs.

Study	Outcome	Sample	Selected features	Methods (AUC)	Findings
Yilmaz and Bozkurt (2012)	Women's IDA	Training: 2000(NA) Test: 600 (122)	RBC, Hb, HCT, MCV, MCH, MCHC	L: supervised, classification FS: NA C: ANN (0.99) V: test-set	ANN and medical diagnosis achieved comparable results. ANN training with several strategies (FFN, CFN, DDN, TDN, LVQ, PNN) produced similar results (accuracy ≥ 0.98). External validation: NA Clinical deployment: NA
Barnhart-Magen et al. (2013)	TM	526 (185)	RBC, Hb, MCV, RDW, MCH, PLT	L: supervised, classification FS: NA C: ANN (specificity=0.967, sensitivity=1) V: test-set	ANN only differentiates TM from the control group based on MCV, RDW, and RBC. External validation: NA Clinical deployment: NA
Çil et al. (2020)	β -thalassemia and IDA distinction	342 (152)	RBC, HCT, MCV, MCH, MCHC, RDW	L: supervised, classification FS: NA C: several weak learners and ANN (>0.90)V: test-set	Different models were best according to gender: SVM for males, RELM for both, and ELM and RELM for females. External validation: NA Clinical deployment: NA

AUC, Area under the ROC (receiver-operating characteristic) curve, IDA, iron deficiency anemia; TM, thalassemia minor; RBC, red blood cells, Hb Hemoglobin; HCT, hematocrit; MCV, mean corpuscular volume; MCH, mean corpuscular hemoglobin; MCHC, mean corpuscular hemoglobin concentration; RDW, red cell distribution width; PLT, platelets, L Learning, FS, feature selection, C Classification, V Validation, NA, not available; ANN, artificial neural networks; FFN, feedforward networks; CFN, cascade forward networks; DDN, distributed delay networks; TDN, time delay networks; LVQ, learning vector quantization; PNN, probabilistic neural network; SVM, support vector machines; RELM, Regular Over-learning Machine; ELM, extreme learning machine.

TABLE 6 Machine-learning-based routine blood tests for the diagnosis (or prognosis) of endocrine, nutritional, or metabolic diseases.

Study	Outcome	Sample	Selected features	Methods (AUC)	Findings
Bernardini et al. (2019)	Clinical factors related to insulin resistance	968	HDL cholesterol, Total cholesterol, Age, Uricemia, WBC, GGT	L: supervised, regression FS: permutation out-of-bag C: Ensemble RF + data imputation (MSE <0.17) V: 10-fold cv	The ensemble approach correlated with insulin resistance based on non-glycemic blood data. External validation: NA Clinical deployment: NA
Kopitar et al. (2020)	Early T2DM diagnosis	3,723	Hyperglycemia, Age, HDL cholesterol, Triglycerides	L: supervised, regression FS: NA C: lm (0.747), glmnet (0.740), lightgbm (0.723), xgboost (0.715), RF (0.723) V: 10-fold cv	No clinically relevant improvement with more sophisticated ML algorithms. Higher variables' stability is preferred for model calibration and clinic interpretation. External validation: NA Clinical deployment: NA
Metsker et al. (2020)	Risk of diabetes polyneuropathy	5,846 (2,342)	Retinopathy, Nephropathy, Hb, Neutrophils, ALT, AST, Glucose	L: supervised, classification FS: NA C: ANN (0.892), SVM (0.864), decision tree (0.898), lm (0.892), logistic regression (0.894) V: 5-fold cv	Different models showed different results in terms of the feature's importance and significance: lm (glucose), rf (neutrophils), and ANN (co-morbidities). Depending on the needs, the choice of the algorithm should vary. External validation: NA Clinical deployment: NA
Mathioudakis et al. (2021)	Risk of iatrogenic hypoglycemia	1 612,425 (50,354)	Basal insulin dose, BG coefficient of variation, Previous hypoglycemic episodes	L: supervised, classification FS: NA C: MLR, RF, NB, SGB (0.90) V: 10-fold cv	Iatrogenic hypoglycemia predicted after short-term blood glucose measurement in-hospital based on EHR data. External validation: Hospital 2 (0.88), Hospital 3 (0.87), Hospital 4 (0.86), Hospital 5 (0.86) Clinical deployment: NA
Kushner et al. (2020)	Blood glucose prediction in T1DM	24	Historic continuous glucose monitoring	L: supervised, regression FS: NA C: shallow neural network (RMSE): t=60 (28 \pm 4), t=90 (33 \pm 4), t=120 (38 \pm 6), t=180 (40 \pm 8), t=240 (43 \pm 12) mg/dL V: test-set	93% of predictions were clinically acceptable, according to the Clarke error grid. External validation: NA Clinical deployment: NA

AUC, Area under the ROC (receiver-operating characteristic) curve, T2DM, Type 2 diabetes mellitus, T1DM, Type 1 diabetes mellitus; HDL, High-Density Lipoprotein; WBC, white blood cells; GGT, Gama-glutamyl Transferase, Hb Hemoglobin; ALT, alanine transaminase; AST, aspartate transferase; BG, blood glucose, L Learning, FS, feature selection, C Classification, V Validation, NA, not available; RF, Random-Forests; MSE, mean squared error; CV, Cross-Validation, lm Linear Regression Model, glmnet Regularized Generalized Linear Model with Lasso (Least Absolute Shrinkage and Selection Operator) Regression, LightGBM, Light Gradient-Boosting Machine, XGBoost Extreme Gradient Boosting, RF, random forests; ANN, artificial neural networks; SVM, support vector machines; MLR, multivariable logistic regression; NB, Naïve-Bayes; SGB, stochastic gradient boosting; EHR, electronic health records.

(Mathioudakis et al., 2021). Performance metrics of the developed model (43 predictors) were slightly lower (c-statistic=0.86:0.90) than previous reports (c-statistic=0.80:0.99) but was the first to be

externally validated in 4 different hospitals with stability in model predictions, working 24 h after each blood glucose measurement.

TABLE 7 Machine-learning-based routine blood tests for the diagnosis (or prognosis) of mental, behavioral, or neurodevelopmental disorders.

Study	Outcome	Sample	Selected features	Methods (AUC)	Findings
Dipnall et al. (2016)	Depression associated biomarkers	5,227	RDW, Glucose, Total bilirubin	L: supervised, regression FS: multiple imputation, boosted regression, imputed weighted logistic regression C: multivariate weighted logistic regression V: cross-validation	The hybrid approach provided a variable selection of three biomarkers for the prediction of depression. External validation: NA Clinical deployment: NA
Hochman et al. (2021)	Depression	Training: 7,702 (522) Validation: 1,752 (117)	Family income, GGT, Glucose, Triglycerides, RDW, Creatinine, BA%, EO%, Bilirubin	L: supervised, classification FS: backward feature selection C: random forests: full (0.83), overweight (0.80), diabetes (0.82), metabolic syndrome (0.82) V: cross-validation	Selected features demonstrated good predictive value in distinguishing depression cases on the four studied datasets. External validation: full (0.69), overweight (0.63), diabetes (0.66), metabolic syndrome (0.64) Clinical deployment: NA

AUC, Area under the ROC (receiver-operating characteristic) curve, RDW, red cell distribution width; GGT, Gama-glutamyl Transferase, BA% basophils count, EO% eosinophils count, L Learning, FS, feature selection, C Classification, V Validation, NA, not available.

Mental, behavioral, or neurodevelopmental disorders (ICD-10 class V)

Using AI, depression was the only condition studied in the context of mental, behavioral, or neurodevelopmental diseases (Table 3). Despite the association between depression and routine blood biomarkers still being under clarification, low HDL-cholesterol values were previously associated with the condition. The studies reviewed approached depression under the NHANES database differently: while Dipnall et al. used data mining, machine learning, and traditional statistics to identify related biomarkers (Dipnall et al., 2016), Hochman et al. aimed to build a low-cost diagnostic tool to perform diagnosis based on blood data (Hochman et al., 2021). The methodology described in the first study explains the feature selection process in three sequential hybrid processes: multiple imputations, ML regression, and traditional statistical regression. From 67 laboratory parameters, the workflow selected 21 after ML regression and only six after univariate analysis. The final multiple logistic regression model suggested two related effects (hemoglobin from bilirubin and cotinine from cadmium), which resulted in the exclusion of Hb and cotinine. The posterior cadmium elimination occurred since only RDW, glucose, and total bilirubin remained significant to several confounder covariates, namely, age ($p < 0.05$). The authors explained related literature associations between the selected biomarkers and depression, yet all with indirect relationships. The subsequent study from Hochman et al. configured a supervised approach for predicting depression using a random forests classifier in four subgroups (Hochman et al., 2021). Feature selection was made using the stepwise backward method, which starts modeling with all features and successively eliminates the least important feature in iterative steps until all features are removed from the model. Results were similar across the four groups with full dataset [ratio of income to poverty (RIP), GGT, glucose, triglyceride and RDW, AUC=0.83], overweight and obesity (GGT, RIP, creatinine, RDW and glucose, AUC=0.80), diabetes (GGT, eosinophils, RIP, basophils and eosinophils, AUC=0.82) and patients with metabolic syndrome (RIP, GGT, eosinophils, bilirubin and basophils, AUC=0.82). Table 7 Despite the developed models accounting for the features selected in the first study, namely, glucose and RDW (full dataset and overweight and obesity), and bilirubin (patients with metabolic syndrome), the results of internal validation did maintain the performance in the

external validation dataset (AUC, average=0.66); this fact compromises the predictive ability of the developed models.

Diseases of the circulatory system (ICD-10 class IX)

Outcomes related to the reviewed circulatory system diseases include the prognosis of postoperative blood coagulation in children with congenital heart disease and the diagnosis of cardiac workload and ischemic stroke. Numerous studies refer to associations between blood analysis and diseases of the circulatory system. However, known routine blood tests associated with heart disease are the low levels of sodium and chloride and the elevated levels of erythrocytes, hematocrit, RDW, urea, and c-reactive protein. The prognosis of postoperative blood coagulation in children was assessed by comparing three different classifiers (decision trees, naïve-Bayes, and support vector machines). Applying recursive feature elimination resulted in seven features, age being the most relevant (Table 8). Traditional statistical tests also evaluated relevant features, which confirmed the significance among the compared groups (abnormal vs. normal blood coagulation). This statistical verification also supports the model's reliability, which achieved accuracy values of 75% in internal validation based on a typical CBC. The cardiac workload is generally measured by the rate pressure product (RRP), which is the product between systolic blood pressure and heart rate. The study from Shou et al. evaluated how blood parameters predicted the biochemical profile related to the resting RRP through the analysis of 55,730 individuals (Shou et al., 2021). The supervised regression task was accomplished by comparing a linear regression model ($r=0.352$) and a tree-based model, XGBoost ($r=0.377$). The authors found that glucose alone predicted rRRP with a Pearson correlation of 0.247 in the linear model and 0.245 in the non-linear model; total protein and neutrophils count were responsible for the additional variance, exhibiting the recognition ability of ML-based approaches to find new biomarkers. Indeed, Zheng et al. followed a similar strategy for ischemic stroke (Zheng et al., 2022). Ischemic stroke is still a major burden due to the high number of miss-diagnosed (or late) cases due to challenges related to the triaging process. Four feature selection techniques (univariate logistic regression, least absolute

TABLE 8 Machine-learning-based routine blood tests for the diagnosis (or prognosis) of circulatory system diseases.

