Universidade do Minho
Escola de Medicina

Miguel Ângelo Alves Gomes dos Santos Silva

Application of machine learning for hematological diagnosis

outubro de 2023

Application of machine learning for hematological diagnosis
Miguel Ângelo Alves Gomes dos Santos Silva
UMinho | 2023

Universidade do Minho
Escola de Medicina

Miguel Ângelo Alves Gomes dos Santos Silva

Application of machine learning for hematological diagnosis

Tese de Doutoramento
Doutoramento em Ciências da Saúde Aplicadas

Trabalho efetuado sob a orientação do
Professor Doutor João Carlos Sousa
e do
Professor Doutor Nuno Sousa

outubro de 2023

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS

Este é um trabalho académico que pode ser utilizado por terceiros desde que respeitadas as regras e boas práticas internacionalmente aceites, no que concerne aos direitos de autor e direitos conexos.

Assim, o presente trabalho pode ser utilizado nos termos previstos na licença abaixo indicada.

Caso o utilizador necessite de permissão para poder fazer um uso do trabalho em condições não previstas no licenciamento indicado, deverá contactar o autor, através do RepositóriUM da Universidade do Minho.

Licença concedida aos utilizadores deste trabalho



Atribuição-CompartilhaIgual
CC BY-SA

https://creativecommons.org/licenses/by-sa/4.0/

ACKNOWLEDGEMENTS

To Professor João Sousa, for leading me on this path and decisively contribute to my scientific development. Also, for all the support, mentorship, and unique humanism.

To Professor Nuno Sousa, for your trust, enthusiasm, and objective vision that allowed me to conduct translational biomedical research. Thank you for believing in me.

To Doctor Marina Majar, for all the clinical, technical, and operational contributions in the scope of clinical pathology. For your constant hospitality, concern, and kindness.

To Professors César Ferreira and António Barros, for sharing their expertise and assisting me in different stages of my technical development.

To all the members of the Clinic Pathology Department and the Clinic Academic Center of the Hospital de Braga, especially Mónica, Ana, Daniela, Joana, Doctor Aurélio, Susana, and Miguel. Also, to Fernanda and Goreti for their operational assistance.

To the School of Medicine for creating the PhDiHES program.

To all my friends for your support.

À minha família, por todo o amor, apoio e motivação que me fizeram chegar até aqui.

À Marisa, presente desde o primeiro dia, por estares sempre comigo.

The work presented in this thesis was performed in the Life and Health Sciences Research Institute (ICVS), University of Minho, Portugal. Financial support was provided by grants from the Portuguese Foundation of Science and Technology - Doctoral Fellowship to Miguel Ângelo Alves Gomes dos Santos Silva (PD/BDE/150622/2020), by the project NORTE-01-0145-FEDER-000039, supported by Norte Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF) and by National funds, through the Foundation for Science and Technology (FCT) - project UIDB/50026/2020 and UIDP/50026/2020.

Cofinanciado por:

COMPETE 2020 PORTUGAL 2020 UNIÃO EUROPEIA Fundo Europeu de Desenvolvimento Regional

FCT Fundação para a Ciência e a Tecnologia MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

NORTE 2020 PROGRAMA OPERACIONAL REGIONAL DO NORTE

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

RESUMO

Aplicação de machine learning para diagnóstico hematológico

A análise de sangue impulsiona a tomada de decisão clínica em cerca de 70% dos atos médicos, influenciando o diagnóstico, prognóstico e tratamentos, segundo critérios demográficos predefinidos, nomeadamente sexo e idade. Os avanços técnico-científicos na ciência dos dados permitiram o estudo e desenvolvimento de sistemas inovadores de apoio à decisão, cujo objetivo fundamental utiliza algoritmos estatísticos e de inteligência artificial para gerar informações clínicas relevantes que promovam melhorias de desempenho no diagnóstico. Nesta tese, abordamos o diagnóstico hematológico incorporando diferentes estratégias de análise de dados para extrair informações clínicas importantes relacionadas com estados patológicos. ***No primeiro capítulo***, aplicamos estatística inferencial para caracterizar uma coorte de pacientes admitidos ao serviço de urgência, analisando Hb e PCR na prevalência de anemia e inflamação. Descobrimos que 36% dos pacientes hospitalizados apresentavam anemia e inflamação, e que níveis elevados de PCR explicavam 15% das alterações de Hb em toda a coorte. Também verificamos que a relação PCR/Hb contribuiu na distinção de alta, em vez de avaliar somente anemia ou inflamação. ***No segundo capítulo***, revimos os estudos de machine learning que utilizaram exames de sangue de rotina para a modelação de diagnóstico ou prognóstico de doenças. No geral, evidenciamos uma diferença significativa entre os parâmetros sanguíneos padrão associados a doenças com os selecionados pelos modelos de machine learning, com métricas de desempenho superiores quando comparados com modelos tradicionais, como APRI e FIB-4. As principais desvantagens foram a falta de validação clínica em centros externos e as dificuldades inerentes à integração operacional clínica. ***No terceiro capítulo***, aplicamos um algoritmo de reconhecimento de padrões para analisar tumores do trato digestivo com base apenas na concentração de Hb e nos níveis de RBC, Htc e RDW. A análise permitiu traçar o perfil de cada tumor, relacionando padrões semelhantes nos tumores que partilhavam a mesma localização anatómica de desenvolvimento. A abordagem descrita aumentou o valor clínico destes parâmetros, não apenas na definição do perfil do tumor, mas também na ajuda à discriminação entre os vários tumores no momento do diagnóstico. Nos três estudos, demonstramos a importância dos parâmetros sanguíneos de rotina em fornecer informações clínicas valiosas. A futura integração destas estratégias na modelação de análises de sangue pode contribuir para o diagnóstico diferenciado, no ponto de atendimento.

Palavras-chave: análise clínica de sangue, aprendizagem de máquina, diagnóstico reconhecimento de padrões

ABSTRACT

Application of machine learning for hematological diagnosis

Blood analysis guides clinical decision-making in nearly 70% of medical acts, influencing diagnosis, prognosis, and follow-up treatments under standardized criteria predefined by patient demographics, namely gender and age. Continuous technical and scientific advances in data science enabled the study and development of innovative decision support systems, whose fundamental aim addresses the use of statistical and artificial intelligence algorithms for delivering relevant insights and providing better performance metrics in the scope of diagnosis. In this thesis, we approached hematological diagnosis by incorporating different data mining strategies for extracting valuable clinical information related to pathological states. ***In the first chapter***, we used inferential statistics to characterize an acute cohort of patients admitted to the emergency department by analyzing Hb and CRP in the prevalence of anemia and inflammation. We found that 36% of hospitalized patients had both anemia and inflammation and that elevated CRP levels explained 15% of Hb changes across the entire cohort. We also found that the CRP/Hb ratio helped to discharge distinction rather than assessing anemia or inflammation only. ***In the second chapter***, we reviewed the machine learning studies that used routine blood tests to model disease diagnosis or prognosis. Overall, we found a significant difference between standard disease-associated blood parameters and those chosen by machine learning models, with improved performance metrics compared to commonly used scores such as APRI and FIB-4. The principal drawbacks were the lack of clinical validation in external centers and the physicians' struggles with operational integration. ***In the third chapter***, we used a pattern recognition algorithm to analyze digestive tract tumors based solely on Hb concentration and RBC, Htc, and RDW levels. Our analysis allowed us to profile each tumor with these blood parameters and demonstrate similar patterns in those with the same anatomical location of cancer development. While this finding agreed with the literature, the described approach enhanced the value of these blood tests for tumor profiling and discrimination at the time of diagnosis. In all three studies, we demonstrated the additional significance of routine blood parameters for delivering valuable clinical insights. Future integration of these strategies in routine blood modeling will contribute to differential diagnosis at the point of care.

Keywords: clinical blood analysis, diagnosis, machine learning, pattern recognition

TABLE OF CONTENTS

Direitos de autor e condições de utilização do trabalho por terceiros ii

Acknowledgements iii

Statement of integrity iv

Resumo v

Abstract vi

Table of contents vii

List of abbreviations viii

List of figures ix

List of tables x

1. Introduction 1

1.1 Background 2

1.2 Goal, Aims, and Approach 12

References 18

2. Experimental work 21

2.1 Correlation Analysis between Hemoglobin and C-Reactive Protein in Patients Admitted to an Emergency Unit 22

2.2 Artificial intelligence in routine blood tests: A review 39

2.3 Pattern recognition of hematological profiles of tumors of the digestive tract: an exploratory study 80

3. Discussion, Conclusion, and Future Perspectives 94

References 101

LIST OF ABBREVIATIONS

ACD – anemia of chronic disease

AI – artificial intelligence

ALP – alkaline phosphatase

ALT – alanine transaminase

ANN – artificial neural network

APRI – aspartate aminotransferase to platelet ratio index

AST – aspartate transferase

AUC – area under the curve

AUROC – area under the receiving operating characteristic curve

CBC – cell blood count

CDC – center for disease control and prevention

CRP – c-reactive protein

CVD – cardiovascular diseases

DL – deep learning

ED – emergency department

EHR – electronic health record

FDA – Food and Drug Administration

FIB-4 – fibrosis Index based on 4 factors

GGT – gamma-glutamyl transferase

Hb – hemoglobin

HDL – high-density lipoprotein

Htc – hematocrit

HIS – health information system

HIV – human immunodeficiency virus

ICD-10 – International Classification of Diseases 10th Revision

ICU – intensive care unit

IDA – iron deficiency anemia

INR – international normalized ratio

IQR – interquartile range

LDL – low-density lipoprotein

LIS – laboratory information system

MCH – mean corpuscular hemoglobin

MCHC – mean corpuscular hemoglobin concentration

MCV – mean corpuscular volume

miRNA – microRNA

ML – machine learning

PCA – principal components analysis

PC1 – first principal component

PC2 – second principal component

PLT – platelets

POC – point-of-care

PTH – parathyroid hormone

RBC – red blood cells

RBF – radial basis function

RDT – rapid diagnostic test

RDW – red cell distribution width

SVM – support vector machine

T1DM – type I diabetes mellitus

T2DM – type II diabetes mellitus

TRIPOD-ML – transparent reporting of a multivariate prediction model for individual prognosis or diagnosis

TSH – thyroid stimulating hormone

WBC – white blood cells

WHO – World Health Organization

LIST OF FIGURES

Chapter 1. Introduction

Chapter 1.1. Goal, Aims, and Approach

Figure 1. Main goal and the four specific aims of this dissertation.....14

Chapter 2. Experimental Work

Chapter 2.1 Correlation Analysis between Hemoglobin and C-Reactive Protein in Patients Admitted to an Emergency Unit

Figure 1. Correlation heat map between each blood parameter and the tumors of the digestive tract.....32

Figure 2. Boxplots show the distribution of hemoglobin and c-reactive protein according to diagnostics group (a) and type of discharge (b).....34

Figure 3. ROC curves of hemoglobin (a), C-reactive protein (b), and C-reactive protein-to-hemoglobin ratio (c).....35

Chapter 2.3 Pattern recognition of hematological profiles of tumors of the digestive tract

Figure 1. The correlation heat map of the laboratory blood tests with the studied disease groups.....87

Figure 2: Principal Components Analysis (PCA) plot discrimination of the studied digestive tumors, based on the four hematologic parameters.....88

Figure 3. ROC curves and performance metrics of the LR prediction models for the classification of each digestive tract tumor.....89

LIST OF TABLES

Chapter 2. Experimental Work

Chapter 2.1 Correlation Analysis between Hemoglobin and C-Reactive Protein in Patients Admitted to an Emergency Unit

Table 1. Descriptive frequencies of the studied cohort.....29

Table 2. Distribution of clinical conditions (anemia only, inflammation only, both anemia and inflammation, and neither of these) across each group of the cohort.....30

Table 3. Correlation between Hb and CRP in the studied subgroups, according to main characteristics and clinical conditions.....32

Table 4. ROC curve results for blood parameters in the cohort.....35

Chapter 2.2 Artificial intelligence in routine blood tests: A review

Table 1. A comprehensive list of supervised ML algorithms for classification, according to the desired learning strategy.....44

Table 2. General health panel required in routine blood examinations.....48

Table 3. Machine-learning-based routine blood tests for the diagnosis (or prognosis) of certain infections or parasitic diseases.....52

Table 4. Machine-learning-based routine blood tests for the diagnosis (or prognosis) of neoplasms.....54

Table 5. Machine-learning-based routine blood tests for the diagnosis (or prognosis) of diseases of the blood or blood-forming organs.....55

Table 6. Machine-learning-based routine blood tests for the diagnosis (or prognosis) of endocrine, nutritional, or metabolic diseases.....57

Table 7. Machine-learning-based routine blood tests for the diagnosis (or prognosis) of mental, behavioral, or neurodevelopmental disorders.....58

Table 8. Machine-learning-based routine blood tests for the diagnosis (or prognosis) of circulatory system diseases.....60

Table 9. Machine-learning-based routine blood tests for respiratory system disease diagnosis (or prognosis).....61

Table 10. Machine-learning-based routine blood tests for the diagnosis (or prognosis) of digestive system diseases.....63

Table 11. Machine-learning-based routine blood tests for the diagnosis (or prognosis) of diseases of the genitourinary system.....64
Table 12. Machine-learning-based routine blood tests for the diagnosis (or prognosis) of Covid-19.....69

Chapter 2.3 Pattern recognition of hematological profiles of tumors of the digestive tract

Table 1. Descriptive statistics of demographics and laboratory blood tests according to each digestive tract tumor.....86

CHAPTER 1

Introduction

CHAPTER 1.1

Background

CLINICAL BLOOD ANALYSIS AND THEIR IMPACT ON THE HEALTH CONDITION

Diagnosis in clinical activities is a continuous learning process characterized by subjective and objective factors that supports the identification of the most probable clinical condition according to the predefined signs, symptoms, clinical background, and clinical measurements of the patient (Badrick, 2013). Because clinical history and symptomatology are subjective, supporting complementary exams, namely clinical laboratory analysis, is considered a critical component of the clinical decision algorithm (Hicks, Carwardine, et al., 2021). Importantly, these tests are vital to discriminate amongst distinct signs or symptoms in different life contexts (e.g., age, gender, and family history) and for various purposes (Morabia & Zhang, 2004). For example: i) in the prenatal period, specific blood tests should be performed to screen anemia, blood group, hepatitis B, HIV, pre-eclampsia, rubella immunity, and syphilis; ii) in newborns, tests vary from amino acid metabolism disorders (i.e., classic phenylketonuria) to hemoglobin disorders (i.e., sickle cell anemia) or cystic fibrosis, among others (El-Hattab et al., 2018); iii) in childhood, adolescence, and during adulthood, tests are performed amongst other purposes for preventive purposes, of which the lipidic profile (total cholesterol) is highly recommended as a blood test, at least once every five years for men aged 35 or higher and women aged 40 or higher (Holland et al., 2006).

These blood examinations emerged as an untargeted monitoring tool that physicians use to assess patients' cellular, metabolic, and lipidic system panels. It is important to note that although they have widespread use, their demonstration of efficacy in many areas is still disputable (Rüttimann et al., 1992). Indeed, a meta-analysis from Krogsbøll L. et al. declared routine testing as a tool for increasing the number of diagnoses but without significant effect on mortality reduction (Krogsbøll et al., 2012). Consequently, most clinical testing has been applied in the symptomatic stage of the disease, either for diagnosis (aimed at early stages), monitoring, or prognosis (time-series) applications, sustaining the claim that the clinical results of blood analysis guide nearly 70% of medical decisions (Hallworth, 2011). This number varies according to the specialty and (even more) with the disease of study. In cardiovascular diseases (CVD), authors concluded that laboratory testing is mandatory in 73% to 94% of CVD guidelines across five international authorities, such as the European Society of Cardiology and the American College of Cardiology, among others (Hicks, Cardwardine, et al., 2021). Importantly, clinical blood analysis such as the blood counts (26%), the renal function (39%), and lipidic (25%) panels comprise the most recommended tests, as they are crucial not only for diagnosis but also in follow-up treatments (Hicks, Carwardine, et al., 2021). The ongoing monitoring of chronic diseases represents nearly

50% of all biochemical tests performed by central laboratories when compared to expenses related to screening or diagnostics (Whiting et al., 2019).

Moreover, continuous technical and scientific developments have promoted this notion of value provided by blood testing for distinct clinical purposes. Recently, these developments comprise both the employment of novel sensing tools and the implementation of innovative clinical decision support systems that enable significant value from collecting, processing, and translating available data, promising precision medicine. Indeed, precision health supported in multimodal clinical data (i.e., demographics, laboratory blood values, time series, omics, imaging, or others) aims to contrast with present clinical decisions supported in rule-based frameworks established through the comparison between individual versus population reference values passively updated under the rule 'one-size-fits-all' (A. Kline et al., 2022).

BLOOD METABOLITES IN DISEASE DIAGNOSIS & MANAGEMENT (GOLD STANDARD)

Whole blood is one of the most complex biologic samples to analyze. Thus, in most cases, there is a previous separation of its cellular (45%) and liquid plasma (55%) components; this permits further appraisal by the specialties of hematology and biochemistry, respectively. While hematology assesses blood cells' concentration, structure, and function, biochemistry evaluates the metabolic, hepatic, renal, lipidic, and electrolyte panels. Other analyses are performed in the serum, i.e., plasma without the clotting factors (i.e., fibrinogen), for coagulation studies (Robert Chatburn, 2010). Complete blood counts with comprehensive metabolic and lipid panels and specific metabolites are commonly carried out in clinical care.

In hematology, total blood counts are usually performed to assess the number of erythrocytes (RBC), leukocytes (WBC), and thrombocytes (PLT), with differential information in white blood cells, namely in the count of neutrophils, eosinophils, basophils, lymphocytes, and monocytes. Erythrocytes structure is measured in four different indexes: mean corpuscular volume (MCV), mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC), and red cell distribution width (RDW). Blood concentration is assessed by quantifying hemoglobin (Hb) (Celkan, 2020).

Typically, the clinical analysis of blood plasma assesses the comprehensive metabolic panel through the quantification of glucose, urea, creatinine, potassium, sodium, chloride, albumin, alanine aminotransferase, aspartate aminotransferase, and total calcium, and the lipidic panel by

the measurement of triglycerides, HDL-cholesterol, LDL-cholesterol, and total cholesterol. Uric acid and TSH are also considered for routine health examinations.