Study	Outcome	Sample	Selected features	Methods (AUC)	Findings
Guo et al. (2021)	Postoperative blood coagulation in children with congenital heart disease	1,690	Age, Sex, MCV, MCH, MCHC, WBC, PLT	L: supervised, classification FS: recursive feature elimination C: DT (0.81), NB (0.82), SVM (0.82) V: 5-fold cv	The accuracy rate of the overall forecast was higher than 75%; Age was the most important feature for the decision-tree model. External validation: NA Clinical deployment: NA
Shou et al. (2021)	Cardiac workload	55,730	Glucose, Total protein, Neutrophil	L: supervised, regression FS: NA C: LR (r=0.352), XGBoost (r=0.377) V: NA	Positive correlation between the measured resting rate pressure (rRRP) with the predicted rRRP based on blood biomarkers. External validation: NA Clinical deployment: NA
Zheng et al. (2022)	Ischemic stroke	15,475 (4,999)	Age, NE%, NE, MO%, MCHC, LY%, RDW-CV, MCV, Hb, Total cholesterol, HDL-cholesterol, uric acid, total protein	L: supervised, classification FS: permutation feature importance C: XGBoost (0.91) V: 5-fold cv	The model was developed based on 15 routine blood tests and externally validated with excellent accuracy. External validation: 5,011 (1,076), XGBoost (0.92) Clinical deployment: available online at istriage.com

AUC, Area under the ROC (receiver-operating characteristic) curve, MCV, mean corpuscular volume; MCH, mean corpuscular hemoglobin; MCHC, mean corpuscular hemoglobin concentration; WBC, white blood cells; PLT, platelets, NE% neutrophils count, NE, neutrophils, MO% monocytes count, MCHC, mean corpuscular hemoglobin concentration, LY% lymphocytes count, RDW-CV, Red Cell Distribution Width-Coefficient of Variation; MCV, mean corpuscular volume, Hb Hemoglobin; DT, decision trees; NB, Naïve-Bayes; SVM, support vector machines; CV, Cross-Validation, L Learning, FS, feature selection, C Classification, V Validation, NA, not available; LR, linear regression, XGBoost Extreme Gradient Boosting, rRRP, resting rate pressure.

shrinkage and selection operator regression, recursive feature elimination, and the spearman correlation) were applied to the training set, reducing 41 to 15 features. Model development was assessed by comparing six algorithms (XGBoost, RF, NN, LR, gaussian NB, KNN); XGBoost showed the best performance with an accuracy of 0.84, 0.83, and 0.86 in training, internal validation, and external validation, respectively. The model was further analyzed by explaining techniques (permutation feature importance, local-interpretable model-agnostic, and Shapley additive explanations) endorsing the importance of neutrophils count, total protein, HDL-cholesterol, and hemoglobin. Aiming for a future clinical deployment, the model was also made available online for prospective validation.

Diseases of the respiratory system (ICD-10 class X)

Regarding respiratory system diseases, we present here one study related to asthma. Given the impact of smoking on respiratory function, we included a study with AI and smoking-related disorders in the review. Routine blood metabolites associated with smoking were the high levels of erythrocytes, hematocrit, leukocytes, triglycerides, and the low levels of HDL-cholesterol, none related to asthma. Indeed, the study of Mamoshina et al. found that HDL-cholesterol was the principal feature for the classification of the smoking status, along with hemoglobin, RDW, and mean cell volume (Mamoshina et al., 2019). These findings were accomplished after an iterative analysis that started with the prediction (regression) of biological age based on routine blood tests. The feature importance shows HbA1C, urea, glucose, and ferritin as the most important (training). The 24 features selected were used to predict age in smokers ($r^2=0.55$) and non-smokers ($r^2=0.57$), showing a potential impact of smoking in the

prediction. The addition of the feature 'smoking status' improved the three tested regression models from 0.56 to 0.57 (23–24 features), 0.54 to 0.58 (20–21 features), and 0.55 to 0.60 (18–19 features) in the prediction of biological age. Authors also found, based on the same models of 23, 20, and 18 features, the ability to predict the 'smoking status' with an accuracy of 0.82 (equivalent for the three models), with HDL-cholesterol, hemoglobin, RDW, and MCV the most relevant features for the prediction. The study from Zhan et al. employed a Mahalanobis-Taguchi system (MTS) to classify asthma patients (Zhan et al., 2020). The algorithm was approached by constructing the Mahalanobis space (collection and distance calculation of the standardized normal and abnormal data), with further identification of useful variables (orthogonal arrays and signal-to-noise ratios for threshold definition and roc curve analysis). Results achieved with the proposed algorithm were compared with an SVM model, where the same features (selected by Pearson correlation) predicted asthma patients with similar accuracy (Table 9). The authors claim a more straightforward interpretability of the model by calculating the Mahalanobis distance (MD) with the values of PDW, MPV, WBC, eosinophils count, lymphocytes count, and MCHC data.

Diseases of the digestive system (ICD-10 class XI)

For digestive system diseases, we focused on studies applying supervised classification methods to diagnose conditions related to liver disease. In general, pathologies related to the liver split into four stages: the inflammation stage [induced by hepatitis B virus (HBV), hepatitis C virus (HCV), alcoholic liver disease (ALD), and nonalcoholic fatty liver disease (NAFL)], the fibrosis stage, the cirrhosis stage and the worst stage related to liver cancer or failure (Tian et al.,

TABLE 9 Machine-learning-based routine blood tests for respiratory system disease diagnosis (or prognosis).

Study	Outcome	Sample	Selected features	Methods (AUC)	Findings
Mamoshina et al. (2019)	Smoking status and aging in smokers	149,000 (49,000)	Smoking: HDL-cholesterol, Hb, RDW, MCV Age: HbA1C, urea, glucose, ferritin	L: supervised, classification (smoking), regression (aging) FS: permutation feature importance C: Feed-forward deep neural networks: Age ($r>0.74$), smoking (accuracy >0.81) V: 5-fold cv	Blood tests could quantify aging caused by smoking; still, this method was less accurate than DNA methylation. External validation: NA Clinical deployment: NA
Zhan et al. (2020)	Asthma	1,835 (355)	PDW, MPV, WBC, EO%, LY %, LY, MCHC	L: supervised, classification FS: person correlation C: MTS, 7 var (sensitivity=0.941); SVM, 7 var (sensitivity=0.935) V: 10-fold cv	MTS showed high classification accuracy on asthma patients (94.15%) and healthy volunteers (97.20%) based on 7 routine blood parameters; SVM achieved similar performance. External validation: NA Clinical deployment: NA

AUC, Area under the ROC (receiver-operating characteristic) curve, HDL, High-Density Lipoprotein, Hb Hemoglobin; RDW, red cell distribution width; MCV, mean corpuscular volume, HbA1C Glycated Hemoglobin, PDW, platelet distribution width; WBC, white blood cells, EO% eosinophils count, LY% lymphocytes count, MCH, mean corpuscular hemoglobin; MTS, Mahalanobis-Taguchi system; DNA, deoxyribonucleic acid, L Learning, FS, feature selection, C Classification, V Validation, NA, not available; CV, Cross-Validation; SVM, support vector machines.

2022). Regarding inflammation, Fialoke et al. studied the discrimination between non-alcoholic steatohepatitis (NASH) and simple steatosis in NAFL (Fialoke et al., 2018). Since NASH is underdiagnosed due to the lack of patient symptoms and relevant biomarkers (high values of AST and ALT), the authors trained 4 ML algorithms with the available data: demographics, the maximum, minimum, and mean values of AST, ALT, AST/ALT, PLT, and the binary diabetes condition. 5-fold cross-validation displayed the model's AUC higher than 0.83, being XGBoost the top classifier (AUC=0.876) with the potential to perform external discrimination with promising results. Ma et al. also approached the inflammation stage by diagnosing NAFLD in a cross-sectional study of 10,030 individuals with a prevalence of 24% (Ma et al., 2018). Four techniques were used for feature selection, and 11 ML algorithms were trained. Selected five biomarkers (Table 10) resulted in different performance metrics across the tested traditional (KNN, SVM, LR, NB, BN, DT), ensemble (AdaBoost, bagging, RF), and extension algorithms (hidden naïve-Bayes, aggregating one-dependence). Since F-measure (harmonic mean between precision and recall) was considered the most important metric, the Bayesian network achieved the best model (F-measure=0.655). Comparisons with current diagnostic scores such as the FLI [calculated with triglycerides, BMI, GGT, waist circumference (F-measure=0.318)], and HIS [estimated with the values of AST, ALT, BMI, diabetic condition, and gender (F-measure=0.524)] demonstrated a superior diagnostic ability of the developed Bayesian network. Cao et al. evaluated HBV-induced liver cirrhosis (inflammation and cirrhosis stages) by studying seven routine blood tests enhanced by a multilayered perceptron and a naïve-Bayes algorithm (Cao et al., 2013). Both classifiers exhibited higher AUC in the internal validation (MLP, AUC=0.942, and NB, AUC=0.899) rather than the training, with better performance for the MLP (MLP, AUC=0.900, and NB, AUC=0.831). This study also compared the ML metrics with the currently used scores APRI (AUC=0.726), gauged with the AST to PLT index, and the FIB-4 (AUC=0.817), calculated with the age, PLT, AST, and ALT levels, with the MLP classifier shown superior performance, enabling a potential reduction in the number of biopsies to perform diagnosis. The worst stage of liver disease—liver failure - was studied by Peng et al. to create a forecast model to predict patient deterioration after hospitalization. This type of prediction is routinely assessed through the model for end-stage liver disease (MELD) calculated

using the values of creatin, total bilirubin, standardized prothrombin ratio (INR), and the etiology of the disease. In opposition, the authors used a database of 15 clinical metabolites kept for modeling determined by hepatologists; only variables with high missing values were discarded. Except for the GLM model, all other models (AUC >0.794) outstand the classification performance of the MELD (AUC=0.699). However, the limitation of the sample size (n=348) reinforces the need to perform validation in a high number of subjects. Finally, Yao et al. approached non-specific liver disease by deep learning on the largest dataset (n=76,914), which comprised 12,688 patients with different stages of liver disease (Yao et al., 2020). The application of a dense deep neural network (DNN) was compared with standard logistic regression and random forests. The network was explored based on the network width (number of neurons per hidden layer) and dropout rate. Widths of 512 (AUC=0.8919) and 1,024 (AUC=8,922) were compared along with dropouts of 0.3 (AUC=0.8812), 0.4 (AUC=0.8891), 0.5 (0.8919), 0.6 (0.8904), and 0.7 (0.8856). Feature importance was assessed with random forests (for reference) since DNN and DenseDNN are black-box algorithms with poor explainability. Global results achieved excellent internal validation (AUC >0.87) except for logistic regression (AUC=0.79). Indeed, a significant improvement was not verifiable between an explainable random forest and the deep learning approaches for diagnosing non-specific liver disease. The fibrosis stage and liver cancer have not been approached yet.

Diseases of the genitourinary system (ICD-10 class XIV)

Concerning disorders of the genitourinary system, we focus on chronic kidney disease (CKD). CKD lacks early diagnosis since obvious symptoms only appear in an advanced stage of the disease wherein the patient's renal function declines with a glomerular filtration rate (GFR) of 60 mL/min/1.73 m² (Tarwater, 2011). The need to create screening procedures that perform early diagnosis motivated several studies for routine blood and urine analysis. Indeed, the study of Mahfuz et al. evaluated 250 CKD patients in a cohort of 400 individuals with information regarding urine (specific gravity, albumin, sugar, red blood cells, pus cell, pus cell clumps, bacteria) and blood (glucose, urea, creatinine, sodium,

TABLE 10 Machine-learning-based routine blood tests for the diagnosis (or prognosis) of digestive system diseases.

Study	Outcome	Sample	Selected features	Methods (AUC)	Findings
Cao et al. (2013)	HBV-induced liver cirrhosis	239 (124)	Age, ALT, AST, PT, PLT, Hb, RDW	L: supervised, classification FS: genetic search C: MLP (0.942), NB (0.899) V: 10-fold cv	Compared to currently used scores for liver cirrhosis prediction (APRI (AUC=0.726) and FIB-4 (AUC=0.817)), the developed MLP achieved excellent performance in the test set. External validation: NA Clinical deployment: NA
Fialoke et al. (2018)	NASH in NAFL	34,949 (17,359)	ALT_mean, ALT_max, AST_max, AST mean	L: supervised, classification FS: genetic search C: LR (0.835), DT (0.842), RF (0.870), XGBoost (0.876) V: 5-fold cv	The model improved by adding longitudinal (temporal) data rather than only using recent values External validation: NA Clinical deployment: NA
Ma et al. (2018)	NAFL	13,030 (2,522)	BMI, Triglycerides, GGT, ALT, Uric acid	L: supervised, classification FS: correlation, redundancy analysis, out-of-bag estimation, Scott-Knot test C: BN (F=0.655) V: 10-fold cv	Tested ML algorithms improved the prediction accuracies from nearly 52% in FLI and HIS rules to >80% for NAFL diagnosis. External validation: NA Clinical deployment: NA
Yao et al. (2020)	Liver disease	76,914 (12,688)	AST, Total bilirubin, Direct bilirubin, Age	L: supervised, classification FS: RF C: LR (0.797), RF (879), DNN (0.886), DenseDNN (0.891) V: 5-fold cv	AUC was slightly higher in deep learning than in weak learners; Selected features were achieved by random forests since DNNs are black-box algorithms. External validation: NA Clinical deployment: NA
Peng et al. (2020)	Exacerbation risk in patients with liver dysfunction	348 (174)	AST, NE, LY, Creatinine, ALT, ALB, Total protein, Total bilirubin	L: supervised, classification FS: manual C: ANN (0.912), CART (0.794), GLM (0.554), SVM (0.853) V: 10-fold cv	While the MELD achieved an AUC of 0.669, ML algorithms enhanced the prediction to nearly 80% (except GLM). External validation: NA Clinical deployment: NA

AUC, Area under the ROC (receiver-operating characteristic) curve, NASH, Non-Alcoholic Steatohepatitis; NAFL, Non-Alcoholic Fatty Liver Disease; ALT, alanine transaminase; AST, aspartate transferase; PT, platelets, Hb Hemoglobin; RDW, red cell distribution width; BMI, body mass index; GGT, Gama-glutamyl Transferase; NE, neutrophils; LY, lymphocytes; ALB, albumin, L Learning, FS, feature selection, C Classification, V Validation, NA, not available; MLP, multilayer perceptron; NB, Naïve-Bayes; CV, Cross-Validation, XGBoost Extreme Gradient Boosting, LR, linear regression; DT, decision trees; RF, random forests; BN, bayesian network; DNN, dense neural networks; CART, classification and regression trees algorithm; GLM, generalized linear models; SVM, support vector machines; APRI, aspartate aminotransferase to platelet ratio index; FIB-4, Fibrosis Index Based on 4 factors; ML, machine learning; MELD, Model for End-Stage Liver Disease.

potassium, hemoglobin, packed cell volume, white blood cell count and red blood cell count) metabolites. The authors performed training on five algorithms and performed feature importance based on the SHAP technique, reducing the number of features from 24 to 13 in concordance between tested gradient boosting, random forest, and extreme gradient boosting. With the 13 selected features, the authors manually split the dataset into six different subsets: all features, blood and others, urine and others, only blood, only urine, and only others. A new train-test cycle applied to these subsets resulted in a classification accuracy ranging from 76% to 99%. Interestingly, results were very similar between all features (RF, AUC=0.99) and only blood (RF, AUC=0.97), with slight variations between the tested classifiers. (Table 11) While this study provides an interesting interpretation approach to CKD screening based on different bundles of metabolites (SHAP explained), it lacks sample size, which limits the findings.

Codes for special purposes (ICD-10 class XXII)

Lastly, we included an analysis for special purposes ICD-10 codes, in which, for instance, the coronavirus 2019 (COVID-19) disease is included. COVID-19 is caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2); noticeably, it

received particular interest in AI-based diagnostics. First reported in November 2019, this virus emerged as a pandemic in March 2020, accounting for 563M infections (Our World in data, 2023) and 6.37M deaths (Data, 2023). The evolution of the virus was irregular, and its spread was facilitated by the struggle to achieve a real-time diagnosis able to distinguish between real positive COVID-19 infections from other viral and bacterial respiratory infections. Indeed, the symptomatology of COVID-19 remains challenging to differentiate from other infections: 40% of patients display mild disease (fever, cough), 40% show moderate disease (pneumonia), 15% are severe (shortness of breath), and 5% refers to critical illness (ICU admission) (Wu and McGoogan, 2020). Additionally, 9–12 days is the average time for ICU admission, with a median length of stay of 9 days. The median length of mechanical ventilation is 8.4 days, and COVID-19 mortality in the ICU is nearly 30% (Auld et al., 2021). Regarding diagnosis, reverse transcription polymerase chain reaction (RT-PCR) and computerized tomography (CT) images are still recognized technologies for determining viral infection. However, both methods comprise disadvantages: CT yields radiation (inherently endorses the risk of cancer development), it is bulk, expensive and hinders the possibility of performing screening. RT-PCR tests are less costly, available in higher volume, and offer a specificity close to 100% and, depending of the primers and strain, a remarkably high sensitivity (Böger et al., 2021). RT-PCR tests require laboratory

TABLE 11 Machine-learning-based routine blood tests for the diagnosis (or prognosis) of diseases of the genitourinary system.

Study	Outcome	Sample	Selected features	Methods (AUC)	Findings
Rashed-Al-Mahfuz et al. (2021)	Chronic kidney disease	400 (250)	Hb, Creatinine, Glucose, Urea, RBC, Sodium	L: supervised, classification FS: SHAP C: RF (0.97), GB (0.96), XGBoost (0.95), LR (0.94), SVM (0.94) V: 10-fold cv	Selected features (SHAP) were consistent with the literature regarding CKD diagnosis, and the performance of ML classifiers was similar for each bundle of features. Hemoglobin was the most important predictor External validation: NA Clinical deployment: NA

AUC, Area under the ROC (receiver-operating characteristic) curve, Hb Hemoglobin, RBC, red blood cells; SHAP, shapley additive explanations; RF, random forests; GB, gradient boosting, XGBoost Extreme Gradient Boosting, LR, linear regression; SVM, support vector machines, L Learning, FS, feature selection, C Classification, V Validation, CV, Cross-Validation; NA, not available; CKD, chronic kidney disease; ML, machine learning.

specialists and infrastructure and produce a 15% false-positive rate on 48–72 h of turnaround time. Rapid diagnostic tests (RDTs) emerged as a point-of-care solution to facilitate access to diagnosis and reduce dependence on laboratory infrastructures. There are more than 400 RDTs commercially available, based on two technologies: antigen-based (immunoassays) to detect domains of the surface proteins of the virus and molecular nucleic acid amplification tests (NAATs) that reveal the presence of viral gene targets (Diagnostics for All, 2023). The criteria for approval and commercialization of RDTs are based on a sensitivity superior to 80% and specificity above 98% (World Health Organization, 2021), tested on a prospective cohort study involving less than 30 persons infected with SARS-CoV-2 and 30 persons without the infection (Food and Drug Administration. Emergency use authorizations for medical devices, 2021). Accepted by the Food and Drug Administration (FDA), these standards do not require independent verification of clinical validation provided by each test manufacturer. Indeed, several studies reported varying degrees of sensitivity (36%–82%) and specificity (98%–100%) when these RDTs are tested in asymptomatic (Prince-Guerra et al., 2021). Notably, most validation studies conducted for RDTs were performed before the appearance of new variants, namely, delta and omicron. The WHO, CDC, and European Center for Disease Prevention and Control guidelines advise using these point-of-care solutions for diagnosing symptomatic persons and screening asymptomatic individuals. Despite the growing need for these solutions, especially in underdeveloped countries, supply-chain limitations hinder the availability and consequent clinical relevance of these tests.