Since hemoglobin is the major constituent of red blood cells, its biological role directly impacts all other body functions by transporting oxygen and carbon dioxide. Hemoglobin is composed of two pairs of polypeptide chains (globins) and four prosthetic heme (iron ferrous) groups, allowing the molecule to reversibly combines with oxygen or carbon dioxide in a tension-mediated process: from the approximately 100 mmHg in the lungs to nearly 20 mmHg in tissues, the heme group releases CO_2 and binds O_2 in continuous cycles. A decrease in hemoglobin concentration indicates a reduction potential of the association between iron and O_2. When the Hb values become lower than 12.0 g/dL for women and 13.0 g/dL for men, anemia arises as a pathological state (McLean et al., 2009). Based on the blood structure, namely the MCV quantification, different types of anemia exist, namely microcytic (<80 fL), normocytic (80-100 fL), and macrocytic (>100 fL). Microcytic anemias include common iron-deficiency anemia (IDA) and thalassemias. In normocytic anemias, the red cells show normal MCVs and MCHCs, and subsequent differential diagnoses include hemolytic anemia, aplastic anemia, myelodysplastic syndrome, or anemia of renal failure, amongst others. Macrocytic anemias are highly associated with nutritional deficiencies (i.e., B12 vitamin or folate) (Matthew R. Pincus, 2011). Polycythemia vera (PV), usually discovered through the incidental measurement of high levels of hematocrit or hemoglobin, is a chronic myeloproliferative disorder characterized by the accumulation of phenotypically normal red cells, leukocytes, and platelets, in the absence of a recognizable physiologic stimulus. β-thalassemia trait and hypoxic erythrocytosis constitute additional examples of microcytic erythrocytosis. The MCV and the RDW indexes allow for the distinction between the three pathologies. While RDW is normal in the β-thalassemia trait, elevated levels arise in hypoxic erythrocytosis and PV (Spivak, 2012).

Abnormalities associated with white blood cell count are studied through the differential count and concentration of each granulocyte. Low levels of white blood cells, usually known as leukopenia, could be justified by a decrease in each cell type, i.e., neutropenia, monocytopenia, or lymphocytopenia. In contrast, an increase in each subtype is termed neutrophilia, basophilia, eosinophilia, monocytosis, and lymphocytosis. Associated differential diagnosis encompasses parasitic infection and allergic reactions in eosinophilia, widespread infection (bacterial or viral), or inflammation processes (trauma) in neutrophilia and hematologic malignancies in basophilia. Monocytosis is associated with hematologic diseases (leukemia) and specific infections

(tuberculosis). Lymphocytosis is usually associated with acute viral infections (mononucleosis), chronic infections (tuberculosis), and hematologic disorders (Marieb & Hoehn, 2012). Therefore, the systematic study of laboratory patterns generated by each blood analysis provides relevant information on the most probable clinical condition.

While the above reflects the general information provided by the analysis of the cellular component in blood, there are numerous possible parameters/panels to study in the plasma. Very briefly, to give some examples, a liver disease, such as hepatitis, cirrhosis, biliary disease, passive congestion, or fulminant hepatic failure, can be assessed by liver function tests, namely the levels of albumin, alkaline phosphatase, alanine aminotransferase, aspartate aminotransferase, direct bilirubin, total bilirubin, and total protein. High values of urea nitrogen and creatinine are suggestive of renal failure. Depending on the patient context, specific biomarkers could be measured to evaluate acute conditions; for example, high levels of c-reactive protein and cardiac troponin suggest inflammatory disease and myocardial infarction, respectively (McPherson & Pincus, 2011).

POINT-OF-CARE ANALYSIS

Point-of-care (POC) refers to the concept that displaces healthcare procedures from centralized units to near-patient sites, where care is delivered. Usually applied in diagnostic tests, the concept fits its deployment at physician offices, nursing facilities, operating rooms, intensive care units, emergency departments, ambulances, and pharmacies. POC testing (POCT) solutions encompass the scope of laboratory tests that use biofluids, such as blood, urine, or saliva, to guide medical decision-making in the areas of urinalysis, general chemistry, general immunology, hematology, immunohematology, endocrinology, toxicology, bacteriology, mycobacteriology, virology, parasitology, mycology, and cytology. These tests are typically characterized to be minimally invasive (e.g., capillary punction), easy to operate (no need for trained professionals), cost-effective, and able to perform real-time detection of the target analyte (Luppa, Müller, Schlichtiger, & Schlebusch, 2011). The COVID-19 pandemic highlighted the critical need for POCT, wherein screening methods were scarce. There was a considerable need to rapidly develop point-of-need tests to accurately screen the disease, independently of the patient's symptomatology or other viral infections as confounders. More than 400 rapid diagnostic tests were developed under a particular criterion of approval and commercialization promoted by the FDA: sensitivity not less than 80% and specificity not less than 98%, with tests carried out on a prospective cohort involving

no more than 60 people (half-only with the infection) (World Health Organization, 2021). These tests significantly reduced the time to diagnosis and endorsed infection control (Brendish et al., 2020).

Following the WHO "ASSURED" recommendations regarding tests design that matches the needs of the developing world, developed solutions should be affordable, sensitive, specific, user-friendly, rapid, robust, equipment-free, and deliverable to end-users (Martinez et al., 2010). In a recent perspective from Land et al., real-time connectivity and the ease of specimen collection (i.e., non-invasive) were proposed to improve future diagnostic solutions under the acronym of REASSURED (Land et al., 2019). Importantly, current implications of the available POCT reveal a positive impact on the pace of diagnosis, a consequent reduction in the overall patient cost, length of stay, readmittance, and morbidity, along with a general improvement in patient satisfaction and life quality (Kankaanpää et al., 2018). This is particularly relevant not only in the screening and controlling of infectious diseases but also in the rise of non-communicable, chronic diseases, such as diabetes and cardiovascular disorders (Mabey et al., 2004).

Available in-vitro rapid tests exist for the screening of specific targets such as SARS-CoV-2 (Roche, 2021), HIV (Roche, 2016), or TSH (Hangzhou AllTest Biotech Co., 2016) and coupled systems with external sensors offer the monitoring of metabolites such as glucose (Abbott, 2015), ethanol (PAL, 2022), lactate, cholesterol or triglycerides (Roche, 2017), bilirubin (PFAFF, n.d.), lipids (Abbott, 2019), hemoglobin (HemoCue, 2000), prothrombin time (INR ratio) (Fernando et al., 2015) albumin/creatinine ratio and HbA1C (Peluso-Lapsley, 2018) and cardiac markers such as troponin, NT-proBNP, CK-MB, D-Dimer and myoglobin (Roche, 2022), among others. In addition, emerging developments in non-invasive technologies allowed the deployment of intelligent, continuous glucose monitoring devices, not only in T2DM (Abbott, 2017) but also in integration with insulin pump therapy, enabling control for T1DM (Medtronic, 2018). In parallel, strategies that facilitate logistics between in-home blood collection with external laboratories (i.e., dried blood spot testing) also provide clinical results in collected blood through volume-specific cartridges (Neoteryx, 2018), spotting targeted metabolites from home.

The deployment of POCT in medical laboratories (operated by trained personnel), in the ward (as point-of-need diagnostic services), or for self-testing (home use) is, therefore, a matter of regulation but also a subject of study related to cost-effectiveness analyses (van der Pol et al., 2021). Such devices should comply with regular procedures, calibrations, controls, and maintenance along with the requirements related to storage and temperature, which depends in

part on the professional motivation to fulfill those tasks (Lenters-Westra & Slingerland, 2010). While central laboratories have desktop standing units supervised by experienced technicians that assure regular, standardized quality-control tests, POCT will need periodic cross-correlation tests with the gold standard. Integration of POCT in-hospital workflow requires a significant clinical benefit relating to the test need for immediate decision-making with improved turn-around time. Indeed, reducing the time for analysis is critical in several situations. For example, i) the quantification of troponin, NT-proBNP, glucose, D-dimer, and CRP could suggest important acute health conditions such as myocardial infarction, heart failure, glycemic dysfunctions, pulmonary embolism, or infection respectively (Cooke et al., 2020) or ii) in trauma patients, it is mandatory to quickly assess coagulopathy and anemia from blood loss through the measurement of hemoglobin, hematocrit, electrolytes, ionized calcium, prothrombin time, partial thromboplastin time, lactates, and arterial blood gases such as pH, pO_2, PCO_2, and HCO_3 (Bhandarkar et al., 2018)(Kasturi Mukherjee, Debojyoti Bhattacharjee, Jayati Roy Choudhury, 2022). The rapid accomplishment of these measurements is only feasible by combining specific cartridges with dedicated detection systems operated by trained healthcare professionals (Abbott, 2021). A note to highlight that important considerations arise after clinical analysis regarding data management and interfaces; for instance, connectivity between POCT and electronic medical records (i.e., HIS, LIS) is fundamental to promoting real-time picturing for clinicians and enables the establishment of specific information profiles for each patient.

Nonetheless, monitoring chronic conditions at home is still challenging due to the lack of POCT solutions that can measure (in clinical grade) several metabolites in real-time at low cost. Unfortunately, existing multi-analyte solutions are not economically viable to address home or even primary care facilities. The latter is the major limitation of current POCT since cheaper solutions are analyte-specific, enabling the direct assessment of one metabolite but hindering the possibility of performing a multi-analyte measurement to assess more complex insights about a specific pathological state. Other limitations derive from potential errors due to insufficient training, user-training cost, insufficient quality-control procedures, inadequate storage of results, or system complexity. Near-future developments aim to address these issues through the simultaneous detection of analytes (multiplexed POCT) by privileging non-invasive measurement systems using wearables (i.e., watches, patches, tattoos, glasses, contact lenses) in several biofluids such as blood, urine, saliva, sweat or tears. Self-monitoring in smartphones will probably be an extension of laboratories, offering more accessible access to health care for every patient. However, all these

developments depend on complex regulatory approvals and lengthy clinical validation studies that will impact market positioning and clinical adoption (Chin et al., 2012).

ARTIFICIAL INTELLIGENCE IN CLINICAL BLOOD ANALYSIS

In parallel and, in most cases, with joint development, artificial intelligence is playing an increasingly important role in diagnosis, either through the enhancement of current state-of-the-art technologies or simply through the learning of EHR data provided by gold-standard technologies in the different domains of clinical analysis (Reardon, 2019). Defined as a subdivision of computer science, artificial intelligence (AI) aims to automatically understand and create intelligent systems based on high amounts of data through different algorithms and learning strategies (Shukla Shubhendu & Vijay, 2013). The application of this technology intrinsically depends on the type, size, and dimensionality of the data. Whether data points appear in a labeled or unlabeled fashion compromises the choice between applying supervised or unsupervised learning. Indeed, supervised learning fits an algorithm on labeled data points for learning and model development used in the label prediction of new data points (Niculescu-Mizil & Caruana, 2005). In the example of anemia, and 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 can be settled in the form of a classification problem that identifies discrete states, such as different stages of anemia, or can be established in the form of 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 not known. In that case, learning might be conducted in an unsupervised fashion where algorithms infer underlying patterns in unlabeled data to find sub-clusters of the original data, identify outliers, or produce low-dimensional representations of the same data. In this way, it could be possible to recognize new associations that were not perceptible before. The example above would potentially separate clusters of patients between anemic and non-anemic, even without knowing the true value of hemoglobin. Other learning strategies such as semi-supervised (accepts both labeled or unlabeled data), reinforcement (rewarding by feedback through trial and error) or deep learning (using high-level interactions on hidden layers) enables data modeling for structured (highly organized in relational databases), unstructured (without predefined format), semi-structured (organized but not on relational databases) and metadata (data properties about data) (Sarker, 2021).

Apart from the data collection and the selected learning strategy, a successful machine learning (ML) pipeline comprises several other stages, such as data preprocessing, modeling, validation, and deployment. Preprocessing evaluates data regarding representativeness and bias concerning the number of samples and features according to the outcome of interest. This is particularly important in classification problems of categorical or binary events, where low representativeness could endorse the existence of 'blind spots', wherein the ML models will fail in predicting the correct outcome due to the inexistence of training data representative of that sample space. In practice, the foundation of an ML model should be accomplished through the data split between three sets of training, validation, and testing. While training allows to fit and determine model coefficients, internal evaluation is used to assess prediction errors and to guide model selection. Usually, at the training stage, models exhibit higher accuracies which can result in overfitting (perfect generalization in the training set with poor or substantially smaller fitting in the test set). This common misuse of ML approaches is tendentially corrected with resampling methods such as cross-validation or bootstrapping, offering higher data usage than partial splitting. The selection of a model depends on its performance metrics, which are also dependent on the nature of the problem, namely RMSE in regression, AUROC in classification, and the concordance index (c-index) in survival or censoring. In medicine, decision curve analysis is an important metric that quantifies the clinical benefit of the solution according to a clinical decision support system and predefined thresholds of probability (Ziobrowski, Hannah N., Sonneville, Kendrin R. Eddy, Kamryn T., Crosby, Ross D., Micali, Nadia, Horton, Nicholas J., Field, 2019). The development pipeline should follow testing or external validation. Ideally, in prospective trials, the selected model is applied for prediction in a different location or population that must be statistically similar to the training data. Indeed, external validation with consistent model accuracy is one of this technology's most challenging aspects due to bias and representativeness risks. Federated learning addresses this issue by connecting multiple individual learning sites that update a central model further deployed at each site, preserving data governance and privacy, and allowing continuous updates, keeping data 'healthy' (Dayan et al., 2021). Deployed systems are offering guidance and 'second opinions' and an increasing list of FDA or CE-approved software applications in the fields of cardiology, psychiatry, endocrinology, radiology, geriatrics, neurology, orthopedics, emergency medicine, ophthalmology, pathology, or oncology is providing aid to clinical diagnosis worldwide (The Medical Futurist, 2022).

ML application to clinical blood analysis showed promising results. Significant developments were accomplished for several groups of diseases (ICD-10). Research has been mostly conducted through supervised learning - especially binary classification, with regression and clustering approaches representing nearly 15% of the work. Relevant findings were accomplished in the features importance stage, not only by confirming the relevance of metabolites that were previously associated with the disease (i.e., age in cardiovascular diseases) but also by showing important associations between the model's performance and other parameters (i.e., RDW in bloodstream infection). In validation, comparing clinically used scores (i.e., APRI or FIB-4) with developed models demonstrated how these tools could leverage diagnosis, with potential integration into clinical decision support systems (Cao et al., 2013). Integrating demographics with an accurate exhibition of time is considered highly relevant for the baseline comparison of each analysis across the studied population. In oncology, colorectal, leukemias differentiation, and non-specified cancer diagnosis in primary care were studied with binary classification models. Kinar et al. demonstrated a stable colorectal diagnostic sensitivity between 480 and 240 days before diagnosis (AUC≈0.76), with a posterior increase in the last 240 days (AUC>0.80), using a decision tree algorithm computed with the values of MCV, MCH, MCHC, RDW, hematocrit, and hemoglobin (Kinar et al., 2016). CBC alone was proven to differentiate six lineages of leukemia (AUC=0.83) (Haider et al., 2022), and Soerensen et al. evaluated two primary care cohorts regarding the probability of cancer in 90 days, selecting albumin and platelets count as the critical features for logistic regression (AUC=0.80) (Soerensen et al., 2021). While these reports sustain the claim related to the intrinsic value of clinical blood analysis leveraged by ML algorithms, its evaluation in prospective trials is scarce, and sufficient accuracy and reliability are rarely achieved, which endorses the slow implementation in clinical settings (Seneviratne et al., 2020). Moreover, the inexistence of a gold-standard approach for developing clinical models (i.e., model architecture) hinders comparisons regarding robustness and trajectories performance in time. The full advantage of longitudinal data could help to tackle some of the above limitations since most models use only a snapshot of the available information. Indeed, artificial intelligence algorithms provide a tool to analyze the biochemical profile of each patient with further evaluation of the developed models in predicting specific outcomes. By testing these models in a preclinical setting, we believe these systems will become broadly accepted in precision medicine.

CHAPTER 1.2

Goal, Aims, and Approach

OVERALL GOAL, AIMS AND APPROACH

The foremost goal in the work developed under this thesis was to study clinical blood analysis with specific data mining frameworks for extracting clinical-grade insights valuable for physicians, in the context of medical diagnosis. The overall investigative approach combined blood analysis data collection and pre-processing with statistical, machine learning, and pattern recognition algorithms for multivariate data analysis. This research dissertation was divided into four aims for which we next provide further detail (**Figure 1**), addressing their rationale and summarizing the ensuing approach.