COVID-19 researchers and clinicians explored AI state-of-the-art learning techniques to find alternatives for COVID-19 forecasting, management, surveillance, and recognizing scalable and cost-effective ways to deal with the pandemic. Regarding diagnosis, several studies emerged in mid-April 2020 through the study of routine blood tests using proprietary datasets (single center), characterized by low sample size ($n < 1,000$) and mostly without external validation. Joshi et al. provided an interesting methodology, modeling diagnosis using an L2-regularized logistic regression trained only with levels of hematocrit, neutrophils, and lymphocytes, achieving an internal validation (AUC=0.78) that was consistent with the evaluation performed in

four different sites (average AUC=0.77) (Joshi et al., 2020). Brinati et al. also used logistic regression but compared the classification performance with a random forest classifier, using 14 features (Table 12) (Brinati et al., 2020). Results were very similar among the internal test set, but the decision tree provided larger comprehension with AST (<25.4) and lymphocytes (<1.3) as major predictors of COVID-19 negativity. Alves et al. also employed a random forest classifier compared with five algorithms in which the ensemble achieved the best internal classification (AUC=0.87) (Alves et al., 2021). A decision tree explained the model, and criteria graphs allowed a visual interpretation of the association between selected blood parameters. An artificial neural network was designed by Banerjee et al. in comparison with a random forest and a lasso-elastic-net regularized generalized linear model (fitting a logistic regression) (Banerjee et al., 2020). The network was tested in community individuals ($n=619$) and patients in the hospital regular ward ($n=69$). While ANN and RF presented the best metrics for hospitalized and non-hospitalized patients, the glmnet identified a decreasing pattern in monocytes, leukocytes, eosinophils, and platelets that was applied to a logistic regression achieving an AUC of 0.85. The ensemble designed by Abayomi-Alli et al. was built under a small dataset ($n=279$), taking the input of 16 features. Comparisons were made between 15 classifiers wherein the ExtraTrees (AUC=0.99) and the AdaBoost (AUC=0.98) outperformed the remaining models. Wu et al. also achieved similar internal classification values using a slightly larger cohort ($n=603$) and a novel dynamic ensemble selection method, first approached with data imbalance techniques and modeled with a hybrid clustering with a posterior bagging classifier (Wu et al., 2021). The authors achieved better results with the hybrid approach rather than by using the bagging approach, tested in divisions 70:30 and 60:40 and with 5-fold cross-validation. Contrary to previous supervised studies, Souza et al. reported an unsupervised clustering approach based on self-organizing maps that detected positive COVID-19 patients with a discrimination power of 83% (LDA model) (Souza et al., 2021). This clustering approach was performed on 599 registers, of which only 81 were COVID-19 positive. It identified WBC, BA, EO, and RDW as features with a strong influence on clustering

TABLE 12 Machine-learning-based routine blood tests for the diagnosis (or prognosis) of COVID-19.

Study	Outcome	Sample	Selected features	Methods (AUC)	Findings
Brinati et al. (2020)	COVID-19 diagnosis	279 (177)	AST, LY, LDH, CRP, WBC, EO, ALT, Age, NE, GGT, MO, BA, ALP, PLT	L: supervised, classification FS: RF C: LR (0.84), RF (0.85) V: nested-cv	AST<25 is predictor of COVID-19 negativity (NPV=83%); AST>25 is predictor of COVID-19 positivity (PPV=76%) External validation: NA Clinical deployment: NA
Joshi et al. (2020)	COVID-19 diagnosis	390 (33)	NE, LY, HTC, Gender	L: supervised, classification FS: manual C: L2-regularized LR (c-statistic 0.78) V: cross-validation	NE and LY were negative predictors, while male and HCT were positive COVID-19 predictors. External validation: c-statistic 0.75, 0.75, 0.81 Clinical deployment: NA
Banerjee et al. (2020)	COVID-19 diagnosis	786 (81)	EO, WBC, RBC, MPV, BA, PLT	L: supervised, classification FS: glmnet; C: RF (0.94), Flexible ANN (0.95); V: 10-fold-cv	LR subtraction model between MO, WBC, EO, and PLT shows AUC=85% (community). External validation: NA Clinical deployment: NA
Plante et al. (2020)	COVID-19 diagnosis	12,183 (2,182)	EO, CA, AST, WBC, BA, RDW, RBC, ALB, TB, MCV, MCH, SO, HCO ₃ , UR, Chloride	L: supervised, classification FS: recursive feature elimination; C: XGBoost (0.91); V: 5-fold-cv	NPV for rule-out-ED >97% for 1%, 10%, and 20% covid-19 prevalence. External validation: XGBoost (0.91) Clinical deployment: NA
Tschoellitsch et al. (2021)	COVID-19 diagnosis	1,537 (65)	WBC, NLR, Hb, CA	L: supervised, classification FS: RF; C: RF (0.74); V: 5-fold-cv	Elevated WBC and NLR improved the model accuracy. External validation: NA Clinical deployment: NA
Alves et al. (2021)	COVID-19 diagnosis	608 (84)	WBC, PLT, EO, MO, CRP	L: supervised, classification FS: decision-tree-based; C: DTX + RF (0.86), LR (0.85), XGBoost (0.85), SVM (0.85), MLP (0.81), Ensemble (0.87); V: nested-cv	Explainable patterns based on selected features, according to previous literature. External validation: NA Clinical deployment: NA
Souza et al. (2021)	COVID-19 diagnosis	599 (81)	WBC, BA, EO, RDW	L: unsupervised, clustering FS: SOM; C: Neural Network SOM, LDA; V: NA.	Unsupervised pattern recognition applied to routine blood tests. External validation: NA Clinical deployment: NA
Cabitz et al. (2021)	COVID-19 diagnosis	1736 (NA)	Age, HCT, Hb, MCH, MCHC, MCV, RBC, WBC, PLT, NE, LY%, MO%, EO%, BA%, NE, LY, MO, EO, BA, Gender	L: supervised, classification FS: NA; C: SVM-RBF kernel (0.76); V: 10-fold-nested-cv	Meta-validation with robustness and cardinality implications in COVID-19 ML models states significant model degradation when tests are performed in different settings (equipment or populations) External validation: SVM-RBF (0.84) Clinical deployment: NA
Babaei et al. (2022)	COVID-19 diagnosis	279 (177) 1,624 (786) 600 (80)	WBC, PLT, MO, EO, Age quantile, CRP, RBC, Hb, LY, BA, CREA, NE, PO, UR, SO, AST, ALT, G	L: supervised, classification FS: SHAP; C: DNN (0.92), SVM (0.87), LR (0.85), NB (0.83), XGBoost (0.81), RNN (0.80), CNN (0.76), DT (0.72), KNN (0.72), LSTM (0.51). V: 4-fold-cv	Deep neural networks performed better than previous studies based on the same dataset; WBC, Age, AST, and LDH were predictors on the three datasets. External validation: NA Clinical deployment: NA
Wu et al. (2021)	COVID-19 diagnosis	603 (83)	Age, HTC, HGB, PLT, RBC, LY, MCHC, WBC, BA, MCH, EO, MCV, MO, RDW, G, CRP	L: supervised, classification FS: recursive feature elimination; C: Dynamic Ensemble Selection (0.99); V: 70–30 and 60–40 training-test and 5-fold-cv	Dynamic ensemble selection application on imbalanced data; External validation: NA Clinical deployment: NA
Campagner et al. (2021)	COVID-19 diagnosis	1,736 (816)	Age, HTC, Hb, MCH, MCHC, MCV, RBC, WBC, PLT, BA%, NE %, LY%, MO%, EO%, BA, NE, LY, MO, EO, BA, covid-19 specific symptoms, Gender	L: supervised, classification FS: recursive feature-elimination; C: SVM (0.975), LR (0.965), E (0.95), RF (0.945), NB (0.935), KNN (0.93); V: 5-fold-nested-cv	The most important predictors were RBC, MCV, NE, EO, and MO. External validation: SVM (0.98), SVM (0.97) Clinical deployment: NA
Abayomi-Alli et al. (2022)	COVID-19 diagnosis	279 (177)	Age, gender, WBC, PLT, CRP, AST, ALT, GGT, ALP, LDH, NE, LY, MO, EO, BA, swab	L: supervised, classification FS: PCA C: Extra-Trees (0.99), Adaboost (0.98), Decision tree (0.98) V: 10-fold cv	Strong comparison between many classifiers, with higher AUC on the proposed ensemble. External validation: NA Clinical deployment: NA

(Continued on following page)

TABLE 12 (Continued) Machine-learning-based routine blood tests for the diagnosis (or prognosis) of COVID-19.

Study	Outcome	Sample	Selected features	Methods (AUC)	Findings
Chadaga et al. (2022)	COVID-19 diagnosis	5,644 (558)	WBC, EO, PLT, MO	L: supervised, classification FS: Pearson correlation C: RF (0.80), LR (0.78), KNN (0.67), XGBoost (0.79) V: NA	Only internal validation was used to evaluate model performance on an imbalanced dataset (sampled with smote). External validation: NA Clinical deployment: NA
Chadaga et al. (2023)	COVID-19 diagnosis	1,169 (270)	ALB, TWBC, BA, SO, AST, PO, TB, DB, UR, TP, LY, NE, Hb, HTC, CREA, MO, NLR	L: supervised, classification FS: Grey wolf optimization (GWO) C: RF (0.99), LR (0.74), DT (0.88), KNN (0.83), STACKA (0.96), Adaboost (0.95), Catboost (0.96), LightGBM (0.98), XGBoost (0.99), STACKB (0.99), STACKC (0.98) V: 5-fold-cv	The RF model's results were interpreted using xAI (Explainable AI): albumin, TWBC, basophil, sodium, and AST are critical for distinguishing COVID-19 from other infections. Increased AST and decreased TWBC and basophils indicate infection with COVID-19
Luo et al. (2021)	COVID-19 severity	196 (129 ICU)	Age, WBC, LY, NE	L: supervised, classification FS: maximum relevance and minimum redundancy C: MCDM (TOPSIS + NB) (0.93) V: 80–20 (train-test)	Advanced age, low immunity, and combined bacterial infections are reasons for COVID-19 severity; The MCDM algorithm is stable on small datasets. External validation: NA Clinical deployment: NA
Benito-León et al. (2021)	COVID-19 severity	853 C1 (58 ICU) C2 (300 H) C3 (495 +)	C1: higher levels of AST, LDH, CRP, NE, and lower levels of MO and LY; C2: intermediate levels; C3: lowest AST, LDH, CRP, NE, and higher levels of MO and LY.	L: unsupervised, clustering FS: unsupervised; C: X-means; V: 80–20 (train-test)	Serum levels of AST, LDH, CRP, and NE were enough to separate patients' severity. External validation: NA Clinical deployment: NA
Famigliani et al. (2022)	COVID-19 severity	1,004 (181)	Age, LY, NE, MCHC, Gender, MCV, MO	L: supervised, classification FS: SHAP; C: MLP (0.71), DT (0.76), SVM (0.85), XGB (0.81); V: hold-out test set	Data consists of literature; CBC data could be used to predict ICU admission on COVID-19 patients. External validation: NA Clinical deployment: NA
Karthikeyan et al. (2021)	COVID-19 prognosis	370 (200 recovered) (170 death)	Age, NE, LY, LDH, hs-CRP	L: supervised, classification FS: NN forward feature selection; C: NN (0.99), LR (0.99), XGBoost (0.98), RF (0.98), SVM (0.99), DT (0.97); V: 80–20 (train-test) with 5-fold-cv	Higher levels of Age, hs-CRP, neutrophils, LDH, and lower levels of lymphocytes predicted mortality with 96% accuracy during the disease span. External validation: NA Clinical deployment: NA
Fernandes et al. (2021)	COVID-19 prognosis	1,040 (288 ICU) (106 MV) (92 M)	Age, LymCRP, CRP, Braden scale	L: supervised, classification FS: SHAP; C: MV: ANN, Extra Trees (0.94), RF, Catboost, Extreme Gradient Boosting M: ANN, Extra Trees (0.97), RF, Catboost, Extreme Gradient Boosting; V: 70–30 (train-test) with 10-fold-cv	ML algorithms could predict untrained outcomes (death) based on other outcomes (ICU + MV), with AUROC higher than 0.91. External validation: NA Clinical deployment: NA
Murri et al. (2021)	COVID-19 prognosis	921 (120 M)	Age, Hb, PLT, NE, SO, UR, CRP, SpO2	L: supervised, classification FS: LR C: LR (0.87) V: 5-fold-cv	Abnormal HGB, PLT, NE, high levels of URE, CRP, SO, and lower SpO2 were associated with an increased risk of death External validation: LR (0.82) Clinical deployment: NA

AUC Area under the ROC (receiver-operating characteristic) curve, AST Aspartate Transferase, ALT Alanine Transaminase, Hb Hemoglobin, MCH Mean Corpuscular Hemoglobin, MCHC Mean Corpuscular Hemoglobin Concentration, HTC Hematocrit, MCV Mean Corpuscular Volume, RDW Red Cell Distribution Width, CBC Cell Blood Count, ALB Albumin, NLR Neutrophil-to-Lymphocyte ratio, HDL High-Density Lipoprotein, WBC White Blood Cells, GGT Gama-glutamyl Transferase, RDW Red Cell Distribution Width, PLT Platelets, CRP C-Reactive Protein, LDH Lactate dehydrogenase, LY Lymphocytes, LY% Lymphocytes Count, EO Eosinophils, EO% Eosinophils Count, NE Neutrophils, NE% Neutrophils Count, MO Monocytes, MO% Monocytes Count, BA Basophils, BA% Basophils Count, ALP Alkaline Phosphatase, MPV Mean Platelet Volume, CA Calcium, ALB Albumin, TB Total Bilirubin, DB Direct Bilirubin, SO Sodium, TP Total Protein, HCO3 Bicarbonate, UR Urea, PO Potassium, CREA Creatinine, G Glucose, hs-CRP High-Sensitivity C-Reactive Protein, LymCRP Lymphocytes to C-Reactive Protein Ratio, SpO2 Oxygen Saturation, NN Neural Network, H Hospitalized, M Mortality, ICU Intensive Care Units, MV Mechanical Ventilation, PPV Positive Predictive Value, NPV Negative Predictive Value, ED Emergency Department, L Learning, FS Feature selection, C Classification, V Validation, CV Cross-Validation, NA Not Available, RF Random Forests, LR linear regression, glmnet Regularized Generalized Linear Model with Lasso (Least Absolute Shrinkage and Selection Operator) Regression, ANN Artificial Neural Networks, XGBoost Extreme Gradient Boosting, DTX Decision Trees Explainer, SVM Support Vector Machines, MPL Multilayer Perceptron, SOM Self-Organizing maps, LDA Linear Discriminant Analysis, RBF Radial Basis Function, DNN Dense Neural Networks, RNN Recurrent Neural Networks, CNN Convolutional Neural Networks, KNN K Nearest Neighbors, DT Decision Trees, LSTM Long Short-term Memory, PCA Principal Component Analysis, MCDM Multi Criteria Decision Making, TOPSIS Technique for Order of Preference by Similarity to Ideal Solution, STACKA Stacked model.

performance but was ambiguous regarding the feature range in outcome prediction. While reviewed studies improved the accuracies supported with more complex ML algorithms, the studies with higher sample sizes ($n > 1,000$) showed a similar increase in classification metrics in concordance with the addition of blood features. In a cohort of 1,537 participants, Tschoellitsch et al. achieved a moderate AUC of 0.74 and a negative predictive value of 98%, which agreed with previous results using random forest (Tschoellitsch et al., 2021). Cabitza et al. described a novel methodology comprising cardinality and similarity as metrics of model's reliability in external validation settings (Cabitza et al., 2021). Considering data regarding demographics and complete blood cell count, the SVM with RBF kernel was applied to eight different external datasets with AUC ranging from 0.66, 0.75, 0.80, 0.83, 0.87, 0.89, 0.97 and 0.98 and similarity values (according to the degree of correspondence) of 0.315, 0.341, 0.348, 0.444, 0.323, 0.447, 0.439, and 0.445, respectively. Babaei et al. compared the performance of 12 ML algorithms in three different datasets. In the third dataset, all algorithms' comparison performance exhibited DNN with the higher classification metrics (Table 12) (Babaei et al., 2022). Interestingly, the previous studies of Brinati et al. (Brinati et al., 2020), and Cabitza et al. (Cabitza et al., 2021), were also compared with DNN surpassing in the first dataset (AUC=0.92 vs. AUC=0.84, from Brinati et al. (Brinati et al., 2020)) and the second dataset (AUC=0.93 vs. AUC=0.84, from Cabitza et al. (Cabitza et al., 2021)), highlight deep neural networks as a promising approach for COVID-19 diagnosis. Plante et al. used a large cohort of 66 hospitals to perform an internal and external validation of an extreme gradient boosting tree based on 15 features. The external validation performed in 23 different hospitals led to the validation of the methodology (AUC=0.91) and allowed a deeper comprehension of the best cutoff score, independently of the disease prevalence (studied for 1%, 10%, and 20%). Campagner et al. validated six algorithms in two different sites (Bergamo, $n=245$ and Desio, $n=337$) with 42% and 48% of COVID-19 positive cases (Campagner et al., 2021). The models achieved an AUC always higher than 93%, with SVM achieving the best results on both external sets. Violin plots of specific key CBC parameters showed high similarity between the training and validation cohorts, namely, in white blood cells, neutrophils, lymphocytes, red blood cells, platelets count, and patient's age, which justifies the stability in the model's performance. Chadaga et al. used similar approaches in two public datasets: the Albert Einstein hospital in Brazil [$n=5,644$, RF (AUC) = 0.80] and the Dr. TMA Pai hospital in India [$n=1,169$, RF (AUC) = 0.99]. While both studies used smote to resolve imbalanced data, the second study used explainable methods to describe how parameters influenced the final decision. Significant improvements in performance metrics were observed (especially through the comparison of the same RF algorithm). Nonetheless, neither study has received external validation.

Regarding the diagnosis of COVID-19 severity, Benito-León et al. used a non-supervised clustering model (X-means) to differentiate intensive-care, hospitalized, and non-hospitalized positive patients (Benito-León et al., 2021). According to the David Bouldin index (lowest value refers to best cluster

distribution with higher intercluster distance and lower intracluster distance), the algorithm defined three clusters (Manhattan distance = 0.701). Relevant features are in Table 12 for the differentiation among clusters, assessed by the p -values and effect size. Famiglini et al. used a supervised classification approach for predicting ICU admission in a cohort of 1004 COVID-19 patients, with only 18.3% admitted to the ICU (imbalanced data) (Famiglini et al., 2022). Data curation (imputation and bias evaluation) and model selection resulted in better AUC score (classification=0.85), lower Brier score (calibration=0.144), and standardized net benefit (clinical utility=0.69), predicting ICU admission with significant importance of the NLR levels (consistent with the literature). Luo et al. also studied this outcome (mild, $n=67$ and severe, $n=129$) with the application of a hybrid system built on multi-criteria decision-making (MCDM) through the combination of a technique for order of preference by similarity to ideal solution (TOPSIS) algorithm and a naïve-Bayes classifier. TOPSIS runs preprocessing and feature ranking while NB performs feature selection. Despite this method achieving a higher AUC (0.93), the sample size was small and did not include external.

Murri et al. developed models for the prognosis of COVID-19 by developing an interpretable logistic regression model constructed with data from 921 hospitalized patients, of which 120 died (prevalence of 13%) (Murri et al., 2021). Despite the discriminatory ability assessed through the levels of hemoglobin, platelets, neutrophils, urea, c-reactive protein, and sodium was higher (AUC=0.87), the subsequent external validation on a population with a prevalence of 22.6%, decreased (AUC=0.81). Fernandes et al. extended the discriminative ability for fatality, invasive mechanical ventilation, and ICU (multipurpose algorithms) (Fernandes et al., 2021). Considering fewer features (age, lymphocyte-to-c-reactive-protein ratio, c-reactive protein, and results from the Braden scale), the authors concluded that each of the studied outcomes (ICU, IMV or fatality) could be predicted using data from the others (outcomes), always with an AUC > 91%. In the study of Karthikeyan et al., higher predictive performance was accomplished by applying an XGBoost for feature importance and a neural network for feature selection on a dataset comprised of deceased ($n=170$) and recovered ($n=200$) patients. Selected features predicted the number of days until the outcome, and accuracy results were consistently higher than 90% for models trained until 12 days before the outcome (with data not only from the closest days—case 2). Notably, the authors also showed blood patterns related to mortality prediction, such as high values of hs-CRP, LDH, and neutrophils and low values of eosinophils, consistent with previous literature.

Challenges

Despite the rising developments in AI reinforced by big data, computational power, and neural networks enhanced the quality of studies relating routine blood analysis with principal diagnosis and prognosis outcomes, the clinical deployment stage remains a foremost challenge. The studies and the pathologies we reviewed confirm the delay in implementing AI-based technology in the clinical setting. The research highlighted in this review was

motivated by the available statistical information expressing significant associations between blood metabolites and numerous pathologies and by the opportunity provided by the high number of general health panels typically performed in a medical health center. As referred, these analytical panels include complete cell blood count, metabolic and lipidic that are currently consistently evaluated with gold-standard, highly stabilized techniques, not prone to systematic errors or bias. A vast amount of non-appraised clinical information cannot be 100% perceived by a single clinician acting in a consultation or emergency setting (especially in longitudinal profiles) that can be processed, patterned, statistically evaluated, and flagged, if necessary. Since current clinical decisions are accomplished in a framework of rule-based systems, i.e., thresholds passively updated according to newer guidelines, the primary reasons that explain the resistance to ML-based solutions are the necessity to use external applications (which require manual data input and consumed extra time), and the non-interpretability of ML algorithms, especially the ones concerned to deep learning ('black-box'). Indeed, a recent study by Henry et al. evaluated the adoption of a AI-based sepsis targeted real-time early warning system (TREWS); these authors found a lack of interpretability of the computation model, but this was not considered a significant barrier, especially after experiencing the system through different patients and following interactions with peers and research team members (Henry et al., 2022). On the other hand, the theoretical 'competing diagnosis' may be perceived as threat to autonomy by some physicians, making them hesitant in adopting these solutions because it may alter their decision-making process with the risk of acting solely on model recommendations, which may not be completely accurate.