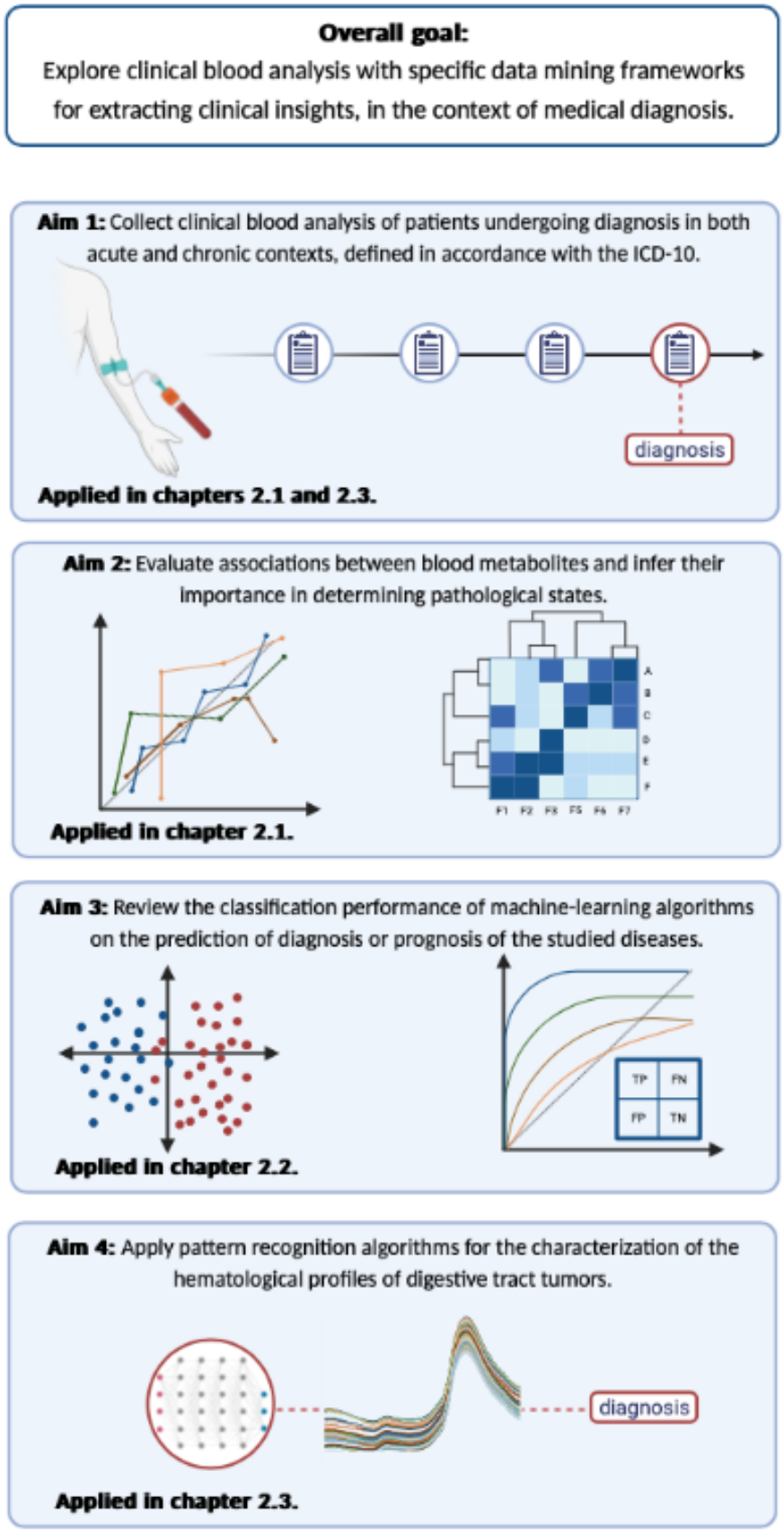


Figure 1. The overall goal and the four specific aims of this dissertation. Created with BioRender.com.

Aim 1: Collect clinical blood analysis of patients undergoing diagnosis in both acute and chronic contexts, defined in accordance with the ICD-10.

Rationale

Clinical blood data comprises the final stage of a gold-standard methodology wherein clinicians guide its decisions, either for screening, diagnosis, or monitoring. This patterned process enables the continuous feed of electronic medical records (EMR) through the generation of high amounts of data, under standardized criteria. The assessment of this consistent information associated with predefined target outcomes, configures an important investigative pathway for the application of data mining frameworks. In this aim, we collected specific data sets according to the patient context, combining demographics, clinical blood analysis, disease codification (ICD-10), and the disease codification date.

Approach

For each patient, demographics (age and gender), hematological (erythrocytes, hemoglobin, hematocrit, rdw, leukocytes, neutrophils, eosinophils, basophils, lymphocytes, platelets, mean globular volume, mean globular hemoglobin, hemoglobin, and the average concentration of globular hemoglobin), and biochemical information (glucose, urea, creatinine, potassium, sodium, chlorides, albumin, ALP, AST, ALT, total calcium, triglycerides, total cholesterol, HDL cholesterol, LDL cholesterol, total bilirubin, uric acid, gamma-GT, c-reactive protein) were acquired. The collection process aggregated data of acute (emergency department admission) and oncologic patients admitted.

Aim 2: Evaluate associations between blood metabolites and infer their importance in determining pathological states.

Rationale

Under basal levels, an intrinsic correlation exists between levels of specific blood parameters. Therefore, statistical associations between metabolites relative to specific clinical contexts configure important insights, especially in variable importance and further modeling. Inferential statistics evaluates correlation and covariance among variables, determines importance relative to the target, and contributes to improved performance metrics with increasing model explainability and interpretability, with reduced risk of overfitting.

Approach

We performed preprocessing and statistical tests to determine correlation, covariance and features importance relative to the target. Bias and representativeness were assessed in demographics using the normality test as an indicator of 'blind spots' existence. Specific associations of the most relevant metabolites were accomplished for each pathological timeframe.

Aim 3: Review the classification performance of machine-learning algorithms on the prediction of diagnosis or prognosis of studied diseases in the context of routine blood tests.

Rationale

Studies integrating routine blood analysis with disease discrimination exist. Still, a complete overview of these studies regarding principal findings for selected blood parameters, tested algorithms, and external validation for specific disease groups was unavailable, which we approached with this aim.

Approach

We reviewed studies between 2012 and 2022 that used artificial intelligence algorithms to analyze routine blood tests for the diagnosis or prognosis of several diseases. Focusing on the principal findings of each study, studies were aggregated according to the ICD-10 disease group, and selected features were discussed in comparison with traditional blood parameters associated with the disease. AUC was used to highlight performance metrics in internal and external validation tests.

Aim 4: Apply pattern recognition algorithms to characterize the hematological profiles of digestive tract tumors.

Rationale

Pattern recognition algorithms identify hidden structures and underlying patterns in data without prior specification of relevant blood parameters. Instead of manually performing correlation analysis to assess potential biomarkers, AI fits the algorithms into the data and creates models that could ultimately identify clinical profiles.

Approach

In the modeling stage, normalization was applied to the feature space, and the used algorithm maximized the variance across the data to highlight patterns that could differentiate groups, either across pathologies or in specific timeframes across the available data. Unsupervised (i.e., principal components analysis) and supervised (i.e., logistic regression) generated models for predicting each disease group from the others, with performance metrics of sensitivity, specificity, and AUC.

Taken together, these aims share the common objective of extracting valuable clinical data to model disease diagnosis, according to the demographics and routine blood analysis of patients. In the following chapters we will provide detail on the methodologies used to address these aims, the results obtained, and the conclusions withdrawn from them.

REFERENCES

Abbott. (2015). *FreeStyle Precision Neo*.

Abbott. (2017). *FreeStyle Libre System*. 22. https://www.myfreestyle.com/provider/sites/all/themes/provider20/pdf/FreeStyle-Libre-In-Service-Guide.pdf

Abbott. (2019). The Accuracy and Reproducibility of a Rapid, Fingerstick Method for Measuring a Complete Lipid Profile Is Comparable to Clinical Diagnostic Laboratory Methods. *Technical Brief*, *Atp Iii*, 8–9.

Abbott. (2021). *i-STAT EG7 + Cartridge*. 1–19.

Badrick, T. (2013). Evidence-based laboratory medicine. *The Clinical Biochemist. Reviews*, *34*(2), 43–46. https://pubmed.ncbi.nlm.nih.gov/24151340

Bhandarkar, P., Pal, R., Munivenkatappa, A., Roy, N., Kumar, V., & Agrawal, A. (2018). Distribution of laboratory parameters in trauma population. *Journal of Emergencies, Trauma and Shock*, *11*(1), 10–14. https://doi.org/10.4103/JETS.JETS-70-17

Brendish, N. J., Poole, S., Naidu, V. V., Mansbridge, C. T., Norton, N. J., Wheeler, H., Presland, L., Kidd, S., Cortes, N. J., Borca, F., Phan, H., Babbage, G., Visseaux, B., Ewings, S., & Clark, T. W. (2020). Clinical impact of molecular point-of-care testing for suspected COVID-19 in hospital (COV-19POC): a prospective, interventional, non-randomised, controlled study. *The Lancet Respiratory Medicine*, *8*(12), 1192–1200. https://doi.org/10.1016/S2213-2600(20)30454-9

Cao, Y., Hu, Z. De, Liu, X. F., Deng, A. M., & Hu, C. J. (2013). An MLP classifier for prediction of HBV-induced liver cirrhosis using routinely available clinical parameters. *Disease Markers*, *35*(6), 653–660. https://doi.org/10.1155/2013/127962

Celkan, T. T. (2020). What does a hemogram say to us? *Turkish Archives of Pediatrics*, *55*(2), 103–116. https://doi.org/10.14744/TurkPediatriArs.2019.76301

Chin, C. D., Linder, V., & Sia, S. K. (2012). Commercialization of microfluidic point-of-care diagnostic devices. *Lab on a Chip*, *12*(12), 2118–2134. https://doi.org/10.1039/c2lc21204h

Cooke, J., Llor, C., Hopstaken, R., Dryden, M., & Butler, C. (2020). Respiratory tract infections (RTIs) in primary care: Narrative review of C reactive protein (CRP) point-of-care testing (POCT) and antibacterial use in patients who present with symptoms of RTI. *BMJ Open Respiratory Research*, *7*(1). https://doi.org/10.1136/bmjresp-2020-000624

Dayan, I., Roth, H. R., Zhong, A., Harouni, A., Gentili, A., Abidin, A. Z., Liu, A., Costa, A. B., Wood, B. J., Tsai, C. S., Wang, C. H., Hsu, C. N., Lee, C. K., Ruan, P., Xu, D., Wu, D., Huang, E., Kitamura, F. C., Lacey, G., … Li, Q. (2021). Federated learning for predicting clinical outcomes in patients with COVID-19. *Nature Medicine*, *27*(10), 1735–1743. https://doi.org/10.1038/s41591-021-01506-3

El-Hattab, A. W., Almannai, M., & Sutton, V. R. (2018). Newborn Screening: History, Current Status, and Future Directions. *Pediatric Clinics of North America*, *65*(2), 389–405. https://doi.org/10.1016/j.pcl.2017.11.013

Fernando, R., Ak, J., Kennedy, S., Lessard, C., Olson, K., Scribner, A., Fernando, R., Ak, J., Kennedy, S., Lessard, C., Olson, K., Scribner, A., Stride, X., Division, G., & Headquarters, G. S. (2015). *PT / INR Test Performance of the Xprecia Stride Coagulation Analyzer on Capillary Blood is Equivalent to a Reference Laboratory Hemostasis System PT / INR Test Performance of the Xprecia Stride Coagulation Analyzer on Capillary Blood*. 2–5.

Haider, R. Z., Ujjan, I. U., Khan, N. A., Urrechaga, E., & 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*(1). https://doi.org/10.3390/diagnostics12010138

Hallworth, M. J. (2011). The “70% claim”: What is the evidence base? *Annals of Clinical Biochemistry*, *48*(6), 487–488. https://doi.org/10.1258/acb.2011.011177

Hangzhou AllTest Biotech Co., L. (2016). *TSH Rapid Test Cassette (Whole Blood / Serum / Plasma) Package Insert*. 402.

HemoCue. (2000). *Hb 801 Analyzer - HemoCue*. 22.

Hicks, A. J., Cardwardine, Z. L., Hallworth, M. J., & Kilpatrick, E. S. (2021). Using clinical guidelines to assess the potential value of laboratory medicine in clinical decision-making. *Biochem Med (Zagreb)*, *1*(25), 717. https://doi.org/10.1001/jama.1883.02390250001001

Hicks, A. J., Carwardine, Z. L., Hallworth, M. J., & Kilpatrick, E. S. (2021). Using clinical guidelines to assess the potential value of laboratory medicine in clinical decision-making. *Biochem Med (Zagreb)*, *31*(1). https://doi.org/10.1097/00005392-199503001-00031

Holland, W. W., Stewart, S., & Masseria, C. (2006). Policy brief: screening in Europe. *Journal of Health and Social Behavior*, *55*(2), 125. http://www.euro.who.int/en/home/projects/observatory

Kankaanpää, M., Holma-Eriksson, M., Kapanen, S., Heitto, M., Bergström, S., Muukkonen, L., & Harjola, V. P. (2018). Comparison of the use of comprehensive point-of-care test panel to conventional laboratory process in emergency department. *BMC Emergency Medicine*, *18*(1), 4–9. https://doi.org/10.1186/s12873-018-0198-x

Kasturi Mukherjee1, Debojyoti Bhattacharjee1, Jayati Roy Choudhury1, R. B. (2022). Association of Serum Biomarkers with the Mortality of Trauma Victims in a Level -1 Trauma Care Centre of Eastern India. *Bull Emerg Trauma 2022;10(1):33-39.*, *10*(1), 33–39. https://doi.org/10.30476/BEAT.2022.89155.1222.Introduction

Kinar, Y., Kalkstein, N., Akiva, P., Levin, B., Half, E. E., Goldshtein, I., Chodick, G., & Shalev, V. (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. *Journal of the American Medical Informatics Association*, *23*(5), 879–890. https://doi.org/10.1093/jamia/ocv195

Kline, A., Wang, H., Li, Y., Dennis, S., Hutch, M., Xu, Z., Wang, F., Cheng, F., & Luo, Y. (2022). Multimodal Machine Learning in Precision Health. *Npj Digital Medicine*, 1–14. https://doi.org/10.1038/s41746-022-00712-8

Krogsbøll, L. T., Jørgensen, K. J., Grønhøj Larsen, C., & Gøtzsche, P. C. (2012). General health checks in adults for reducing morbidity and mortality from disease: Cochrane systematic review and meta-analysis. *BMJ (Online)*, *345*(7884), 1–13. https://doi.org/10.1136/bmj.e7191

Land, K. J., Boeras, D. I., Chen, X. S., Ramsay, A. R., & Peeling, R. W. (2019). REASSURED diagnostics to inform disease control strategies, strengthen health systems and improve patient outcomes. *Nature Microbiology*, *4*(1), 46–54. https://doi.org/10.1038/s41564-018-0295-3

Lenters-Westra, E., & Slingerland, R. J. (2010). Six of eight hemoglobin A1c point-of-care instruments do not meet the general accepted analytical performance criteria. *Clinical Chemistry*, *56*(1), 44–52. https://doi.org/10.1373/clinchem.2009.130641

Mabey, D., Peeling, R. W., Ustianowski, A., & Perkins, M. D. (2004). Diagnostics for the developing world. *Nature Reviews Microbiology*, *2*(3), 231–240. https://doi.org/10.1038/nrmicro841

Marieb, E. N., & Hoehn, K. (2012). Blood Composition and Functions. *Human Anatomy & Physiology*, 634–657. http://www.phschool.com/atschool/florida/pdfbooks/sci_Marieb/pdf/Marieb_ch17.pdf

Martinez, A. W., Phillips, S. T., Whitesides, G. M., & Carrilho, E. (2010). *Diagnostics for the Developing World: Microfluidic Paper-Based Analytical Devices*. *82*(1), 3–10.

Matthew R. Pincus, N. Z. A. J. (2011). INTERPRETING LABORATORY RESULTS. In *Henry's Clinical Diagnosis and Management22th Edition.pdf* (pp. 92–94). Elsevier.

McLean, E., Cogswell, M., Egli, I., Wojdyla, D., & De Benoist, B. (2009). Worldwide prevalence of anaemia, WHO Vitamin and Mineral Nutrition Information System, 1993-2005. *Public Health Nutrition*, *12*(4), 444–454. https://doi.org/10.1017/S1368980008002401

McPherson, R. A., & Pincus, M. R. (2011). Disease/Organ Panels. In *Henry's Clinical Diagnosis and Management22th Edition.pdf* (pp. 1513–1514).

Medtronic. (2018). *Guardin connect* (Issue November).

Morabia, A., & Zhang, F. F. (2004). History of medical screening: From concepts to action. *Postgraduate Medical Journal*, *80*(946), 463–469. https://doi.org/10.1136/pgmj.2003.018226

Neoteryx. (2018). *Blood Collection Made Easy: Anywhere, Anytime, by Anyone Blood Collection Devices Powered by Volumetric Absorptive Microsampling (VAMS™) Technology*.

Niculescu-Mizil, A., & Caruana, R. (2005). Predicting good probabilities with supervised learning. *ICML 2005 - Proceedings of the 22nd International Conference on Machine Learning*, *May*, 625–632.

https://doi.org/10.1145/1102351.1102430
PAL. (2022). *Albio Alcohol Biometer*. https://www.palfinland.com
Peluso-Lapsley, M. (2018). *DCA Vantage Analyzer and HbA1c*. *September*, 1–45.
PFAFF. (n.d.). *Bilimeter 3D*. https://www.bilimeter.de/en-gb/page-1
Reardon, S. (2019). Rise of Robot Radiologists. *Nature*, *576*(7787), S54–S58. https://doi.org/10.1038/d41586-019-03847-z
Robert Chatburn, E. M.-C. (2010). Hematology. In *Handbook of Respiratory Care Third Edition* (pp. 54–63).
Roche. (2016). *cobas HIV-1*. https://doi.org/10.12968/prtu.2016.54.51
Roche. (2017). *AccuTrend Plus*. https://www.corporate.roche.pt/pt/produtos-e-servicos/solucoes-diagnostico/point-of-care/autocontrolo/accutrend–plus.html
Roche. (2021). *SARS-CoV-2 Rapid Antigen Test*.
Roche. (2022). *Roche CARDIAC IQC*.
Rüttimann, S., Clémençon, D., & Dubach, U. C. (1992). Usefulness of complete blood counts as a case-finding tool in medical outpatients. *Annals of Internal Medicine*, *116*(1), 44–50. https://doi.org/10.7326/0003-4819-116-1-44
Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, *2*(3). https://doi.org/10.1007/s42979-021-00592-x
Seneviratne, M. G., Shah, N. H., & Chu, L. (2020). Bridging the implementation gap of machine learning in healthcare. *BMJ Innovations*, *6*(2), 45–47. https://doi.org/10.1136/bmjinnov-2019-000359
Shukla Shubhendu, S., & Vijay, J. (2013). Applicability of Artificial Intelligence in Different Fields of Life. *International Journal of Scientific Engineering and Research (IJSER)*, *1*(1), 28–35. www.ijser.in
Soerensen, P. D., Christensen, H., Gray Worsoe Laursen, S., Hardahl, C., Brandslund, I., & Madsen, J. S. (2021). Using artificial intelligence in a primary care setting to identify patients at risk for cancer: A risk prediction model based on routine laboratory tests. *Clinical Chemistry and Laboratory Medicine*, 1–12. https://doi.org/10.1515/cclm-2021-1015
Spivak, J. (2012). Polycythemiavera and other myeloproliferative diseases. In K. D. et al editors Longo D., Fauci A. (Ed.), *Harrison's principles of internal medicine* (18th ed., pp. 898–904). McGraw-Hill.
Svensson, C. M., Hübler, R., & Figge, M. T. (2015). Automated Classification of Circulating Tumor Cells and the Impact of Interobsever Variability on Classifier Training and Performance. *Journal of Immunology Research*, *2015*. https://doi.org/10.1155/2015/573165
The Medical Futurist. (2022). *FDA approved AI-based algorithms*. https://medicalfuturist.com/fda-approved-ai-based-algorithms/
van der Pol, S., Rojas Garcia, P., Antoñanzas Villar, F., Postma, M. J., & van Asselt, A. D. I. (2021). Health-Economic Analyses of Diagnostics: Guidance on Design and Reporting. *PharmacoEconomics*, *39*(12), 1355–1363. https://doi.org/10.1007/s40273-021-01104-8
Whiting, D., Croker, R., Watson, J., Brogan, A., Walker, A. J., & Lewis, T. (2019). Optimising laboratory monitoring of chronic conditions in primary care: A quality improvement framework. *BMJ Open Quality*, *8*(1), 1–8. https://doi.org/10.1136/bmjoq-2018-000349
World Health Organization. (2021). Antigen-detection in the diagnosis of SARS-CoV-2 infection. *Interim Guidance*, *October*.
Ziobrowski, Hannah N., Sonneville, Kendrin R. Eddy, Kamryn T., Crosby, Ross D., Micali, Nadia, Horton, Nicholas J., Field, A. E. (2019). Reporting and Interpreting Decision Curve Analysis: A Guide for Investigators. *European Urology*, *65*(4), 139–148. https://doi.org/10.1016/j.eururo.2018.08.038.Reporting

CHAPTER 2

Experimental work

CHAPTER 2.1

Correlation Analysis between Hemoglobin and C-Reactive Protein in Patients Admitted to an Emergency Unit

The work presented in this chapter has been published in the Journal of Clinical Medicine.