Regarding routine blood analysis, data sources (i.e., equipment, disease incidence, patient demographics) with different reference values should also be evaluated and discussed. The study should distinctly report information regarding the data source type (cohort, randomized control trial, or other), data source quality (representativeness, bias, features, and outcome with the exact time of measurement and associated medication or treatment), and data source quantity. People's biochemical fingerprint varies in basal conditions for several reasons; most have little to do with their clinical condition. Re-test studies could ultimately elucidate if the AI model's predictions connect to features that correlate with the problem of interest or if they only capture external variabilities, such as sensor noise, ambient temperatures, user manipulation, etc. (Stegmann et al., 2020)

Considering the evaluation of the reviewed medical applications, only a few were performed in external centers, and most were conducted with retrospective data. Therefore, working with data matching the same conditions met in traditional clinical settings is essential, principally user interface (i.e., healthcare professionals or patients) and technology integration into the clinical workflow (physical conditions such as illumination, temperature, humidity, and others). Curiously, a recent evaluation in a prospective assessment regarding the performance of a deep-learning system for the detection of diabetic retinopathy demonstrated a 'larger-than-expected proportional of the retinal images as ungradable owing to blurring or darkening', caused by poor ambient lighting during the measurement procedure (Co-operation, 2021). Regarding

COVID-19 prediction models, one study found that the underlying data distribution, known as domain shifts, significantly impacts anticipated performance and dependability, resulting in model failure in clinical applications. Domain shifts, which can be induced by changes in disease prevalence, adjustments to RT-PCR testing protocols, or viral mutations, suggest that machine learning models may lose reliability and performance over time, underlining the importance of constant monitoring and updating (Roland et al., 2022). These examples emphasize how training should incorporate the original conditions to generate truthful coefficients for the desired problem-solving.

These challenges should be revised and improved by consistently implementing the described ML pipeline to develop federated learning (training in multiple institutions) and the deployment in ETL (i.e., extract, transform and load), keeping data 'healthy'.

Future perspectives

The deployment of a cost-free real-time blood augmentation diagnostic tool, based on longitudinal data and source-stable (gold-standard), should address probabilistic metrics of diagnosis and provide the clinician with a landscape view for each individual. AI can play a key role in delivering explainable decision support systems to assure that patterns are correctly identified, and biomarkers are accurately measured, directly influencing the outcome. Measures of clinical effectiveness, such as user feedback, clinical reliance, and interpretability, must improve and be better described, particularly in the upcoming guidelines for model development and reporting (TRIPOD-ML). Although this protocol is still under development, The Transparent Reporting of a multivariate prediction model for Individual Prognosis or Diagnosis (TRIPOD, 2015) standard (Collins et al., 2015) should be considered since It provides guidance and recommendations for reporting a multivariate prediction model for diagnosis or prognosis. Representativeness, in particular, should always be addressed because it is an essential concept in data quality, covering the necessary heterogeneity of the studied population in a balanced proportion, which is especially important when models aim to predict categorical or binary events in the context of medical problems.

In the future, deploying ML models will still face data shifts across time, hindering representativeness and compromising the model's performance. A paramount example refers to the performance of the covid-19 models trained during the initial alpha strain with the current disease condition, led by several other variants of the virus, and the influence of the addition of vaccines that altered disease outcomes for the vast majority of infected individuals. Therefore, deployment should start with isolated pilot studies to receive feedback from healthcare experts on user experience, interface, efficiency, and real-time evaluation performance.

Outlook

This review summarizes the application of artificial intelligence algorithms in the diagnosis and prognosis of ICD-

10 disorders using routine blood tests only. Reports herein analyzed differ in data source type, quality, and quantity and describe a multitude of ML algorithms for outcome prediction. Principal findings indicate virtuous performance metrics in validation studies and a clear gap between standard disease-associated metabolites and those chosen machine learning models, resulting in higher performance metrics than traditional clinical practice scores.

Although there is still a sizable gap between reviewed studies and their clinical application, AI is changing the practice of medicine, and digital tools are key for helping physicians evaluate patients more personalized, rapid, and efficiently. The use of routine blood parameters as exclusive input features for model development could allow the translation of high-level diagnosis from primary or secondary care to point-of-care, making these analyses more valuable in lowering time to diagnosis and overall healthcare costs.

Author contributions

MS-S: Data curation, Investigation, Methodology, Writing—original draft. NS: Conceptualization, Formal Analysis, Funding acquisition, Supervision, Writing—review and editing. JS: Conceptualization, Data curation, Formal Analysis, Methodology, Supervision, Validation, Writing—review and editing.

References

- Abayomi-Alli, O. O., Damaševičius, R., Maskeliūnas, R., and Misra, S. (2022). An ensemble learning model for COVID-19 detection from blood test samples. *Sensors* 22, 2224. doi:10.3390/s22062224
- Ahmad, Z., Rahim, S., Zubair, M., and Abdul-Ghaffar, J. (2021). Artificial intelligence (AI) in medicine, current applications and future role with special emphasis on its potential and promise in pathology: present and future impact, obstacles including costs and acceptance among pathologists, practical and philosoph. *Diagn. Pathol.* 16, 1–16. doi:10.1186/s13000-021-01085-4
- Alsuliman, T., Humaidan, D., and Sliman, L. (2020). Machine learning and artificial intelligence in the service of medicine: necessity or potentiality? *Curr. Res. Transl. Med.* 68, 245–251. doi:10.1016/j.retram.2020.01.002
- Alves, M. A., Castro, G. Z., Oliveira, B. A. S., Ferreira, L. A., Ramirez, J. A., Silva, R., et al. (2021). Explaining machine learning based diagnosis of COVID-19 from routine blood tests with decision trees and criteria graphs. *Comput. Biol. Med.* 132, 104335. doi:10.1016/j.compbiomed.2021.104335
- Auld, S. C., Harrington, K. R. V., Adelman, M. W., Robichaux, C. J., Overton, E. C., Caridi-Scheible, M., et al. (2021). Trends in ICU mortality from coronavirus disease 2019: a tale of three surges. *Crit. Care Med.* 50, 245–255. doi:10.1097/ccm.0000000000005185
- Azarkhish, I., Raoufy, M. R., and Gharibzadeh, S. (2012). Artificial intelligence models for predicting iron deficiency anemia and iron serum level based on accessible laboratory data. *J. Med. Syst.* 36, 2057–2061. doi:10.1007/s10916-011-9668-3
- Babaei, S., Sorayaie, A., Ghafari, A., and Bagherzadeh, J. (2022). COVID-19 diagnosis from routine blood tests using artificial intelligence techniques. *Biomed. Signal Process Control.* doi:10.1016/j.bspc.2021.103263
- Badrick, T. (2013). Evidence-based laboratory medicine. *Clin. Biochem. Rev.* 34, 43–46.
- Bajwa, J., Munir, U., Nori, A., and Williams, B. (2021). Artificial intelligence in healthcare: transforming the practice of medicine. *Future Healthc. J.* 8, e188–e194. doi:10.7861/fhj.2021-0095
- Banerjee, A., Ray, S., Vorselaars, B., Kitson, J., Mamalakis, M., Weeks, S., et al. (2020). Use of machine learning and artificial intelligence to predict SARS-CoV-2 infection from full blood counts in a population. *Int. Immunopharmacol.* 86, 106705. doi:10.1016/j.intimp.2020.106705
- Barnhart-Magen, G., Gotlib, V., Marilus, R., and Einav, Y. (2013). Differential diagnostics of thalassemia minor by artificial neural networks model. *J. Clin. Lab. Anal.* 27, 481–486. doi:10.1002/jcla.21631
- Benito-León, J., del Castillo, M. D., Estirado, A., Ghosh, R., Dubey, S., and Serrano, J. I. (2021). Using unsupervised machine learning to identify age- and sex-independent severity subgroups among patients with COVID-19: observational longitudinal study. *J. Med. Internet Res.* 23, 259888–e26014. doi:10.2196/25988
- Benjamins, S., Dhunoo, P., and Meskó, B. (2020). The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *npj Digit. Med.* 3, 118. doi:10.1038/s41746-020-00324-0
- Bernardini, M., Morettini, M., Romeo, L., Frontoni, E., and Burattini, L. (2019). TyG-er: an ensemble Regression Forest approach for identification of clinical factors related to insulin resistance condition using Electronic Health Records. *Comput. Biol. Med.* 112, 103358. doi:10.1016/j.compbiomed.2019.103358
- Böger, B., Fachi, M. M., Vilhena, R. O., Cobre, A. F., Tonin, F. S., and Pontarolo, R. (2021). Systematic review with meta-analysis of the accuracy of diagnostic tests for COVID-19. *Am. J. Infect. Control* 49, 21–29. doi:10.1016/j.ajic.2020.07.011
- Brendan McMahan, H., Moore, E., Ramage, D., and Hampson, S., Communication-efficient learning of deep networks from decentralized data. 54, 10 (2017).
- Brinati, D., Campagner, A., Ferrari, D., Locatelli, M., Banfi, G., and Cabitza, F. (2020). Detection of COVID-19 infection from routine blood exams with machine learning: a feasibility study. *J. Med. Syst.* 44, 135. doi:10.1007/s10916-020-01597-4
- Bruckert, S., Finzel, B., and Schmid, U. (2020). The next generation of medical decision support: a roadmap toward transparent expert companions. *Front. Artif. Intell.* 3, 507973–508013. doi:10.3389/frai.2020.507973
- Cabitza, F., Campagner, A., Soares, F., García de Guadiana-Romualdo, L., Challa, F., Sulejmani, A., et al. (2021). The importance of being external. methodological insights for the external validation of machine learning models in medicine. *Comput. Methods Programs Biomed.* 208, 106288. doi:10.1016/j.cmpb.2021.106288
- Campagner, A., Carobene, A., and Cabitza, F. (2021). External validation of machine learning models for COVID-19 detection based on complete blood count. *Health Inf. Sci. Syst.* 9, 37–15. doi:10.1007/s13755-021-00167-3
- Cao, Y., Hu, Z., De, Liu, X. F., Deng, A. M., and Hu, C. J. (2013). An MLP classifier for prediction of HBV-induced liver cirrhosis using routinely available clinical parameters. *Dis. Markers* 35, 653–660. doi:10.1155/2013/127962
- Celkan, T. T. (2020). What does a hemogram say to us? *Turk pediatri arsivi* 55, 103–116. doi:10.14744/TurkPediatriArs.2019.76301
- Chadaga, K., Prabhu, S., Bhat, V., Sampathila, N., Umakanth, S., and Chadaga, R. (2023). A decision support system for diagnosis of COVID-19 from non-COVID-

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. MS-S received a PhD fellowship from the Foundation for Science and Technology (FCT, Portugal)/FEDER.

Conflict of interest

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

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

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

- 19 influenza-like illness using explainable artificial intelligence. *Bioengineering* 10, 439. doi:10.3390/bioengineering10040439
- Chadaga, K., Prabhu, S., Vivekananda Bhat, K., Umakanth, S., and Sampathila, N. (2022). Medical diagnosis of COVID-19 using blood tests and machine learning. *J. Phys. Conf. Ser.* 2161, 012017. doi:10.1088/1742-6596/2161/1/012017
- Chatburn, R., and Hematology, E. M.-C. (2010). *Handbook of respiratory care*. Third Edition, 54–63.
- Çil, B., Ayyıldız, H., and Tuncer, T. (2020). Discrimination of β -thalassemia and iron deficiency anemia through extreme learning machine and regularized extreme learning machine based decision support system. *Med. Hypotheses* 138, 109611. doi:10.1016/j.mehy.2020.109611
- Collins, G. S., Reitsma, J. B., Altman, D. G., and Moons, K. G. M. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Eur. Urol.* 67, 1142–1151. doi:10.1016/j.eururo.2014.11.025
- Co-operation, E. (2021). Machine learning in translation. *Nat. Biomed. Eng.* 5, 485–486. doi:10.1038/s41551-021-00758-1
- Data, O. W. (2023). *Covid-19 deaths*.
- Dayan, I., Roth, H. R., Zhong, A., Harouni, A., Gentili, A., Abidin, A. Z., et al. (2021). Federated learning for predicting clinical outcomes in patients with COVID-19. *Nat. Med.* 27, 1735–1743. doi:10.1038/s41591-021-01506-3
- Demichev, V., Tober-Lau, P., Lemke, O., Nazarenko, T., Thibeault, C., Whitwell, H., et al. (2021). A time-resolved proteomic and prognostic map of COVID-19. *Cell Syst.* 12, 780–794.e7. doi:10.1016/j.cels.2021.05.005
- Diagnostics for All (2023). *Foundation for innovative new diagnostics*. Test Directory.
- Dipnall, J. F., Pasco, J. A., Berk, M., Williams, L. J., Dodd, S., Jacka, F. N., et al. (2016). Fusing data mining, machine learning and traditional statistics to detect biomarkers associated with depression. *PLoS One* 11, 01481955–e148223. doi:10.1371/journal.pone.0148195
- Famigliani, L., Campagner, A., Carobene, A., and Cabitza, F. (2022). A robust and parsimonious machine learning method to predict ICU admission of COVID-19 patients. *Med. Biol. Eng. Comput.*, 1–13. doi:10.1007/s11517-022-02543-x
- Fernandes, F. T., de Oliveira, T. A., Teixeira, C. E., Batista, A. F. M., Dalla Costa, G., and Chiavegatto Filho, A. D. P. (2021). A multipurpose machine learning approach to predict COVID-19 negative prognosis in São Paulo, Brazil. *Sci. Rep.* 11, 3343–3347. doi:10.1038/s41598-021-82885-y
- Fialoke, S., Malarstig, A., Miller, M. R., and Dumitriu, A. (2018). Application of machine learning methods to predict non-alcoholic steatohepatitis (NASH) in non-alcoholic fatty liver (NAFL) patients. *AMIA Annu. Symp. Proc. AMIA Symp.* 2018, 430–439.
- Food and Drug Administration. Emergency use authorizations for medical devices (2021). *Template for developers of molecular and antigen diagnostic COVID-19 tests for home use*.
- Fujiwara, K. (2021). Sparse Modeling delivers fast, energy efficient and explainable AI solutions for cutting-edge medical applications. *Nature*, 50–51.
- Gunčar, G., Kukar, M., Notar, M., Brvar, M., Černelč, P., Notar, M., et al. (2018). An application of machine learning to haematological diagnosis. *Sci. Rep.* 8, 411. doi:10.1038/s41598-017-18564-8
- Guo, K., Fu, X., Zhang, H., Wang, M., Hong, S., and Ma, S. (2021). Predicting the postoperative blood coagulation state of children with congenital heart disease by machine learning based on real-world data. *Transl. Pediatr.* 10, 33–43. doi:10.21037/tp-20-238
- Haider, R. Z., Ujjan, I. U., Khan, N. A., Urrechaga, E., and Shamsi, T. S. (2022). Beyond the in-practice CBC: the research CBC parameters-driven machine learning predictive modeling for early differentiation among leukemias. *Diagnostics* 12, 138. doi:10.3390/diagnostics12010138
- Hao, K. (2023). *Training a single AI model can emit as much carbon as five cars in their lifetimes*. United States: MIT Technology Review.
- Henry, K. E., Kornfield, R., Sridharan, A., Linton, R. C., Groh, C., Wang, T., et al. (2022). Human – machine teaming is key to AI adoption: clinicians’ experiences with a deployed machine learning system. *NPJ Digit. Med.* 5, 97–106. doi:10.1038/s41746-022-00597-7
- Ho, T. S., Weng, T. C., Wang, J. D., Han, H. C., Cheng, H. C., Yang, C. C., et al. (2020). Comparing machine learning with case-control models to identify confirmed dengue cases. *PLoS Negl. Trop. Dis.* 14, 00088433–e8921. doi:10.1371/journal.pntd.0008843
- Hochman, E., Feldman, B., Weizman, A., Krivoy, A., Gur, S., Barzilay, E., et al. (2021). Development and validation of a machine learning-based postpartum depression prediction model: a nationwide cohort study. *Depress. Anxiety* 38, 400–411. doi:10.1002/da.23123
- Hornbrook, M. C., Goshen, R., Choman, E., O’Keeffe-Rosetti, M., Kinar, Y., Liles, E. G., et al. (2017). Early colorectal cancer detected by machine learning model using gender, age, and complete blood count data. *Dig. Dis. Sci.* 62, 2719–2727. doi:10.1007/s10620-017-4722-8
- Joshi, R. P., Pejaver, V., Hammarlund, N. E., Sung, H., Lee, S. K., Furmanchuk, A., et al. (2020). A predictive tool for identification of SARS-CoV-2 PCR-negative emergency department patients using routine test results. *J. Clin. Virol.* 129, 104502. doi:10.1016/j.jcv.2020.104502
- Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Nitin Bhagoji, A., et al. (2021). Advances and open problems in federated learning. *Found. Trends® Mach. Learn.* 14, 1–210. doi:10.1561/22000000083
- Karthikeyan, A., Garg, A., Vinod, P. K., and Priyakumar, U. D. (2021). Machine learning based clinical decision support system for early COVID-19 mortality prediction. *Front. Public Health* 9, 626697–626713. doi:10.3389/fpubh.2021.626697
- Kerr, W. T., Lau, E. P., Owens, G. E., and Treffer, A. (2012). The future of medical diagnostics: large digitized databases. *Yale J. Biol. Med.* 85, 363–377.
- Kinar, Y., Kalkstein, N., Akiva, P., Levin, B., Half, E. E., Goldshtein, I., et al. (2016). Development and validation of a predictive model for detection of colorectal cancer in primary care by analysis of complete blood counts: a binational retrospective study. *J. Am. Med. Inf. Assoc.* 23, 879–890. doi:10.1093/jamia/ocv195
- Kline, R. R. (2011). Cybernetics, automata studies, and the dartmouth conference on artificial intelligence. *IEEE Ann. Hist. Comput.* 33, 5–16. doi:10.1109/mahc.2010.44
- Kocbek, P., Fijacko, N., Soguero-Ruiz, C., Mikalsen, K. Ø., Maver, U., Povalej Brzan, P., et al. (2019). Maximizing interpretability and cost-effectiveness of surgical site infection (SSI) predictive models using feature-specific regularized logistic regression on preoperative temporal data. *Comput. Math. Methods Med.* 2019, 1–13. doi:10.1155/2019/2059851
- Kopitar, L., Kocbek, P., Cilar, L., Sheikh, A., and Stiglic, G. (2020). Early detection of type 2 diabetes mellitus using machine learning-based prediction models. *Sci. Rep.* 10, 11981–12012. doi:10.1038/s41598-020-68771-z
- Krizhevsky, A., Sutskever, I., and Hinton, G. (2012). ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* 25, 1097–1105. doi:10.1145/3383972.3383975
- Kushner, T., Brette, M. D., and Sankaranarayanan, S. (2020). Multi-hour blood glucose prediction in type 1 diabetes: a patient-specific approach using shallow neural network models. *Diabetes Technol. Ther.* 22, 883–891. doi:10.1089/dia.2020.0061
- Li, W. T., Ma, J., Shende, N., Castaneda, G., Chakladar, J., Tsai, J. C., et al. (2020). Using machine learning of clinical data to diagnose COVID-19: a systematic review and meta-analysis. *BMC Med. Inf. Decis. Mak.* 20, 247–313. doi:10.1186/s12911-020-01266-z
- Lin, J. K., Chien, T. W., Wang, L. Y., and Chou, W. (2021). An artificial neural network model to predict the mortality of COVID-19 patients using routine blood samples at the time of hospital admission: development and validation study. *Med. Baltim.* 100, e26532. doi:10.1097/md.00000000000026532
- Luo, J., Zhou, L., Feng, Y., Li, B., and Guo, S. (2021). The selection of indicators from initial blood routine test results to improve the accuracy of early prediction of COVID-19 severity. *PLoS One* 16, 02533299–e253418. doi:10.1371/journal.pone.0253329
- Ma, H., Xu, C. F., Shen, Z., Yu, C. H., and Li, Y. M. (2018). Application of machine learning techniques for clinical predictive modeling: a cross-sectional study on nonalcoholic fatty liver disease in China. *BioMed Res. Int.* 2018, 1–9. doi:10.1155/2018/4304376
- Mahmood, N., Shahid, S., Bakhshi, T., Riaz, S., Ghufuran, H., and Yaqoob, M. (2020). Identification of significant risks in pediatric acute lymphoblastic leukemia (ALL) through machine learning (ML) approach. *Med. Biol. Eng. Comput.* 58, 2631–2640. doi:10.1007/s11517-020-02245-2
- Mamoshina, P., Kochetov, K., Cortese, F., Kovalchuk, A., Aliper, A., Putin, E., et al. (2019). Blood biochemistry analysis to detect smoking status and quantify accelerated aging in smokers. *Sci. Rep.* 9, 142–210. doi:10.1038/s41598-018-35704-w
- Marieb, E. N., and Hoehn, K. (2012). Blood composition and functions. *Hum. Anat. Physiol.*, 634–657.
- Mathioudakis, N. N., Abusamaan, M. S., Shakarchi, A. F., Sokolinsky, S., Fayzullin, S., McGready, J., et al. (2021). Development and validation of a machine learning model to predict near-term risk of iatrogenic hypoglycemia in hospitalized patients. *JAMA Netw. Open* 4, e2030913–e2030915. doi:10.1001/jamanetworkopen.2020.30913
- Matthew, R., and Pincus, N. Z. A. J. (2011). *Henry’s clinical diagnosis and management*. 22th Edition. Amsterdam, Netherlands: Elsevier.
- Meiseles, A., Paley, D., Ziv, M., Hadid, Y., Rokach, L., and Tadmor, T. (2022). Explainable machine learning for chronic lymphocytic leukemia treatment prediction using only inexpensive tests. *Comput. Biol. Med.* 145, 105490. doi:10.1016/j.combiomed.2022.105490
- Metsker, O., Magoev, K., Yakovlev, A., Yanishevskiy, S., Kopanitsa, G., Kovalchuk, S., et al. (2020). Identification of risk factors for patients with diabetes: diabetic polyneuropathy case study. *BMC Med. Inf. Decis. Mak.* 20, 201–215. doi:10.1186/s12911-020-01215-w
- Mooney, C., Eogan, M., Ni Áinle, F., Cleary, B., Gallagher, J. J., O’Loughlin, J., et al. (2021). Predicting bacteraemia in maternity patients using full blood count parameters: a supervised machine learning algorithm approach. *Int. J. Laboratory Hematol.* 43, 609–615. doi:10.1111/ijlh.13434
- Moranga, C. M., Amenga-Etego, L., Bah, S. Y., Appiah, V., Amuzu, D. S. Y., Amoako, N., et al. (2020). Machine learning approaches classify clinical malaria outcomes based on haematological parameters. *BMC Med.* 18, 375–416. doi:10.1186/s12916-020-01823-3