Miguel A. Santos-Silva, Nuno Sousa, João Carlos Sousa. (2021). **Correlation analysis between Hemoglobin and C-Reactive Protein in patients admitted to an emergency unit.** *Journal of Clinical Medicine, 10, 5411.*

https://doi.org/10.3390/jcm10225411

ABSTRACT

Anemia and inflammation are common clinical conditions in emergency departments. This study explored a cohort of patients admitted to the emergency department with a particular interest in determining the frequency of anemia and inflammation and the association between hemoglobin (Hb) and C-reactive protein (CRP) concentrations. The study included 125 patients categorized according to their demographic (gender and age) and clinical condition (Hb and CRP concentrations, pathological background, and diagnostic). We found that anemia and inflammation were simultaneously present in 36.0% of the cohort, reaching 67.0% in patients that were subsequently hospitalized. The Hb level was significantly lower in patients with elevated concentration of CRP when compared to individuals with normal CRP levels (11.58 ± 2.23 vs. 13.25 ± 1.80, $p = 0.001$); furthermore, we found a significantly negative correlation between Hb concentration and the CRP level ($rs = -0.42$, $p < 0.001$). The linear regression model applied to the cohort showed that CRP levels explain 15% of Hb variations. The sensitivity of the CRP/Hb ratio (cut-off = 1.32) as a predictor of hospitalization was 80.0%, with a specificity of 68.4% for all patients. These findings confirmed the prevalence of anemia and inflammation and identified a moderate but significant association between Hb and serum CRP in a heterogeneous group of patients admitted to the emergency department.

KEYWORDS

hemoglobin; C-reactive protein; emergency; correlation

INTRODUCTION

Admissions to emergency departments (EDs) comprise a very heterogeneous population, from patient demographics to the type of symptoms and conditions (chronic or acute) that require immediate care [1]. Although many studies are conducted for specific pathologies with highly isolated groups of individuals, many of the patients admitted to EDs present multiple symptoms and comorbidities, often associated with chronic pathologies that lack characterization [2]. The association between the clinical history, symptoms, and data from exams such as blood laboratory analysis is relevant to understanding the patient's health problem [3]. Furthermore, the information retrieved from blood laboratory analysis influences 70% of medical decisions [4]. Thus, this work focused on addressing whether there is an association between two of the most common blood metabolites evaluated in the context of ED admissions, i.e., hemoglobin (Hb) and C-reactive protein (CRP), with the diagnostic made on a patient's admission.

According to the World Health Organization (WHO), the prevalence of anemia varies between 22.9% and 26.7% (95% CI), affecting almost 1.62 billion people [5]. Typically, anemia encompasses low levels of circulating red blood cells (RBCs) or Hb concentration and is defined with Hb threshold values of <13 g/dL and <12 g/dL for men and women, respectively [5]. Laboratory characterization of anemia is accomplished either by a morphologic or kinetic approach. The morphological approach categorizes anemia according to RBC mean corpuscular volume (MCV), namely microcytic (MCV < 80 fL), normocytic (80–100 fL), and macrocytic (>100 fL) anemias. The kinetic approach differentiates anemias caused by the increased loss of RBCs (due to chronic or acute hemorrhages), increased destruction of RBCs (caused by acquired or inherited factors), and decreased production of RBCs (due, for example, to the lack of essential vitamins or hormones, or when the patient undergoes chemotherapy or exposure to radiation) [6].

The most prevalent cause of anemia is iron deficiency (IDA), followed by anemia of chronic disease (ACD) [7]. ACD, also named anemia of inflammation, has a high prevalence, particularly in the elderly (70%), due to the continuous increase of chronic diseases. ACD is mainly caused by infections (chronic or acute), cancer, autoimmune disorders, chronic rejection after solid-organ transplantation, and chronic kidney disease and inflammation [7]. Recently, congestive heart failure, chronic pulmonary disease, and obesity have also been associated with ACD [8]. Briefly, these pathologies stimulate the production of proinflammatory cytokines that enable iron sequestration and cause shortened RBC survival [9]. In the case of emergency departments specifically, existing studies reported significant variations of Hb concentration, with an average

prevalence of 44% [10] anemic individuals, independent of age and gender. Notably, this value increases in the elderly cohort (+70 years), and it is higher in males than females [11].

CRP is an acute-phase protein and pattern-recognition marker of systemic inflammation. CRP is frequently measured in emergency settings and synthesized in the liver in response to cytokines or the tumor necrosis factor produced during infections or inflammatory mechanisms [12,13]. The physiological behavior of CRP involves a 4–6 h increase after inflammation initiates, doubling its value every 8 h with a peak at 36–50 h, ending with a sharp decrease once the inflammation is resolved [14]. However, increased CRP concentrations over time, rather than spikes, are intrinsically related to several chronic pathologies, namely cardiovascular diseases, atherosclerosis, hemorrhagic stroke, Alzheimer's disease, and Parkinson's disease [15].

Some studies describe the relationship between Hb and CRP. While Ziv-Baran et al. found a marginal ($r_s = -0.025$, $p < 0.001$) association between both proteins in 16,095 healthy individuals undergoing a routine health examination, Heidari et al. described a moderate linear association ($r_s = -0.607$, $p = 0.001$) between Hb and CRP in 73 hemodialysis patients [16]. Although these studies are not comparable as they focus on substantially different cohorts, the association between Hb and CRP was analogous, confirming its possible relevance for clinical diagnosis and prognosis. This aspect was mostly visible in the hemodialysis study, where the inflammatory process, as evaluated by CRP levels, reduced Hb levels linearly [17]. Additionally, it is well-known that decreasing Hb levels linked to inflammatory states highly correlate to higher mortality [18]. Monitoring Hb in patients with inflammatory conditions is, thus, of relevance for patient and treatment monitoring [19].

Nevertheless, whether Hb and CRP associate in patients admitted to the emergency department has not yet been studied, which is what we approached in this work. Additionally, we explored how this potential association would relate to patient diagnosis at admission and whether it could support the clinical decision in the context of emergency department admission.

MATERIALS AND METHODS

Study design and Population

This work is a retrospective observational study of Portuguese individuals who attended the emergency department of the Hospital de Braga between May and December of 2019. The hospital assigns an identification code to compile the patient's clinical information longitudinally. The database contains demographic information (gender and age), clinical information (clinical analyses and exams results), and information regarding past and future appointments or procedures. This identification code allows for a direct and anonymous connection to each patient, according to the ethical guidelines of the Braga Hospital Ethics Committee. The study was approved by the Portuguese data protection commission, and the Braga Hospital and the University of Minho Ethics Committees (SECVS-011/2013) under the project "My Health DoIT".

The study focused on a sample of adult patients (18 years or older) with entire clinical history, including pathological background, symptoms description, length of stay, and blood analysis. The sample collection method was random, namely on non-consecutive days and on different time intervals over the various days, to avoid repeated causes of emergency department arrivals (for instance, acute myocardial accidents, which are prevalent in the morning). The blood analysis comprised a complete blood count (CBC) and chemistry panel. Participants with incomplete information regarding clinical history were excluded. Only blood tests with complete cell counts and Hb and CRP concentrations were considered since one of the study objectives was an evaluation of the relation between Hb and CRP concentrations.

Data Source, Measurement, and Variables

Patients' anonymized demographic (gender and age) and clinical (pathological background, symptoms description, length of stay, and blood analysis) data were retrieved from the Clinic Academic Center (2CA, Braga, Portugal) of the Hospital de Braga. CBC, including erythrocytes, platelets, and white blood cells count, and Hb concentration, hematocrit ratio, and red cell distribution width, were analyzed in the hospital clinical pathology laboratory using standard methods (Sysmex XE-2100, Sysmex Inc, Mundelein, IL, USA). Serum CRP was measured with traditional immunoturbidimetric assay by the ADVIA system (Siemens Healthcare Diagnostics Inc, Tarrytown, NY, USA).

Symptoms, pathological background, and primary diagnostic performed in the emergency department were merged according to the International Statistical Classification of Diseases and

Related Health problems 10th revision (ICD-10). Length of stay was calculated as the difference between the time the patient was screened in the emergency department and the final hospital discharge, including exits to hospitalization, primary health care centers for follow-up by the family medicine physician, or home stays.

Statistical Analysis

Statistical analysis of the data grouped patients based on gender, age, primary diagnosis, and Hb and CRP concentrations. Furthermore, the most frequently diagnosed ICD-10 groups were explored as a function of age and Hb and CRP concentrations. This analysis was performed with SPSS (IBM SPSS Statistics for Mac, version 26, IBM Corp., NY, USA).

Continuous variables were evaluated for normal distribution with histograms (skewness and kurtosis) and described using the mean and standard deviation (SD). Skewed continuous variables were reported with median and interquartile range (IQR). Categorical variables were described using frequency and percentage. The association between Hb and CRP was evaluated with Spearman's correlation coefficient. The comparison between CRP in different categorical groups was accomplished with Kruskal–Wallis and Mann–Whitney tests. Receiver operating characteristic (ROC) analysis was used to evaluate the predictive ability of both Hb and CRP for the hospitalization discharge. The area under the ROC curve (AUC) was calculated together with 95% confidence intervals and assessed accordingly: 0.5–0.6: fail, 0.6–0.7: poor, 0.7–0.8: fair, 0.8–0.9: good, and 0.9–1.0: excellent. The Youden index (maximum sensitivity and specificity) was applied to determine the best cut-off value for the significant studied parameters in ROC analysis and evaluated with sensitivity and specificity. Statistical significance was determined at $p < 0.05$.

RESULTS

Participants and Clinical Condition

The study cohort (**Table 1**) included data of 125 individuals (50.4% males and 49.6% females) with a median age of 69 years (IQR 49–81 years). Anemia was defined according to the WHO threshold, with values of Hb of less than 12 g/dL for women and less than 13 g/dL for men. An active inflammatory process was considered when CRP was above 5 mg/L. Diagnostic groups of diseases (ICD-10) with less than 5% of cases, such as hematological, metabolic, as well as

mental conditions, and conditions of the ear, skin, musculoskeletal, injury or poisoning, and external causes of morbidity and mortality, were not included for further analysis.

Table 1. Descriptive frequencies of the studied cohort.

Characteristics		Overall *n* = 125 (%)
Gender	Female	62 (49.6%)
	Male	63 (50.4%)
Age Group	< 65	52 (41.6%)
	≥ 65	73 (58.4%)
Clinical condition	Anemia only	11 (8.8%)
	Inflammation only	31 (24.8%)
	Anemia and inflammation	45 (36.0%)
	No anemia either inflammation	38 (30.4%)
Diagnosis	I—Infectious diseases	12 (9.6%)
	II—Neoplasms (active cancer)	7 (5.6%)
	III—Hematological diseases	1 (0.8%)
	IV—Endocrine, nutritional, and metabolic disorders	1 (0.8%)
	V—Mental and behavioral disorders	6 (4.8%)
	VI—Diseases of the nervous system	7 (5.6%)
	VII—Diseases of the ear and mastoid process	3 (2.4%)
	IX—Diseases of the circulatory system	20 (16.0%)
	X—Diseases of the respiratory system	16 (12.8%)
	XI—Diseases of the digestive system	15 (12.0%)
	XII—Diseases of the skin and subcutaneous tissue	3 (2.4%)
	XIII—Diseases of the musculoskeletal system	6 (4.8%)
	XIV—Diseases of the genitourinary system	23 (18.4%)
	XIX—Injury and poisoning	2 (1.6%)
	XX—External causes of morbidity and mortality	3 (2.4%)
Discharge	Hospitalization	30 (24.0%)
	Primary health care	66 (52.8%)
	Home	29 (23.2%)
Length of stay	< 6	66 (52.8%)
	≥ 6	59 (47.2%)

Data represent number and (frequency). Discharge includes exit categories of the emergency department and length of stay in hours.

According to the criteria indicated above for the definition of each clinical condition, total anemia (irrespective of CRP levels) and total inflammation (irrespective of Hb levels) displayed a prevalence of 44.8% and 63.2%, respectively. Anemia was more present in males (57.1%) rather than females (42.9%) and was five times higher in patients aged 65 or older than in younger subjects (84.0% against 16.0%). Considering the diagnostic groups, the prevalence of anemia decreased from neoplasms (100%) to diseases of digestive (53.3%), genitourinary (43.5%), nervous (42.9%), infectious (41.6%), respiratory (30%), and circulatory (14.3%) systems. Only 7.2% of anemic patients were discharged to have health care at home.

Active inflammation was found in 63.2% of the cohort, balanced between gender but with approximately twice the rate in senior patients (age > 65 with 70.0%). Patients with CRP ≥ 5 mg/L were found in more than 50.0% for each diagnostic group of diseases. Respiratory diseases had 81.2% of patients with elevated CRP, followed by neoplasms (71.4%), infectious (66.6%), and cardiovascular and digestive diseases (both with 60.0%), diseases of the genitourinary system (65.0%), and diseases of the nervous system (57.1%).

When considering anemia only (no alterations in CRP levels), the prevalence was 8.8%. It affected more males (54.5%) than females (45.5%) and was nearly three times higher in patients aged 65 or older than in younger subjects (73% vs. 27%). When considering the diagnostic groups, the prevalence of anemia was higher in neoplasms (28.6%), infectious disorders (16.7%), digestive diseases (13.3%), and genitourinary diseases (13.0%) (**Table 2**). In our small cohort, patients with diseases of the nervous and respiratory systems were not anemic. Only 9.1% of anemic patients were hospitalized, while 18.2% were discharged to follow health care at home, and the remaining patients were signaled to be monitored at a primary health care center.

Table 2. Distribution of clinical conditions (anemia only, inflammation only, both anemia and inflammation, and neither of these) across each group of the cohort.

Clinical Condition		Anemia Only	Inflammation Only	Anemia and Inflammation	No Anemia or Inflammation
Gender	Female	5 (8.1%)	19 (30.6%)	19 (30.6%)	19 (30.6%)
	Male	6 (9.5%)	12 (19.0%)	26 (41.3%)	19 (30.2%)
Age Group	<65	3 (5.8%)	17 (32.7%)	6 (11.5%)	26 (50.0%)
	≥65	8 (11.0%)	14 (19.2%)	39 (53.4%)	12 (16.4%)
Diagnosis	I	2 (16.7%)	5 (41.7%)	3 (25.0%)	2 (16.7%)
	II	2 (28.6%)	0 (0.0%)	5 (71.4%)	0 (0.0%)
	VI	0 (0.0%)	2 (28.6%)	2 (28.6%)	3 (42.9%)
	IX	1 (5.0%)	7 (35.0%)	5 (25.0%)	7 (35.0%)
	X	0 (0.0%)	3 (18.8%)	10 (62.5%)	3 (18.8%)
	XI	2 (13.3%)	3 (20.0%)	6 (40.0%)	4 (26.7%)
	XIV	3 (13.0%)	8 (34.8%)	7 (30.4%)	5 (18.8%)
Discharge	Hospitalization	1 (3.3%)	7 (23.3%)	20 (66.7%)	2 (6.7%)
	Primary health care	8 (12.1%)	19 (28.8%)	18 (27.3%)	21 (31.8%)
	Home	2 (6.9%)	5 (17.2%)	7 (24.1%)	15 (51.7%)
Length of stay	<6	6 (9.1%)	14 (21.2%)	20 (30.3%)	26 (39.4%)
	≥6	5 (8.5%)	17 (28.8%)	25 (42.4%)	12 (20.3%)

Data represent number and (frequency). Discharge includes exit categories of the emergency department and length of stay in hours.

Active inflammation only (no alterations in Hb levels) was found in 24.8% of the samples, evenly distributed between age groups but with approximately 1.5 times more cases in female

patients (61.3% against 38.7%). Patients with CRP ≥ 5 mg/L were found in all diagnostic groups except for neoplasms. Elevated CRP levels were detected in 41.7% of patients with infectious diseases, followed by circulatory (35.0%), genitourinary (34.8%), digestive (20.0%), and respiratory diseases (18.8%), and diseases of the nervous system (28.6%). In this cohort, patients with neoplasms did not display CRP levels above 5 mg/L. In opposition to anemia, an active inflammatory condition was more prevalent in patients with a high length of stay (>6 h) in the emergency department (54.8% against 45.2%).

Patients with both clinical conditions (anemia and active inflammatory process) represented nearly 36% of the sample population. When considering gender and age distribution, respectively, the existence of both anemia and an active inflammatory process was more prevalent in males (57.8%) and patients aged ≥65 years (86.7%). Neoplasms were the most prevalent (71.4%), followed by diseases of the respiratory system (62.5%), digestive system (40.0%), genitourinary system (30.4%), and infectious diseases and diseases of the cardiovascular system (25.0%). Interestingly, patients without any of these clinical conditions corresponded to the second most prevalent group of the cohort (30.4%); in this group, only two patients were hospitalized, and the most common diagnostic was related to diseases of the circulatory system (18.4%).