- Murri, R., Lenkiewicz, J., Masciocchi, C., Iacomini, C., Fantoni, M., Damiani, A., et al. (2021). A machine-learning parsimonious multivariable predictive model of mortality risk in patients with Covid-19. *Sci. Rep.* 11, 21136–21210. doi:10.1038/s41598-021-99905-6
- Our World in data (2023). Covid-19 cases. Available at: <https://ourworldindata.org/grapher/covid-cases-income>.
- Predregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825.
- Peng, J., Zhou, M., Chen, C., Xie, X., and Luo, C. H. (2020). Identification of exacerbation risk in patients with liver dysfunction using machine learning algorithms. *PLoS One* 15, 02392666–e239310. doi:10.1371/journal.pone.0239266
- Plante, T. B., Blau, A. M., Berg, A. N., Weinberg, A. S., Jun, I. C., Tapson, V. F., et al. (2020). Development and external validation of a machine learning tool to rule out COVID-19 among adults in the emergency department using routine blood tests: a large, multicenter, real-world study. *J. Med. Internet Res.* 22, 240488–e24112. doi:10.2196/24048
- Prince-Guerra, J. L., Almdares, O., Nolen, L. D., Gunn, J. K. L., Dale, A. P., Buono, S. A., et al. (2021). Evaluation of abbot BinaxNOW rapid antigen test for SARS-CoV-2 infection at two community-based testing sites — pima county, Arizona, november 3–17, 2020. *MMWR. Morb. Mortal. Wkly. Rep.* 70, 100–105. doi:10.15585/mmwr.mm7003e3
- Rashed-Al-Mahfuz, M., Haque, A., Azad, A., Alyami, S. A., Quinn, J. M. W., and Moni, M. A. (2021). Clinically applicable machine learning approaches to identify attributes of chronic kidney disease (CKD) for use in low-cost diagnostic screening. *IEEE J. Transl. Eng. Health Med.* 9, 1–11. doi:10.1109/jtehm.2021.3073629
- Ratzinger, F., Dedeyan, M., Rammerstorfer, M., Perkmann, T., Burgmann, H., Makristathis, A., et al. (2014). A risk prediction model for screening bacteremic patients: a cross sectional study. *PLoS One* 9, e106765. doi:10.1371/journal.pone.0106765
- Rawson, T. M., Hernandez, B., Moore, L. S. P., Blandly, O., Herrero, P., Gilchrist, M., et al. (2019). Supervised machine learning for the prediction of infection on admission to hospital: a prospective observational cohort study. *J. Antimicrob. Chemother.* 74, 1108–1115. doi:10.1093/jac/dky514
- Reardon, S. (2019). Rise of robot radiologists. *Nature* 576, S54–S58. doi:10.1038/d41586-019-03847-z
- Richard, A., McPherson, M., and Pincus, M. R. (2011). Clinical diagnosis and management by laboratory methods. doi:10.1136/jcp.34.2.228-a
- Roland, T., Böck, C., Tschollitsch, T., Maletzky, A., Hochreiter, S., Meier, J., et al. (2022). Domain shifts in machine learning based covid-19 diagnosis from blood tests. *J. Med. Syst.* 46, 23. doi:10.1007/s10916-022-01807-1
- Sarbaz, M., Pournik, O., Ghalichi, L., Kimiafar, K., and Razavi, A. R. (2013). Designing a Human T-Lymphotropic Virus Type 1 (HTLV-I) diagnostic model using the complete blood count. *Iran. J. Basic Med. Sci.* 16, 247–251.
- Sarker, I. H. (2021). Machine learning: algorithms, real-world applications and research directions. *SN Comput. Sci.* 2, 160. doi:10.1007/s42979-021-00592-x
- Sarker, I. H. (2022). AI-based modeling: techniques, applications and research issues towards automation, intelligent and smart systems. *SN Comput. Sci.* 3, 158–220. doi:10.1007/s42979-022-01043-x
- Shou, L., Huang, W. W., Barszczyk, A., Wu, S. J., Han, H., Waese-Perlman, A., et al. (2021). Blood biomarkers predict cardiac workload using machine learning. *BioMed Res. Int.* 2021, 1–5. doi:10.1155/2021/6172815
- Shukla Shubhendu, S., and Vijay, (2013). J. Applicability of artificial intelligence in different fields of life. *Int. J. Sci. Eng. Res.* 1, 28–35.
- Soerensen, P. D., Christensen, H., Gray Worsoe Laursen, S., Hardahl, C., Brandslund, I., and Madsen, J. S. (2022). Using artificial intelligence in a primary care setting to identify patients at risk for cancer: a risk prediction model based on routine laboratory tests. *Clin. Chem. Laboratory Med. (CCLM)* 60, 2005–2016. doi:10.1515/cclm-2021-1015
- Soguero-Ruiz, C., Fei, W. M. E., Jenssen, R., Augestad, K. M., Álvarez, J. L. R., Jiménez, I. M., et al. (2015). Data-driven temporal prediction of surgical site infection. *AMIA Annu. Symp. Proc. AMIA Symp.* 2015, 1164–1173.
- Souza, A. A., Almeida, D. C., Barcelos, T. S., Bortoletto, R. C., Munoz, R., Waldman, H., et al. (2021). Simple hemogram to support the decision-making of COVID-19 diagnosis using clusters analysis with self-organizing maps neural network. *Soft Comput.* 27, 3295–3306. doi:10.1007/s00500-021-05810-5
- Stegmann, G. M., Hahn, S., Liss, J., Shefner, J., Rutkove, S., Kawabata, K., et al. (2020). Repeatability of commonly used speech and language features for clinical applications. *Digit. Biomarkers* 4, 109–122. doi:10.1159/000511671
- Svensson, C. M., Hübler, R., and Figge, M. T. (2015). Automated classification of circulating tumor cells and the impact of interobserver variability on classifier training and performance. *J. Immunol. Res.* 2015, 1–9. doi:10.1155/2015/573165
- Tarwater, K. (2011). Estimated glomerular filtration rate explained. *Mo. Med.* 108, 29–32.
- The Medical Futurist (2022). FDA approved AI-based algorithms. Available at: <https://medicalfuturist.com/fda-approved-ai-based-algorithms/>.
- Tian, J., Yu, S., Wang, L., Kong, D., Zhao, W., Tian, Z., et al. (2022). Molecular pathogenesis: connections between viral hepatitis-induced and non-alcoholic steatohepatitis-induced hepatocellular carcinoma. *Nat. Prod. Res.* 13, 1–9. doi:10.1080/14786419.2022.2134864
- Tschollitsch, T., Dünser, M., Böck, C., Schwarzbauer, K., and Meier, J. (2021). Machine learning prediction of SARS-CoV-2 polymerase chain reaction results with routine blood tests. *Lab. Med.* 52, 146–149. doi:10.1093/labmed/lmaa111
- World Health Organization (2021). Antigen-detection in the diagnosis of SARS-CoV-2 infection. *Interim guid.*
- Wu, J., Shen, J., Xu, M., and Shao, M. (2021). A novel combined dynamic ensemble selection model for imbalanced data to detect COVID-19 from complete blood count. *Comput. Methods Programs Biomed.* 211, 106444. doi:10.1016/j.cmpb.2021.106444
- Wu, Z., and McGoogan, J. M. (2020). Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72314 cases from the Chinese center for disease control and prevention. *JAMA - J. Am. Med. Assoc.* 323, 1239–1242. doi:10.1001/jama.2020.2648
- Yao, Z., Li, J., Guan, Z., Ye, Y., and Chen, Y. (2020). Liver disease screening based on densely connected deep neural networks. *Neural Netw.* 123, 299–304. doi:10.1016/j.neunet.2019.11.005
- Yilmaz, Z., and Bozkurt, M. R. (2012). Determination of women iron deficiency anemia using neural networks. *J. Med. Syst.* 36, 2941–2945. doi:10.1007/s10916-011-9772-4
- Zhan, J., Chen, W., Cheng, L., Wang, Q., Han, F., and Cui, Y. (2020). Diagnosis of asthma based on routine blood biomarkers using machine learning. *Comput. Intell. Neurosci.* 2020, 1–8. doi:10.1155/2020/8841002
- Zheng, H., Zhu, J., Xie, W., and Zhong, (2021). J. Reinforcement learning assisted oxygen therapy for COVID-19 patients under intensive care. *BMC Med. Inf. Decis. Mak.* 21, 1–8. doi:10.1186/s12911-021-01712-6
- Zheng, Y., Guo, Z., Zhang, Y., Shang, J., Yu, L., Fu, P., et al. (2022). Rapid triage for ischemic stroke: a machine learning-driven approach in the context of predictive, preventive and personalised medicine. *EPMA J.* 13, 285–298. doi:10.1007/s13167-022-00283-4
- Zoabi, Y., Kehat, O., Lahav, D., Weiss-Meilik, A., Adler, A., and Shomron, N. (2021). Predicting bloodstream infection outcome using machine learning. *Sci. Rep.* 11, 20101–20111. doi:10.1038/s41598-021-99105-2