The median age of the study population also represented differences by diagnosis group: patients under the age of 65 years had a higher prevalence of diseases of the nervous system (VI, 9.6% vs. 2.7%), while the aged group (≥65 years) had a higher proportion of diseases of the genitourinary system (XIV, 20.5% vs. 15.4%).

Correlation between Hemoglobin and C-Reactive Protein

Given that the level of Hb was significantly lower in patients with elevated concentration of CRP when compared to the normal group (11.5 ± 2.2 vs. 13.2 ± 1.8, $p = 0.001$), we next evaluated how strongly these metabolites were correlated.

When considering the entire cohort, we found a significantly negative correlation between the Hb concentration and the CRP levels ($rs = -0.42$, $r^2 = 0.15$, $p < 0.001$) (**Figure 1**). This association was maintained after adjustments for both gender and age groups. Interestingly, the correlation was stronger in males than females ($rs = -0.56$ vs. $rs = -0.37$, $p < 0.05$) and in the aged group ($rs = -0.43$ vs. $rs = -0.30$, $p < 0.05$).

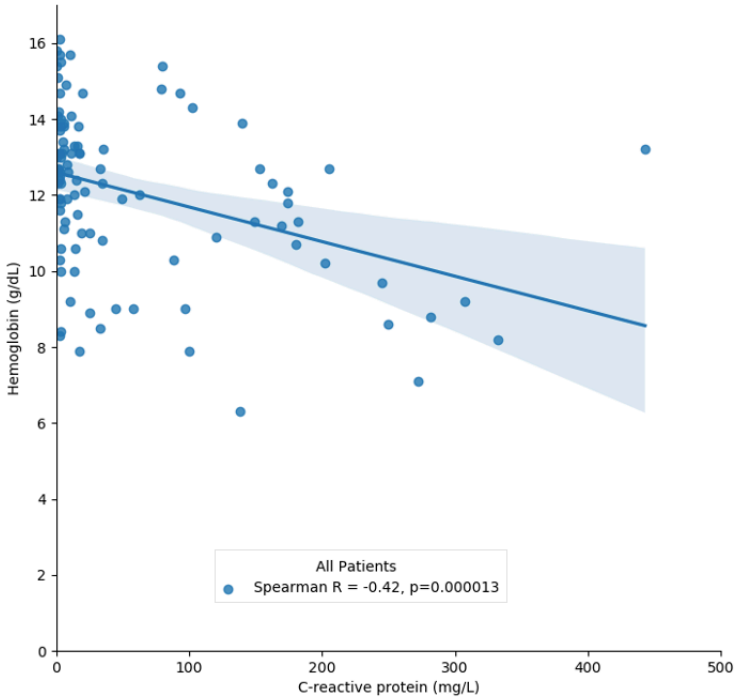


Figure 1. Correlation between hemoglobin and C-reactive protein concentrations in the entire cohort.

In patients grouped by diagnostics, the Spearman's correlation was significant for patients with diseases of the nervous system ($rs = -0.82$, $p < 0.05$) and respiratory system ($rs = -0.77$, $p < 0.05$). Patients with digestive system diseases displayed a correlation of –0.509 that nevertheless failed to reach statistical significance ($p = 0.053$).

When considering the different types of discharge and lengths of stay, a significant correlation between Hb and CRP levels was also verified. However, this was not confirmed when the cohort was analyzed by clinical condition. In fact, only patients with both anemia and inflammation presented a significant association between Hb and CRP, particularly in hospitalized patients and in patients diagnosed with respiratory diseases (**Table 3**).

Table 3. Correlation between Hb and CRP in the studied subgroups, according to main characteristics and clinical conditions.

Clinical Condition		Anemia Only	Inflammation Only	Anemia and Inflammation	No Anemia or Inflammation	All
Gender	Female	0.44 (0.450)	–0.26 (0.284)	–0.37 (0.112)	0.04 (0.867)	–0.37 (0.003)
	Male	–0.58 (0.222)	0.48 (0.108)	–0.26 (0.206)	–0.14 (0.567)	–0.56 (<0.001)
Age group	<65	0.50 (0.667)	–0.08 (0.736)	–0.03 (0.957)	–0.17 (0.402)	–0.30 (0.03)

	≥65	−0.36 (0.386)	0.01 (0.976)	−0.28 (0.087)	0.36 (0.252)	−0.43 (<0.001)
Diagnosis	I	-	−0.15 (0.805)	−0.50 (0.667)	-	0.28 (0.931)
	II	-	-	−0.46 (0.434)	-	−0.05 (0.908)
	VI	-	-	-	-	−0.82 (0.023)
	IX	-	0.05 (0.908)	−0.80 (0.104)	0.07 (0.878)	−0.28 (0.228)
	X	-	-	−0.69 (0.029)	-	−0.77 (0.001)
	XI	-	-	−0.37 (0.468)	−0.32 (0.684)	−0.51 (0.053)
	XIV	-	0.24 (0.570)	−0.18 (0.702)	0.56 (0.322)	−0.22 (0.325)
Discharge	Hospitalization	-	0.02 (0.969)	−0.45 (0.047)	-	−0.38 (0.038)
	Primary health care	−0.11 (0.797)	−0.08 (0.745)	−0.21 (0.413)	−0.08 (0.726)	−0.41 (0.001)
	Home	-	−0.70 (0.188)	−0.18 (0.699)	0.058 (0.837)	−0.47 (0.009)
Length of stay	<6	0.13 (0.802)	−0.34 (0.242)	−0.33 (0.156)	−0.12 (0.569)	−0.44 (<0.001)
	≥6	−0.80 (0.588)	0.08 (0.768)	−0.29 (0.154)	0.11 (0.741)	−0.48 (<0.001)
	All	−0.18 (0.588)	−0.10 (0.598)	−0.30 (0.044)	−0.05 (0.749)	−0.42 (<0.001)

Data represent the Spearman's rank correlation coefficients (rs) with the respective level of significance (p).

Hemoglobin and CRP Distribution

We next analyzed the distribution of the concentration of Hb. The study population displayed a normal distribution of Hb levels with an average of 12.2 ± 2.2 g/dL and a median value of 12.6 (10.8–13.9) g/dL. It did not differ between genders [$t(113.6) = -0.839$, $p = 0.403$], but varied significantly between age groups [$t(123) = 4.558$, $p = 0.001$]. When considering the diagnostic groups (ICD-10), the concentration of Hb was similar across pathologies, except for neoplasms [$t(123) = 2.105$, $p = 0.037$] with patients with oncologic conditions presenting lower Hb concentrations (Group II) (**Figure 2a**).

CRP concentration in the study population displayed a left-skewed distribution and a median value of 10.8 (IQR 2.8–63.7 mg/L). It was not statistically different between genders [$H(1) = 0.098$, $p = 0.754$] but differed between age groups ($U = 1202.5$, $z = -3.564$, $p < 0.001$). Furthermore, the median levels of CRP were only significantly different for patients with diseases of the respiratory system ($U = 578.5$, $z = -2.171$, $p = 0.030$).

Discharge groups exhibited a similar behavior regarding the comparison between the levels of Hb and CRP. Patients that were admitted for hospitalization presented lower Hb [$F(2,18) = 4.338$, $p = 0.016$] and higher CRP levels [$H(2) = 26.31$, $p = 0.001$] than patients oriented for primary health care or home care. No statistical differences were observed in the levels of both blood metabolites on the two latter groups of patients (**Figure 2b**). Therefore, this suggests that

Hb and CRP concentrations at admission might act as discharge predictors for hospitalization when a patient visits the emergency department.

Additionally, the concentration of CRP was also different according to the length of stay in the emergency department [H(1) = 5.298, p = 0.021], with a cut-off point of 6 h. Hb levels did not differ between the different length of stay groups [t(123) = 1.431, p = 0.155].

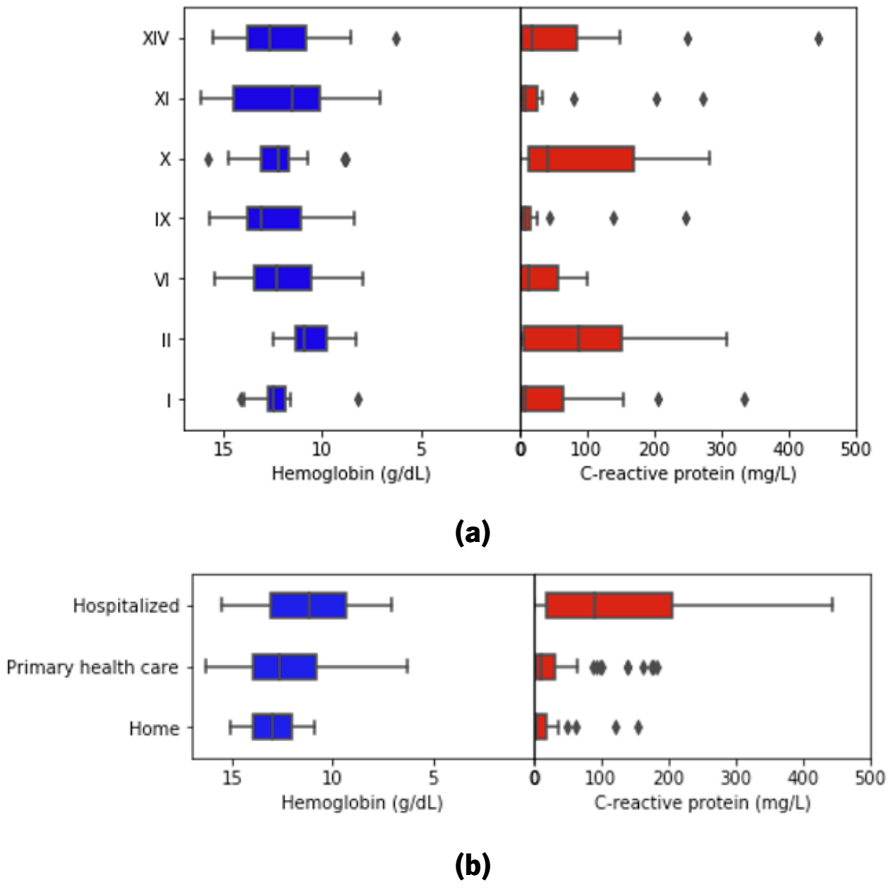
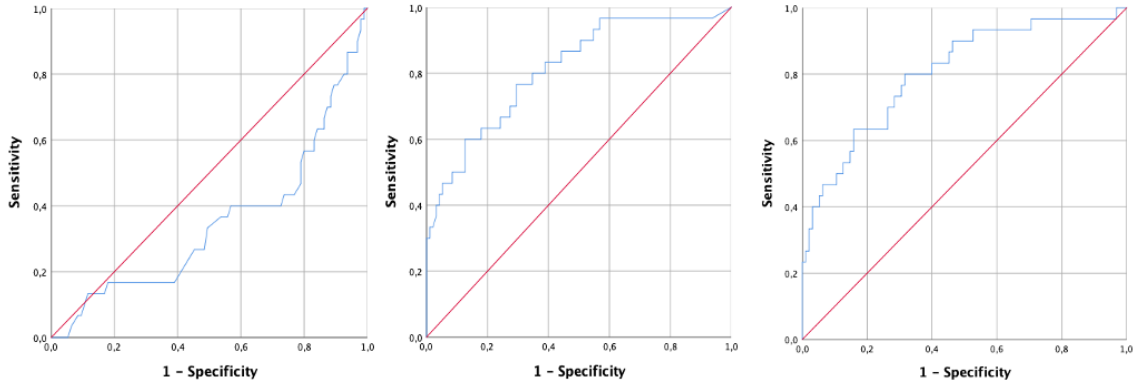


Figure 2. Boxplots show the distribution of hemoglobin and c-reactive protein according to diagnostics group **(a)** and type of discharge **(b)**. Rectangles represent the second and third quartiles with a vertical line inside to indicate the median value. The lower and upper quartiles are shown as horizontal lines on either side of the rectangle and outliers are plotted outside of it, individually.

ROC Analysis

Considering the significant differences in the Hb and CRP levels between discharge groups, ROC analysis was performed to evaluate the area under the curve (AUC) for the classification metrics regarding the destination of patients after emergency room discharge. The AUC was found to be statistically significant across the three analyses but decreased from CRP (0.811) to CRP/Hb (0.804) and Hb (0.347) (**Figure 3**).



(a) **(b)** **(c)**

Figure 3. ROC curves of hemoglobin **(a)**, C-reactive protein **(b)**, and C-reactive protein-to-hemoglobin ratio **(c)**.

The sensitivity of the CRP/Hb score (cut-off = 1.32) in the discharge for hospitalization was 80.0%, and the specificity was 68.4% for patients arriving at the emergency department. The analysis performed for Hb revealed that it did not have predictive value for discharge destination from the ER (AUC < 0.5) (**Table 4**).

Table 4. ROC curve results for blood parameters in the cohort.

Blood Parameters	Cut Off	AUC (95% CI)	*p*	Sensitivity	Specificity
CRP	67.55	0.811 (0.719–0.902)	<0.001	60.0%	87.4%
CRP/Hb	1.32	0.804 (0.711–0.898)	<0.001	80.0%	68.4%

AUC: area under the curve, CI: confidence interval.

DISCUSSION

Data regarding the relationship between serum CRP and Hb concentration are limited. To the best of our knowledge, their association was not determined in heterogeneous patients admitted to emergency settings. This retrospective study performed an exploratory evaluation of patients admitted to an emergency department considering the concentration of Hb and CRP as markers of anemia and inflammation, respectively. The association between both metabolites was studied globally and sub-segmented according to their demographics, diagnostics ICD-10 classification, type of discharge, and length of stay in the emergency department.

The results of this study indicate that Hb and CRP blood concentrations strongly associate. Average Hb concentration was significantly lower in patients with a high concentration of CRP. Serum CRP negatively correlated with Hb concentration, and its relationship remained significant

after adjustment for gender, age group, specific group of pathologies, discharge, and length of stay in the emergency department. Interestingly, a significant correlation between both blood parameters was verified in the 125 patients studied, and the linear regression model applied showed that the levels of CRP explained 15% of Hb variations. Considering each patient's multitude of symptoms and pathological background, the correlation herein described is of clinical significance. Reports from studies exploring the association between these blood metabolites found differences according to the cohort type. Ziv-Baran et al. (2020) studied apparently healthy individuals undergoing a routine health examination. They found a general significant, albeit weak, relationship between these blood metabolites, namely an inverse correlation between serum CRP and Hb ($r^2 = -0.025$, $p < 0.001$). Furthermore, a correlation between the blood concentration of these metabolites was found across different age groups and in males [16]. However, this study lacked detailed clinical characterization of the cohort.

In the present study, we categorized patients according to the diagnostics made on emergency department admission and conducted linear associations between the metabolites studied. Hb levels displayed a normal distribution, without significant differences across ICD groups, except for patients with active cancer (neoplasms). Furthermore, the prevalence of anemia (levels of Hb below 13 mg/dL or 12 mg/dL in males and females, respectively) in the cohort agrees with the literature [10], with an average of 40%. Additionally, and in accordance with the reported by others [20], we found that anemia is more prevalent in males and in the elderly group (>65 years). Different from Hb, CRP concentration exhibited a left-skewed distribution that was only significantly higher for aged patients and those diagnosed with diseases of the respiratory system.

Considering the type of discharge, we found that both anemia and inflammation frequencies increased from home to primary health care and hospitalization. Hospitalized patients displayed significantly different levels of both Hb and CRP, with symmetric (negative) variation (correlation). This correlation was still significant for 67% of patients with both anemia and inflammation, suggesting that the increase in CRP explains nearly 45% of the decrease in Hb. These findings follow the current knowledge regarding the impact of anemia of inflammation in determining long-term hospitalization and increased mortality outcomes [19].

The evaluation of Hb and CRP as predictors of discharge from the emergency unit to hospitalization showed relevant predictive efficiency for CRP and the CRP/Hb ratio. This test confirms the importance of the CRP concentration as a predictor of hospitalization. We suggest that the cut-off value of 1.32 for the CRP/Hb ratio (with a sensitivity of 80.0%) might be considered

as a relevant biochemical index that could provide support at admission to emergency units in discriminating between patients with a need for hospitalization.

This retrospective observational study presents limitations such as the reduced sample size, which limited in-depth statistical associations between gender, age, and disease groups and clinical outcomes regarding the longitudinal behavior of blood metabolites. A collection of anemia-related biomarkers, such as ferritin and transferrin saturation, would also be relevant to differentiate patients according to the type of anemia. Correlations between studied metabolites could also be more personalized and valuable in that context.

The strength of this study is justified by the heterogeneity of the data, which allowed for an exploration according to blood indexes and diagnostic made on the ED. Moreover, significant associations were found for a non-specific cohort and raised for two specific groups of pathologies, which could enable focused research for the diseases under these groups of pathologies.

CONCLUSIONS

This study confirmed how anemia and inflammation are prevalent in emergency scenarios, especially for elderly patients with chronic conditions or comorbidities that endorse extra challenges on clinical outcomes. We verified a significant linear correlation between Hb and CRP on the entire cohort of patients admitted to the emergency department; this association was kept significant after age, gender, and pathology group adjustments. Our data provide support for the use of CRP and the CRP/Hb ratio as predictors of hospitalization following ER discharge.

REFERENCES

1. Chudasama, Y.V.; Khunti, K.; Davies, M.J. Clustering of comorbidities. Futur. Healthc. J. 2021, 8, e224–e229. [CrossRef] [PubMed]
2. James, N.J.; Hussain, R.; Moonie, A.; Richardson, D.; Waring, W.S. Patterns of admissions in an acute medical unit: Priorities for service development and education. Acute Med. 2012, 11, 74–80. [CrossRef] [PubMed]
3. Badrick, T. Evidence-Based Laboratory Medicine. Clin. Biochem. Rev. 2013, 34, 43–46. [PubMed]
4. Hallworth, M.J. The '70% claim': What is the evidence base? Ann. Clin. Biochem. 2011, 48, 487–488. [CrossRef] [PubMed]
5. McLean, E.; Cogswell, M.; Egli, I.; Woidyla, D.; De Benoist, B. Worldwide prevalence of anaemia, WHO Vitamin and Mineral Nutrition Information System, 1993–2005. Public Health Nutr. 2009, 12, 444–454. [CrossRef] [PubMed]
6. Cascio, M.J.; DeLoughery, T.G. Anemia: Evaluation and Diagnostic Tests. Med. Clin. N. Am. 2017, 101, 263–284. [CrossRef] [PubMed]
7. Guenter, G.; Lawrence, T. Anemia of chronic disease. N. Engl. J. Med. 2005, 352, 1011–1023.
8. Weiss, G.; Ganz, T.; Goodnough, L.T. Anemia of inflammation. Blood 2019, 133, 40–50. [CrossRef] [PubMed]
9. Madu, A.J.; Ughasoro, M.D. Anaemia of Chronic Disease: An In-Depth Review. Med Princ. Pract. 2017, 26, 1–9. [CrossRef] [PubMed]
10. Röhrig, G.; Klossok, W.; Becker, I.; Benzing, T.; Schulz, R. Prevalence of anemia among elderly patients in an emergency room setting. Eur. Geriatr. Med. 2014, 5, 3–7. [CrossRef]
11. Lippi, G.; Bonelli, P.; Meschi, T.; Nouvenne, A.; Borghi, L.; Cervellin, G. Prevalence of anemia and critical anemia in elderly patients admitted to a large urban emergency department. Eur. Geriatr. Med. 2014, 5, 214–215. [CrossRef]
12. Dyer, E.M.; Waterfield, T.; Baynes, H. How to use C-reactive protein. Arch. Dis. Child. Educ. Pract. Ed. 2019, 104, 150–153. [CrossRef] [PubMed]
13. Marsik, C.; Kazemi-Shirazi, L.; Schickbauer, T.; Winkler, S.; Joukhadar, C.; Wagner, O.F.; Endler, G. C-Reactive Protein and All-Cause Mortality in a Large Hospital-Based Cohort. Clin. Chem. 2008, 54, 343–349. [CrossRef] [PubMed]
14. McWilliam, S.; Riordan, A. How to use: C-reactive protein. Arch. Dis. Child. Educ. Pract. Ed. 2010, 95, 55–58. [CrossRef] [PubMed]
15. Luan, Y.-Y.; Yao, Y.-M. The Clinical Significance and Potential Role of C-Reactive Protein in Chronic Inflammatory and Neurodegenerative Diseases. Front. Immunol. 2018, 9, 1302. [CrossRef] [PubMed]
16. Ziv-Baran, T.; Wasserman, A.; Goldiner, I.; Stark, M.; Shenhar-Tsarfaty, S.; Shapira, I.; Zeltser, D.; Mailis, I.; Berliner, S.; Rogowski, O. The association between C-reactive protein and common blood tests in apparently healthy individuals undergoing a routine health examination. Clin. Chim. Acta 2020, 501, 33–41. [CrossRef] [PubMed]
17. Heidari, B.; Fazli, M.R.; Misaeid, M.A.G.; Heidari, P.; Hakimi, N.; Zeraati, A.A. A linear relationship between serum high-sensitive C-reactive protein and hemoglobin in hemodialysis patients. Clin. Exp. Nephrol. 2014, 19, 725–731. [CrossRef] [PubMed]
18. Moeller, M.; Pink, C.; Endlich, N.; Endlich, K.; Grabe, H.-J.; Völzke, H.; Dörr, M.; Nauck, M.; Lerch, M.M.; Köhling, R.; et al. Mortality is associated with inflammation, anemia, specific diseases and treatments, and molecular markers. PLoS ONE 2017, 12, e0175909. [CrossRef] [PubMed]
19. Balasubramanian, A.; Henderson, R.J.; Putcha, N.; Fawzy, A.; Raju, S.; Hansel, N.N.; MacIntyre, N.R.; Jensen, R.L.; Kinney, G.L.; Stringer, W.W.; et al. Haemoglobin as a biomarker for clinical outcomes in chronic obstructive pulmonary disease. ERJ Open Res. 2021, 7, 3. [CrossRef] [PubMed]
20. Gaskell, H.; Derry, S.; Moore, R.A.; McQuay, H.J. Prevalence of anaemia in older persons: Systematic review. BMC Geriatr. 2008, 8, 1. [CrossRef] [PubMed]

CHAPTER 2.2

Artificial intelligence in routine blood tests: A review

The work presented in this chapter has been submitted and is currently under revision for publication in the Frontiers in Digital Health, section Personalized Medicine.

Artificial intelligence in routine blood tests

Miguel A. Santos-Silva[1,2], Nuno Sousa[1,2,3], João Carlos Sousa[1,2,*]

[1] Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho, Campus de Gualtar, 4750-057 Braga, Portugal

[2] ICVS/3B's – PT Government Associate Laboratory, Braga/Guimarães, Portugal

[3] Clinical Academic Center-Braga (2CA), 4710-243 Braga, Portugal

[*] Correspondence: jcsousa@med.uminho.pt

ABSTRACT

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 that can be integrated and automated at scale, allowing for in-depth clinical inference and cost-effectiveness when 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 still lacks harmonization 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.

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[1]. Automatic-integrated systems mimicking human cognitive functions and deploying supervised applications replicating previous tasks with higher accuracy levels transformed time-consuming and high-volume repetitive tasks[2]. 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[3]. The COVID-19 pandemic is the paramount example of how AI applications enable new screening tools and achieve early diagnosis by measuring disease severity[4], progression[5], and mortality prediction[6], 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% of specificity[7]. 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[8]. Since then, progress was slow and kept stationary until 2012, when an ImageNet-DL-Algorithm triggered significant attention for the technology, with high-accuracy performance classification metrics that disrupted current state-of-the-art[9]. Currently, AI defines as a computer science subdivision that aims to automatically understand and create intelligent systems based on high amounts of data[10]. 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[11].

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. Concepts such as precision medicine, ranging from diagnostics to prognostics and therapeutics with connected care, are under development[12]. 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[13]. 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[14]. Also, legal and regulatory matters are under development, which will be crucial to regulate how to build AI algorithms and evaluate continuous learning.

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 studied pathologies based on the included general health parameters (summarized according to their function, and associated causes of variation), their inherent classification performance, and principal findings associated to 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 that combined AI and routine blood tests to perform disease diagnosis 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 modeling, fuzzy logic, knowledge representation, case-based reasoning, text mining, visual analytics, and optimization, among others[15]. We will next briefly explain these learning strategies.

ML is a pattern recognition method that performs automatic detection of 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 hypothesis[16]. Supported and leveraged by the continuous increase of data availability (big data), ML has become the preferred framework for deploying AI applications. 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[17]. Data is considered mandatory for the model's development, and its 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)[1].

ML algorithms such as gaussian naïve Bayes (GNB), k-nearest neighbors (KNN), support vector machines (SVM), decision trees (DT), linear regression (LR), or ensemble 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 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, or established in the form of a regression problem that evaluates continuous variables and predicts, for example, the numeric value of hemoglobin[18]. However, suppose the true state of the data is not known. In that case, learning might be conducted in an unsupervised fashion where algorithms infer underlying patterns in unlabeled data to find sub-clusters of the original data, identify outliers, or produce low-dimensional representations of the data. This way, it could be possible to recognize new associations that were not perceptible. In the example above, the algorithm would potentially 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 answer found by the algorithm is correct or not; it is like supervised learning, but it 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[19], 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.

Table 1: A comprehensive list of supervised ML algorithms for classification, according to the desired learning strategy[20].

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.
	Logistic 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 tree (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.
	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.

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 create accurate models from raw data. In medical diagnosis, this method has an application on 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[21]. Deep learning requires minimal human intervention (except for sample labeling) but large amounts of curated data sets. Additionally, computational power is also relevant to conduct these tasks. Some researchers estimated that a life cycle to train several large AI models could emit nearly five times the CO2 an average American car[22]. These challenges conducted to the study and development of a new learning approach based on sparse modeling (SM) 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[23]. 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) presents itself as a new learning paradigm aiming to correct some 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 the training of 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 McMahan et al[24]. This workflow includes the distribution of the global model on independent 'clients' that train the model in its data and sends the adjusted local model to the global server to perform the trained models' aggregation; this cycle repeats until the global model converges. Dayan I. et al. [25] implemented an FL approach for the Covid-19 prognosis of oxygen supplies on symptomatic patients using vital signs, laboratory data, and chest-X-rays inputs 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[26], 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.

Box.1: Glossary of key terms.

AUROC	The area under the receiving operating characteristic curve computed with the true positive versus 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.

Blood and routine blood tests

Blood is the only fluid tissue present in the human body. Typically, an average adult has 6 to 7 liters 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[27]. 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 to 10 mL of venous blood.

In hematology, cell blood count (CBC) is the most performed exam and 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)[28].

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[29]. The serum refers to plasma without clotting factors, i.e., fibrinogen, and is commonly used for chemistry testing and coagulation studies[30]. A general health blood parameters panel typically includes CBC with differential, comprehensive metabolic and lipid panels, uric acid, GGT, and TSH[31].

Table 2: General health panel required in routine blood examinations.

	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
	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 metabolization	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.

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. Studies that included other biofluids parameters or the ones that reported reviews, systematic reviews, meta-analyses, protocols, commentaries, or book chapters were not included for analysis.

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); tables 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, 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[32,33]. 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[34]. 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[35]. 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. pneumonie* (n=83) and others (n=500)[36]. 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)[37]. 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)[38]. The evaluation of the model was made available, differentiating this study from the previous and enabling a prospective evaluation 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[35]	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[36]	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[33]	Surgical-site infection (post-operative)	1005 (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[39]	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 72h after admission. External validation: NA Clinical deployment: NA
Kocbek et al., 2019[32]	Surgical site infection	1137 (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
Morang'a et al., 2020[40]	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[34]	Dengue	4,894 (2,942)	Age, Temperature, WBC	L: supervised, classification FS: NA C: DNN (0.858)	For all three models, pre-peak sensitivities (<35 weeks) were higher than 90%. External validation: NA

				V: 10-fold cv	Clinical deployment: NA
Mooney et al., 2021[37]	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[38]	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, *HLTV-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 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 performance (AUC=0.81, AUC=0.87, and AUC=0.85). This consistency endorses the relevance of its findings. Noteworthily, 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)[41], Hornbrook et al. identified sub-regions of colorectal cancer that were better diagnosed, namely the cecum and the ascending colon[42].

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[43]. 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)[44]. 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 two years for patients with lymphocytic leukemia with a dataset of 109 patients[45]. 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[46]. 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.

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[41]	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[42]	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[43]	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., 2021[46]	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[44]	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[45]	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.

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

The approach to blood diseases was fundamentally through the combination of 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[47]. The accuracy results obtained were highly similar between the studied strategies (accuracies≥0.98), and comparison with previous studies[48] 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[49]. 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)[50]. 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.

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 et al., 2012[47]	Women's IDA	Training: 2000(NA)	RBC, Hb, HCT, MCV, MCH, MCHC	L: supervised, classification FS: NA	ANN and medical diagnosis achieved comparable results.

		Test: 600(122)		C: ANN (0.99) V: test-set	ANN training with several strategies (FFN, CFN, DDN, TDN, LVQ, PNN) produced similar results (accuracy ≥ 0.98). External validation: NA Clinical deployment: NA
Magen et al., 2013[49]	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[50]	β-thalassemia and IDA distinction	342 (152)	RBC, HCT, MCV, MCH, MCHC, RDW	L: supervised, classification FS: NA C: several week 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, *HTC* 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.

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). 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[51]. 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 by applying a shallow neural network based on historical continuous blood glucose monitoring[52]. The results improved the current condition through a more extended prediction (t=240min vs. 120min) with lower error (RMSE,60min=28mg/dL vs. 43mg/dL). Bernardini et al. initially featured disease complications with an evaluation of clinical factors associated with insulin resistance[53]. 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≤70mg/dL) was approached by Mathioudakis et al. using a stochastic gradient boosting ML model in an extensive data study (n=1 612,425)[54]. 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 24h after each blood glucose measurement.

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[53]	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[51]	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[55]	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[54]	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[52]	Blood glucose prediction in T1DM	24	Historic continuous glucose monitoring	L: supervised, regression FS: NA C: shallow neural network (RMSE): t=60 (28±4), t=90 (33±4), t=120 (38±6), t=180 (40±8), t=240 (43±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.

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 7**). 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[56], Hochman et al. aimed to build a low-cost diagnostic tool to perform diagnosis based on blood data[57]. 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 confounders 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[57]. 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). 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.

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[56]	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., 2020[57]	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.

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 reliability of the model, 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[58]. 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[59]. 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 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.

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., 2020[60]	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[58]	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., 2021[59]	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* Logistic Regression, *XGBoost* Extreme Gradient Boosting, *rRRP* Resting Rate Pressure

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[61]. 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 (r2=0.55) and non-smokers (r2=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 to 24 features), 0.54 to 0.58 (20 to 21 features), and 0.55 to 0.60 (18 to 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[62]. 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.

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[61]	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[62]	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.

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[63]. Regarding inflammation, Fialoke et al. studied the discrimination between non-alcoholic steatohepatitis (NASH) and simple steatosis in NAFL[64]. Since NASH is underdiagnosed due to the lack of patient symptoms and relevant biomarkers (high values of AST and ALT), the authors trained four 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 NAFDL in a cross-sectional study of 10,030 individuals with a prevalence of 24%[65]. Four techniques were used for feature selection, and eleven 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 & cirrhosis stages) by studying seven routine blood tests enhanced by a multilayered perceptron and a naïve-Bayes algorithm[66]. 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[67]. 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 1024 (AUC=8922) 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 were not approached yet.

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[66]	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[64]	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[65]	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[67]	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[68]	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* Logistic 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.

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 60mL/min/1.73m^2 [69].The need to create screening procedures that perform early diagnosis motivated several studies for routine analysis, mostly performed in combination between blood and urine. 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, 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. 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.

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
Mahfuz et al., 2021[70]	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* Logistic 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.

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[71] and 6.37M deaths[72]. 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)[73]. 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%[74]. 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[75]. RT-PCR tests require laboratory specialists and infrastructure and produce a 15% false-positive rate on 48 to 72h 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[76]. The criteria for approval and commercialization of RDTs are based on a sensitivity superior to 80% and specificity above 98%[77], tested on a prospective cohort study involving less than 30 persons infected with SARS-CoV-2 and 30 persons without the infection[78]. 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 (from 36% to 82%) and specificity (from 98% to 100%) when these RDTs are tested in asymptomatics[79]. 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<1000) 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)[80]. Brinati et al. also used logistic regression but compared the classification performance with a random forest classifier, using 14 features (**Table 12**)[81]. 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)[82]. 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 Banarjee et al. in comparison with a random forest and a lasso-elastic-net regularized generalized linear model (fitting a logistic regression)[83]. 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[84]. 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)[85]. 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 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>1000) showed a similar increase in classification metrics in concordance with the addition of blood features. In a cohort of 1537 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[86]. Cabitza et al. described a novel methodology comprising cardinality and similarity as metrics of model's reliability in external validation settings[87]. 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, the comparison performance between all the algorithms exhibited DNN with the higher classification metrics (**Table 12**)[88]. Interestingly, the previous studies of Brinati et al.[81], and Cabitza et al.[87], were also compared with DNN surpassing in the first dataset (AUC=0.92 vs. AUC=0.84, from Brinati et al.[81]) and the second dataset (AUC=0.93 vs. AUC=0.84, from Cabitza et al.[87]), 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[89]. 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[90]. 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)[91]. 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%)[92]. 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 and invasive mechanical ventilation and ICU (multipurpose algorithms)[93]. 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%. Higher predictive performance was accomplished in the study of

Karthikeyan et al., through the application of 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.

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[81]	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[94]	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
Banarjee et al., 2020[83]	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[95]	Covid-19 diagnosis	12,183 (2182)	EO, CA, AST, WBC, BA, RDW, RBC, ALB, TB, MCV, MCH, SO, HCO3, 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[86]	Covid-19 diagnosis	1537 (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[82]	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[85]	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
Cabitza et al., 2021[87]	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., 2021[96]	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[97]	Covid-19 diagnosis	603 (83)	Age, HTC, HGB, PLT, RBC, LY, MCHC, WBC, BA, MCH, EO,	L: supervised, classification FS: recursive feature elimination;	Dynamic ensemble selection application on imbalanced data; External validation: NA

			MCV, MO, RDW, G, CRP	C: Dynamic Ensemble Selection (0.99); V: 70-30 and 60-40 training-test and 5-fold-cv	Clinical deployment: NA
Campagner et al., 2021[89]	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[98]	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
Chadaga et al., 2022[100]	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[101]	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[4]	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[90]	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
Famiglini et al., 2022[91]	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[99]	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[93]	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[92]	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* Logistic 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.

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[100]. On the other hand, the theoretical 'competing diagnosis' was referred to as a possible 'adverse engine' for some clinicians to stop assessing and simply act as a consequence of the recommendations issued by the model.

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 variates 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, and others[101]

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[102]. This example elucidates how training should encompass 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). Despite this protocol is still under development, The Transparent Reporting of a

multivariate prediction model for Individual Prognosis or Diagnosis (TRIPOD, 2015) standard[103] should be considered since It provides guidance and recommendations for the reporting of 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, and efficiency, as well as 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 terms of 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, as well as a clear gap between standard disease-associated metabolites and those chosen machine learning models, which resulted 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 clearly changing the practice of medicine, and digital tools are key for helping physicians evaluate patients in a more personalized, rapid, and efficient manner. 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.

REFERENCES

1. Sarker, I. H. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Comput. Sci.* **2**, (2021).
2. Bajwa, J., Munir, U., Nori, A. & Williams, B. Artificial intelligence in healthcare: transforming the practice of medicine. *Futur. Healthc. J.* **8**, e188–e194 (2021).
3. Reardon, S. Rise of Robot Radiologists. *Nature* **576**, S54–S58 (2019).
4. Luo, J., Zhou, L., Feng, Y., Li, B. & Guo, S. The selection of indicators from initial blood routine test results to improve the accuracy of early prediction of COVID-19 severity. *PLoS One* **16**, 1–18 (2021).
5. Demichev, V. *et al.* A time-resolved proteomic and prognostic map of COVID-19. *Cell Syst.* **12**, 780-794.e7 (2021).
6. Lin, J. K., Chien, T. W., Wang, L. Y. & Chou, W. 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. *Medicine (Baltimore).* **100**, e26532 (2021).
7. Li, W. T. *et al.* Using machine learning of clinical data to diagnose COVID-19: A systematic review and meta-analysis. *BMC Med. Inform. Decis. Mak.* **20**, 1–13 (2020).
8. Kline, R. R. Cybernetics, automata studies, and the dartmouth conference on artificial intelligence. *IEEE Ann. Hist. Comput.* **33**, 5–16 (2011).
9. Krizhevsky, A., Sutskever, I. & Hinton, G. ImageNet Classification with Deep Convolutional Neural Networks. *Adv. Neural Inf. Process. Syst. 25 (NIPS 2012)* 1097–1105 (2012) doi:10.1145/3383972.3383975.
10. Shukla Shubhendu, S. & Vijay, J. Applicability of Artificial Intelligence in Different Fields of Life. *Int. J. Sci. Eng. Res.* **1**, 28–35 (2013).
11. Gunčar, G., Kukar, M., Notar, M., Brvar, M. & Černelč, P. Application of machine learning for hematological diagnosis. 1–15.
12. The Medical Futurist. FDA approved AI-based algorithms. https://medicalfuturist.com/fda-approved-ai-based-algorithms/ (2022).
13. Ahmad, Z., Rahim, S., Zubair, M. & Abdul-Ghafar, J. 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 (2021).
14. Bruckert, S., Finzel, B. & Schmid, U. The Next Generation of Medical Decision Support: A Roadmap Toward Transparent Expert Companions. *Front. Artif. Intell.* **3**, 1–13 (2020).
15. Sarker, I. H. AI-Based Modeling: Techniques, Applications and Research Issues Towards Automation, Intelligent and Smart Systems. *SN Comput. Sci.* **3**, 1–20 (2022).
16. Kerr, W. T., Lau, E. P., Owens, G. E. & Trefler, A. The future of medical diagnostics: large digitized databases. *Yale J. Biol. Med.* **85**, 363–377 (2012).
17. Alsuliman, T., Humaidan, D. & Sliman, L. Machine learning and artificial intelligence in the service of medicine: Necessity or potentiality? *Curr. Res. Transl. Med.* (2020) doi:10.1016/j.retram.2020.01.002.
18. Svensson, C. M., Hübler, R. & Figge, M. T. Automated Classification of Circulating Tumor Cells and the Impact of Interobsever Variability on Classifier Training and Performance. *J. Immunol. Res.* **2015**, (2015).
19. Zheng, H., Zhu, J., Xie, W. & Zhong, J. Reinforcement learning assisted oxygen therapy for COVID-19 patients under intensive care. *BMC Med. Inform. Decis. Mak.* **21**, 1–8 (2021).
20. Pedregosa, F. and Varoquaux, G. and Gramfort, A. and Michel, V., and Thirion, B. and Grisel, O. and Blondel, M. and Prettenhofer, P., and Weiss, R. and Dubourg, V. and Vanderplas, J. and Passos, A. and & Cournapeau, D. and Brucher, M. and Perrot, M. and Duchesnay, E. Scikit-learn: Machine

Learning in Python. *Journal of Machine Learning Research 12, pp. 2825-2830* 2825–2830 at (2011).
21. Benjamens, S., Dhunnoo, P. & Meskó, B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *npj Digit. Med.* **3**, (2020).
22. Hao, K. Training a single AI model can emit as much carbon as five cars in their lifetimes. *MIT Technology Review.*
23. Fujiwara, K. Sparse Modeling delivers fast, energy efficient and explainable AI solutions for cutting-edge medical applications. 50–51 (2021).
24. H. Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, B. A. y A. Communication-Efficient Learning of Deep Networks from Decentralized Data. **54**, 10 (2017).
25. Dayan, I. *et al.* Federated learning for predicting clinical outcomes in patients with COVID-19. *Nat. Med.* **27**, 1735–1743 (2021).
26. Kairouz, P. *et al.* Advances and open problems in federated learning. *Found. Trends Mach. Learn.* **14**, 1–210 (2021).
27. Badrick, T. Evidence-based laboratory medicine. *Clin. Biochem. Rev.* **34**, 43–46 (2013).
28. Celkan, T. T. What does a hemogram say to us? *Turkish Arch. Pediatr.* **55**, 103–116 (2020).
29. Marieb, E. N. & Hoehn, K. Blood Composition and Functions. *Hum. Anat. Physiol.* 634–657 (2012).
30. Robert Chatburn, E. M.-C. Hematology. in *Handbook of Respiratory Care Third Edition* 54–63 (2010).
31. Richard A. McPherson, M. & Matthew R. Pincus, MD, P. *Clinical Diagnosis and Management by Laboratory Methods*. (2011). doi:10.1136/jcp.34.2.228-a.
32. Matthew R. Pincus, N. Z. A. J. INTERPRETING LABORATORY RESULTS. in *Henry's Clinical Diagnosis and Management22th Edition.pdf* 92–94 (Elsevier, 2011).
33. Kocbek, P. *et al.* 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**, (2019).
34. Soguero-Ruiz, C. *et al.* Data-driven Temporal Prediction of Surgical Site Infection. *AMIA ... Annu. Symp. proceedings. AMIA Symp.* **2015**, 1164–1173 (2015).
35. Ho, T. S. *et al.* Comparing machine learning with case-control models to identify confirmed dengue cases. *PLoS Negl. Trop. Dis.* **14**, 1–21 (2020).
36. Sarbaz, M., Pournik, O., Ghalichi, L., Kimiafar, K. & Reza Razavi, A. 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 (2013).
37. Ratzinger, F. *et al.* A risk prediction model for screening bacteremic patients: A cross sectional study. *PLoS One* **9**, (2014).
38. Mooney, C. *et al.* Predicting bacteraemia in maternity patients using full blood count parameters: A supervised machine learning algorithm approach. *Int. J. Lab. Hematol.* **43**, 609–615 (2021).
39. Zoabi, Y. *et al.* Predicting bloodstream infection outcome using machine learning. *Sci. Rep.* **11**, 1–11 (2021).
40. Rawson, T. M. *et al.* Supervised machine learning for the prediction of infection on admission to hospital: A prospective observational cohort study. *J. Antimicrob. Chemother.* **74**, 1108–1115 (2019).
41. Morang'a, C. M. *et al.* Machine learning approaches classify clinical malaria outcomes based on haematological parameters. *BMC Med.* **18**, 1–16 (2020).
42. Kinar, Y. *et al.* 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. Informatics Assoc.* **23**, 879–890 (2016).
43. Hornbrook, M. C. *et al.* Early Colorectal Cancer Detected by Machine Learning Model Using Gender, Age, and Complete Blood Count Data. *Dig. Dis. Sci.* **62**, 2719–2727 (2017).

44. Mahmood, N. *et al.* Identification of significant risks in pediatric acute lymphoblastic leukemia (ALL) through machine learning (ML) approach. *Med. Biol. Eng. Comput.* **58**, 2631–2640 (2020).
45. Haider, R. Z., Ujjan, I. U., Khan, N. A., Urrechaga, E. & Shamsi, T. S. Beyond the In-Practice CBC: The Research CBC Parameters-Driven Machine Learning Predictive Modeling for Early Differentiation among Leukemias. *Diagnostics* **12**, (2022).
46. Meiseles, A. *et al.* Explainable machine learning for chronic lymphocytic leukemia treatment prediction using only inexpensive tests. *Comput. Biol. Med.* **145**, 105490 (2022).
47. Soerensen, P. D. *et al.* 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. Lab. Med.* 1–12 (2021) doi:10.1515/cclm-2021-1015.
48. Yılmaz, Z. & Bozkurt, M. R. Determination of women iron deficiency anemia using neural networks. *J. Med. Syst.* **36**, 2941–2945 (2012).
49. Azarkhish, I., Raoufy, M. R. & Gharibzadeh, S. Artificial intelligence models for predicting iron deficiency anemia and iron serum level based on accessible laboratory data. *J. Med. Syst.* **36**, 2057–2061 (2012).
50. Barnhart-Magen, G., Gotlib, V., Marilus, R. & Einav, Y. Differential diagnostics of thalassemia minor by artificial neural networks model. *J. Clin. Lab. Anal.* **27**, 481–486 (2013).
51. Çil, B., Ayyıldız, H. & Tuncer, T. Discrimination of β-thalassemia and iron deficiency anemia through extreme learning machine and regularized extreme learning machine based decision support system. *Med. Hypotheses* **138**, (2020).
52. Kopitar, L., Kocbek, P., Cilar, L., Sheikh, A. & Stiglic, G. Early detection of type 2 diabetes mellitus using machine learning-based prediction models. *Sci. Rep.* **10**, 1–12 (2020).
53. Kushner, T., Breton, M. D. & Sankaranarayanan, S. Multi-Hour Blood Glucose Prediction in Type 1 Diabetes: A Patient-Specific Approach Using Shallow Neural Network Models. *Diabetes Technol. Ther.* **22**, 883–891 (2020).
54. Bernardini, M., Morettini, M., Romeo, L., Frontoni, E. & Burattini, L. 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 (2019).
55. Mathioudakis, N. N. *et al.* Development and Validation of a Machine Learning Model to Predict Near-Term Risk of Iatrogenic Hypoglycemia in Hospitalized Patients. *JAMA Netw. Open* **4**, 4–15 (2021).
56. Metsker, O. *et al.* Identification of risk factors for patients with diabetes: Diabetic polyneuropathy case study. *BMC Med. Inform. Decis. Mak.* **20**, 1–15 (2020).
57. Dipnall, J. F. *et al.* Fusing data mining, machine learning and traditional statistics to detect biomarkers associated with depression. *PLoS One* **11**, 1–23 (2016).
58. Hochman, E. *et al.* Development and validation of a machine learning-based postpartum depression prediction model: A nationwide cohort study. *Depress. Anxiety* **38**, 400–411 (2021).
59. Shou, L. *et al.* Blood Biomarkers Predict Cardiac Workload Using Machine Learning. *Biomed Res. Int.* **2021**, (2021).
60. Zheng, Y. *et al.* Rapid triage for ischemic stroke: a machine learning-driven approach in the context of predictive, preventive and personalised medicine. *EPMA J.* 285–298 (2022) doi:10.1007/s13167-022-00283-4.
61. Guo, K. *et al.* 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 (2021).
62. Mamoshina, P. *et al.* Blood Biochemistry Analysis to Detect Smoking Status and Quantify Accelerated Aging in Smokers. *Sci. Rep.* **9**, 1–10 (2019).
63. Zhan, J. *et al.* Diagnosis of Asthma Based on Routine Blood Biomarkers Using Machine Learning. *Comput. Intell. Neurosci.* **2020**, (2020).
64. Tian, Z. *et al.* Molecular pathogenesis: Connections between viral hepatitis-induced and non-alcoholic steatohepatitis-induced hepatocellular carcinoma. *Front. Immunol.* **13**, 1–12 (2022).

65. Fialoke, S., Malarstig, A., Miller, M. R. & Dumitriu, A. Application of Machine Learning Methods to Predict Non-Alcoholic Steatohepatitis (NASH) in Non-Alcoholic Fatty Liver (NAFL) Patients. *AMIA ... Annu. Symp. proceedings. AMIA Symp.* **2018**, 430–439 (2018).
66. Ma, H., Xu, C. F., Shen, Z., Yu, C. H. & Li, Y. M. Application of Machine Learning Techniques for Clinical Predictive Modeling: A Cross-Sectional Study on Nonalcoholic Fatty Liver Disease in China. *Biomed Res. Int.* **2018**, (2018).
67. Cao, Y., Hu, Z. De, Liu, X. F., Deng, A. M. & Hu, C. J. An MLP classifier for prediction of HBV-induced liver cirrhosis using routinely available clinical parameters. *Dis. Markers* **35**, 653–660 (2013).
68. Yao, Z., Li, J., Guan, Z., Ye, Y. & Chen, Y. Liver disease screening based on densely connected deep neural networks. *Neural Networks* **123**, 299–304 (2020).
69. Peng, J., Zhou, M., Chen, C., Xie, X. & Luo, C. H. Identification of exacerbation risk in patients with liver dysfunction using machine learning algorithms. *PLoS One* **15**, 1–10 (2020).
70. Tarwater, K. Estimated glomerular filtration rate explained. *Mo. Med.* **108**, 29–32 (2011).
71. Rashed-Al-Mahfuz, M. *et al.* Clinically Applicable Machine Learning Approaches to Identify Attributes of Chronic Kidney Disease (CKD) for Use in Low-Cost Diagnostic Screening. *IEEE J. Transl. Eng. Heal. Med.* **9**, 1–11 (2021).
72. Our World in data. Covid-19 cases. https://ourworldindata.org/grapher/covid-cases-income.
73. Data, O. W. in. Covid-19 deaths.
74. Wu, Z. & McGoogan, J. M. 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 (2020).
75. Auld, S. C. *et al.* Trends in ICU Mortality From Coronavirus Disease 2019. *Crit. Care Med.* **Publish Ah**, 245–255 (2021).
76. Böger, B. *et al.* Systematic review with meta-analysis of the accuracy of diagnostic tests for COVID-19. *Am. J. Infect. Control* **49**, 21–29 (2021).
77. Diagnostics for All. Foundation for Innovative New Diagnostics. *Test Directory*.
78. World Health Organization. Antigen-detection in the diagnosis of SARS-CoV-2 infection. *Interim Guid.* (2021).
79. Food and Drug Administration. 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*.
80. Prince-Guerra, J. L. *et al.* Evaluation of Abbott 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 (2021).
81. Joshi, R. P. *et al.* A predictive tool for identification of SARS-CoV-2 PCR-negative emergency T department patients using routine test results. (2020).
82. Brinati, D. *et al.* Detection of COVID-19 Infection from Routine Blood Exams with Machine Learning: A Feasibility Study. *J. Med. Syst.* **44**, (2020).
83. Alves, M. A. *et al.* Explaining machine learning based diagnosis of COVID-19 from routine blood tests with decision trees and criteria graphs. *Comput. Biol. Med.* **132**, (2021).
84. Banerjee, A. *et al.* Use of Machine Learning and Artificial Intelligence to predict SARS-CoV-2 infection from Full Blood Counts in a population. *Int. Immunopharmacol.* **86**, 106705 (2020).
85. Wu, J., Shen, J., Xu, M. & Shao, M. A novel combined dynamic ensemble selection model for imbalanced data to detect COVID-19 from complete blood count. (2020).
86. Souza, A. A. d. *et al.* Simple hemogram to support the decision-making of COVID-19 diagnosis using clusters analysis with self-organizing maps neural network. *Soft Comput.* (2021) doi:10.1007/s00500-021-05810-5.
87. Tschoellitsch, T., Dünser, M., Böck, C., Schwarzbauer, K. & Meier, J. Machine Learning Prediction

of SARS-CoV-2 Polymerase Chain Reaction Results with Routine Blood Tests. *Lab. Med.* **52**, 146–149 (2021).

88. Cabitza, F. *et al.* The importance of being external. methodological insights for the external validation of machine learning models in medicine. *Comput. Methods Programs Biomed.* **208**, 106288 (2021).
89. Babaei, S., Sorayaie, A., Ghafari, A. & Bagherzadeh, J. COVID-19 diagnosis from routine blood tests using artificial intelligence techniques. (2020).
90. Campagner, A., Carobene, A. & Cabitza, F. External validation of Machine Learning models for COVID-19 detection based on Complete Blood Count. *Heal. Inf. Sci. Syst.* **9**, 1–15 (2021).
91. Benito-León, J. *et al.* 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**, 1–14 (2021).
92. Famiglini, L., Campagner, A., Carobene, A. & Cabitza, F. A robust and parsimonious machine learning method to predict ICU admission of COVID-19 patients. *Med. Biol. Eng. Comput.* (2022) doi:10.1007/s11517-022-02543-x.
93. Murri, R. *et al.* A machine-learning parsimonious multivariable predictive model of mortality risk in patients with Covid-19. *Sci. Rep.* **11**, 1–10 (2021).
94. Fernandes, F. T. *et al.* A multipurpose machine learning approach to predict COVID-19 negative prognosis in São Paulo, Brazil. *Sci. Rep.* **11**, 1–7 (2021).
95. Joshi, R. P. *et al.* A predictive tool for identification of SARS-CoV-2 PCR-negative emergency T department patients using routine test results. (2020).
96. Plante, T. B. *et al.* 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**, 1–12 (2020).
97. Babaei, S., Sorayaie, A., Ghafari, A. & Bagherzadeh, J. COVID-19 diagnosis from routine blood tests using artificial intelligence techniques. (2021).
98. Jiachao Wu , Jiang Shen , Man Xu, M. S. A novel combined dynamic ensemble selection model for imbalanced data to detect COVID-19 from complete blood count. *Computer Methods and Programs in Biomedicine* (2021).
99. Abayomi-Alli, O. O., Damaševičius, R., Maskeliūnas, R. & Misra, S. An Ensemble Learning Model for COVID-19 Detection from Blood Test Samples. *Sensors* **22**, (2022).
100. Chadaga, K., Prabhu, S., Vivekananda Bhat, K., Umakanth, S. & Sampathila, N. Medical diagnosis of COVID-19 using blood tests and machine learning. *J. Phys. Conf. Ser.* **2161**, (2022).
101. Chadaga, K. *et al.* A Decision Support System for Diagnosis of COVID-19 from Non-COVID-19 Influenza-like Illness Using Explainable Artificial Intelligence. *Bioengineering* **10**, (2023).
102. Karthikeyan, A., Garg, A., Vinod, P. K. & Priyakumar, U. D. Machine Learning Based Clinical Decision Support System for Early COVID-19 Mortality Prediction. *Front. Public Heal.* **9**, 1–13 (2021).
103. Henry, K. E. *et al.* Human – machine teaming is key to AI adoption : clinicians ' experiences with a deployed machine learning system. 1–6 (2022) doi:10.1038/s41746-022-00597-7.
104. Stegmann, G. M. *et al.* Repeatability of Commonly Used Speech and Language Features for Clinical Applications. *Digit. Biomarkers* **4**, 109–122 (2020).
105. Co-operation, E. Machine learning in translation. *Nat. Biomed. Eng.* **5**, 485–486 (2021).
106. Collins, G. S., Reitsma, J. B., Altman, D. G. & Moons, K. G. M. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *BMJ* **350**, 1–9 (2015).

AUTHOR CONTRIBUTIONS All authors contributed to the conception and design of the study. M.A.S.-S. contributed to the literature search and data extraction with N.Sousa support in coordination and validation. J.C.S. and M.A.S.-S. contributed to data analysis and interpretation. M.A.S.-S. drafted the manuscript with J.C.Sousa's feedback, proof reading, and content curation. All authors approved the manuscript. All authors guaranteed the integrity of the work.

COMPETING INTERESTS The authors declare no competing interests.

CHAPTER 2.3

Pattern recognition of hematological profiles of tumors of the digestive tract: an exploratory study

The work presented in this chapter has been published in Frontiers in Medicine, section Translational Medicine.

Pattern recognition of hematological profiles of tumors of the digestive tract: an exploratory study

Miguel A. Santos-Silva[1,2], Nuno Sousa[1,2,3,5], Marina Majar[4], Miguel Machado[4], Joana Reis[1,2,3], João Carlos Sousa[1,2,*]

[1] Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho, Campus de Gualtar, 4750-057 Braga, Portugal

[2] ICVS/3B's – PT Government Associate Laboratory, Braga/Guimarães, Portugal

[3] Clinical Academic Center-Braga (2CA), 4710-243 Braga, Portugal

[4] Clinical Pathology Department, Hospital de Braga, 4710-243 Braga, Portugal

[5] Association P5 Digital Medical Center (ACMP5), 4710-057, Braga, Portugal

[*] Correspondence: jcsousa@med.uminho.pt

ABSTRACT

Aims: In this study, we aimed to apply laboratory blood analysis to identify the hematological (based on hemoglobin concentration, erythrocytes, hematocrit, and RDW count) profiles associated with the most prevalent forms of digestive tract malignancies. Furthermore, we aimed to evaluate how these profiles contributed to distinguishing these tumors at diagnosis.

Methods: We collected data from the date of ICD-10 diagnostic coding for C15 esophagus, C16 stomach, C18 colon, and C19 rectum tumors of 184 individuals. The statistical analysis and data visualization approaches, notably the heat map and principal components analysis (PCA), allowed for creating a summary hematological profile and identifying the most associated parameters for each pathologic state. Univariate and multivariate data modeling and ROC analysis were performed in SPSS and Python.

Results: Our data reveal unique patterns based on tumor development anatomical location, clustering the C18 colon and C19 rectum from the C15 esophagus and C16 stomach. We found a significant difference between C16 stomach carcinoma and the other tumors, which substantially correlated with raised RDW in conjunction with low hemoglobin concentration, erythrocytes, and hematocrit counts. In contrast, C18 colon carcinoma had the higher red blood cell count, allowing for the best classification metrics in the test set of the binary logistic regression (LR) model, accounting for an AUC of 0.77 with 94% sensitivity and 52% specificity.

Conclusions: This study emphasizes the significance of adding hematological patterns in diagnosing these malignancies, which could path further investigations regarding profiling and monitoring at the point-of-care.

Keywords: cancer, pattern recognition, hematological profile, anatomical location

INTRODUCTION

Esophageal, gastric, and colorectal carcinomas are among the most prevalent malignancies and account for a significant portion of cancer-related morbidity and death globally, with a considerable burden on healthcare[1]. In fact, the global growth of digestive tract cancers has impacted millions of people. According to the International Agency for Research on Cancer (IARC), colorectal cancer (CRC) has a 6.1% incidence rate and a 9.2% fatality rate, followed by stomach (8.2%), esophagus (5.3%), and rectum (3.2%)[2]. Diagnosis of these malignancies is well standardized in current clinical decision support systems, through the use of several technologies,

from DNA sequencing, chromosomal and immunology analysis, endoscopy[3], colonoscopy[4], histology, and blood tests, among others[5]. Nevertheless, blood tests remain the primary method used by clinicians to assess the hematological profile and conduct subsequent investigations, influencing nearly 70% of the medical decisions[6].

Generally, low levels of RBC, Hb, HTC, and high values of WBC, PLT and RDW associate with cancer diagnosis[7]. RDW was previously identified as a biomarker of right-sided CRC cases, with an 84% sensitivity and an 88% specificity[8]. It was also discovered to exhibit elevated values in patients with esophageal[9] and gastric[10] cancers, with a significant correlation with the digestive tract tumor stage. As a prevalent clinical condition in digestive tract tumors, anemia recognition by measuring Hb over time is also considered an important factor for improving CRC detection and further diagnosis[11].

Artificial intelligence (AI) driven research, has been developing remarkable results in feasibility studies involving routine blood tests with cancer diagnosis[12,13,14] and prognosis[15,16]. The machine-learning ColonFlag® model developed by Kinar et al., used a decision-trees algorithm to predict the risk of CRC based on age, gender and cell blood counts (CBC), demonstrating increased sensitivity (AUC=0.81) than anemia guidelines (AUC=0.76), especially considering a 6-month period before diagnosis[12]. Hornbrook et al., validated the ColonFlag® model in a US insured population by confirming model applicability in CRC diagnosis (AUC=0.80), and also highlighting model performance based on the anatomical location of the carcinoma, with better metrics in the cecum and ascending colon rather than in transverse, sigmoid, and rectum[13],

Pattern recognition (PR) refers to the AI ability to infer underlying patterns (regularities, trends, or anomalies) in the data[17]. Differently from statistics, PR automatically extracts actionable knowledge from complex datasets. When leveraged by machine or deep learning algorithms, trained models could be used to predict similar structures. Nonetheless, the degree of model explainability decreases in deep learning, limiting model's applicability[18]. In blood data, there is a vast amount of non-appraised clinical information that cannot be 100% perceived by clinicians, endorsing the use of PR in processing, patterning, and flagging if necessary[19].

Thus, we thought of interest to go beyond the classic studies evaluating potential biomarkers for diagnosis and explore hematological data from patients with these pathologies from a distinct perspective. Indeed, herein, we assessed whether erythrocytes, hemoglobin, hematocrit, and RDW enable profile distinction between the location of digestive tract tumors and if they can (and contribute to) distinguish them at the time of diagnosis.

METHODS

Study design and population

This study is a retrospective observational study of Portuguese people diagnosed with the most prevalent oncological diseases at the Hospital de Braga between January 2018 and 2021. The Hospital's oncological registry issues an identification code to combine the patient's clinical information longitudinally. We collected anonymous data from adult patients (18+ years of age) at random prior to the diagnosis codification date, specifically: demographics (gender and age), laboratory blood tests (complete blood count and routine biochemistry), and the diagnosis (ICD-10). We gathered information on 184 patients with digestive system cancers, including the esophagus, stomach, colon, and rectum. Laboratory blood tests were manually filtered to identify four distinct parameters conducted in every case, in a total of 760 tests. The study was approved by the Braga Hospital Ethics Committee under the project "Application of machine learning for hematological diagnosis" (protocol code 191_2022).

Data source, measurement, and features

Patient anonymized demographics and laboratory blood test data were retrieved in the Clinic Academic Center (2CA, Braga, Portugal) of the Hospital de Braga. Cell blood count, including erythrocytes, hematocrit, red cell distribution width, and hemoglobin concentration, were analyzed in the clinical pathology laboratory using standard methods (Sysmex XE-2100, Sysmex Inc, Mundelein IL, USA). Disease diagnosis was codified according to the International Statistical Classification of Diseases and Related Health Problems 10th revision (ICD-10). Time before disease codification, i.e., timeframe, was converted in a scalar of days to comply with anonymization requirements, and it was calculated as the absolute difference between the date of disease codification (0) with the date of analysis, i.e., 731 – starting period of the retrospective analysis.

Statistical analysis

The statistical analysis explored the ICD-10 studied diseases as a function of gender, age, and metabolites. This analysis was performed with SPSS (IBM SPSS Statistics for Mac, version 26, IBM Corp., NY, USA). Continuous variables were evaluated for normal distribution with histograms (skewness and kurtosis) and described using mean and standard deviation (SD). Skewed continuous variables were reported with median and interquartile range (IQR). The timeframe of analysis was selected to zero, corresponding to the codification date. One-way ANOVA (with

Tukey's HSD as a posthoc test) and Kruskal Wallis were used to compare groups for parametric and non-parametric variables, respectively. The statistical significance level was set at 0.05.

Pattern recognition and data modeling

After z-score normalization, PCA and heat map (supervised visualization tool) was used to explore and visualize patterns across the research laboratory blood tests within the study groups. The "Clustvis" online tool, found at https://biit.cs.ut.ee/clustvis/, was used to create the scores plot of the PCA and the heat map, which was computed using the correlation function for 'clustering distance', the 'tightest cluster first' for clustering and functionalized with the RdBu palette, from -2 to 2. Principal components (PC1 and PC2) with respective loadings were also made available for interpretation and parameters influence in diseases discrimination. Preprocessing was accomplished in Python 3.10.2 (VS 1.64.2) through the exploration of specific libraries for data acquisition (accessing database's raw data), curation (removal of incorrect values or characters and merging of separate intraday analysis) and normalization (i.e., z-score). Univariate and multivariate data modeling, along with ROC analysis, were computed both in SPSS and Python. The classification performance was measured by the area under the receiver operator curve (AUROC), which ranged from zero to one, with one being a perfect classifier. Because our experiments had four classes, the AUROC was calculated one-versus-rest for each class. The F1-score, which ranges from 0 to 1, represents the balanced mean of precision and recall.

RESULTS

The study cohort included 184 patients (60% males and 40% females, median age of 69 years, IQR 58-81) with complete blood analysis performed on the diagnostic date (t0). C18 colon (n=84) was the most prevalent group, followed by C16 stomach (n=66), C19+C20 rectum (n=19), and C15 esophagus (n=15). On the day of codification, each illness had 89, 66, 19, and 16 clinical blood analyses completed, respectively.

Except for age and RDW, which are shown with the median and interquartile range due to their continuous skewed distribution, Table 1 depicts the quantitative features of patients in terms of mean and standard deviation. Indeed, neither age nor RDW were found to have significant differences across groups; however, this assumption was only confirmed for age since the Mann-Whitney tests revealed significant differences in RDW between C15 esophagus and C16 stomach (p=.035), and between C16 stomach and C18 colon (p=.026).

Considering the remaining hematological parameters, statistically significant differences in erythrocytes, hemoglobin, and hematocrit were found, with superior levels in C18 colon, followed by C19 rectum, C15 esophagus, and C16 stomach. Post-hoc analysis using Tukey's HSD revealed further significant differences in erythrocytes between C16 stomach and C19 rectum (p=.001), as well as statistically significant variances in hemoglobin between C16 stomach and C18 colon (p<.001), and C19 rectum (p=.034), and also in hematocrit between C16 stomach and C18 colon (p<.001), and C19 rectum (p=.016).

Table 1: Descriptive statistics of demographics and laboratory blood tests according to each digestive tract tumor.

	Study Groups				
	C15 Esophagus	**C16 Stomach**	**C18 Colon**	**C19+C20 Rectum**	**Sig.**
Number of blood tests	16	66	89	19	
Number of patients	15	66	84	19	
Gender number (f/m)	3/12	24/42	40/44	8/11	
Age (years) *	68 (57–82)	69 (61–81)	68 (58–83)	71 (53–78)	0.919
Erythrocytes (x $10^6/\mu L$)	3.80 ± 0.58	3.73 ± 0.75	4.33 ± 0.81	4.07 ± 0.87	<0.005
Hemoglobin (g/dL)	11.38 ± 1.79	9.96 ± 2.65	12.05 ± 2.45	11.72 ± 2.43	<0.005
Hematocrit (%)	34.07 ± 4.94	30.50 ± 7.03	36.39 ± 6.56	35.71 ± 6.92	<0.005
RDW (%) *	13.0 (12.3–15.2)	15.0 (12.8–18.4)	13.7 (12.5–15.2)	14.1 (13.1–14.6)	0.056

*The normal distribution was not confirmed. Non-parametric tests (Kruskal-Wallis) were used to compare groups, and results are provided as the median and interquartile range (IQR).

The heat map allows for the visualization of laboratory blood test patterns (rows) in relation to the researched disease categories (columns) (**Figure 1**). The matrix profiles each disease location based on its correlation with each metabolite, with a strong positive correlation shown in red and a significant negative correlation shown in blue. Moreover, the map groups diseases and metabolites by nodding (branching), approaching C15 esophagus and C19 rectum to C18 colon, and separating C16 stomach from the preceding. Indeed, C16 stomach displays the strongest correlated profile, with high values of RDW and low hematocrit, hemoglobin, and erythrocytes levels. C18 colon displays a strong correlation between high levels of erythrocytes and moderate association with hemoglobin and hematocrit, with a practically negligible association with RDW. Low RDW levels closely associate with C15 esophagus, which is also characterized by low erythrocyte count. The C19 rectum displays the weakest correlation profile.

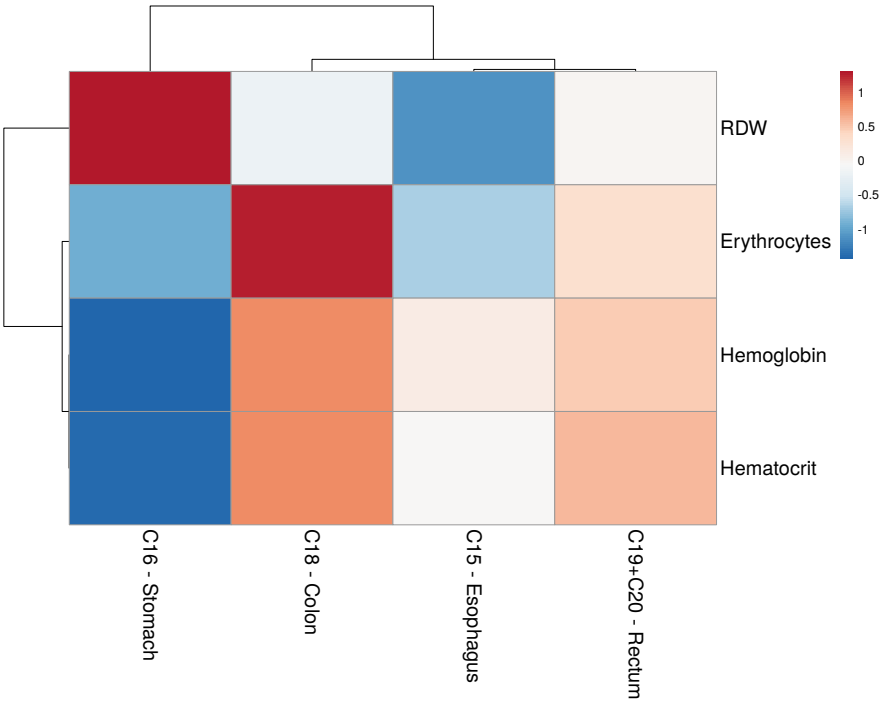


Figure 1: Correlation heat map between each blood parameter and the tumors of the digestive tract.

Considering the previously exhibited correlations, each laboratory blood test was subjected to a ROC analysis to assess its predictive ability among the studied groups. We calculated the confidence intervals at a 95% confidence level for the associated AUC values of each parameter. Only RDW in C16 stomach (0.52 – 0.70), erythrocytes (0.62 – 0.77), hemoglobin (0.59 – 0.74), and hematocrit (0.60 – 0.76) in C18 colon revealed potential diagnostic value (AUC > 0.5) in distinguishing the respective tumor groups. The studied parameters displayed a lower ability to distinguish C15 esophagus or C19 rectum from the others. Both sensitivity and specificity maximized by the Youden index were computed for the parameters with diagnostic ability: while RDW in C16 stomach achieved 52% and 73%, C18 colon was predictable by erythrocytes with 60% and 71%, hemoglobin with 76% and 52%, and by hematocrit with 80% and 51%, respectively. Principal components analysis is a dimensionality reduction algorithm that was used to maximize the variance between disease groups through the linear merging of the blood parameters, and extract information regarding the latent variables (principal components) that explain the distribution of the scores. The first principal component of the PCA plot (**PC1, Figure 2**) separates C16 stomach from the other tumors, with a direct influence of the higher values of RDW (-0.40) together with the low levels of hemoglobin (0.56), hematocrit (0.56), and erythrocytes (0.47). The second principal component, which explains the remaining 20% of the PCA, is strongly determined

by erythrocytes (-0.61) and RDW (-0.79), wherein C15 esophagus is clustered from the remaining tumors due to the low levels of RDW and erythrocytes. A small overlap between C19 rectum and C18 colon was also verified, which should not be sufficient for tumors separation.

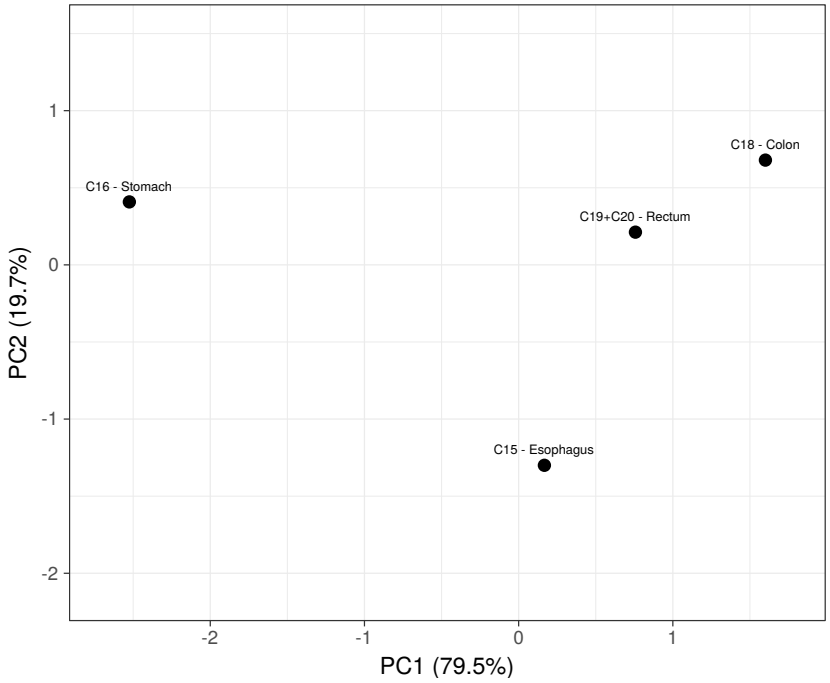


Figure 2: Principal Components Analysis (PCA) plot discrimination of the studied digestive tumors, based on the four hematologic parameters.

LR was computed to evaluate the predictive ability of the studied hematological parameters for the classification of each digestive tract tumor. To preserve representative distribution between training and test sets, data was split in a 70:30 ratio using stratified group k-fold, ensuring similar proportions in class distribution for each subset. **Figure 3** highlights the predictive performance of each class by using data from the test set. As expected, the aggregation of the studied metabolites enabled higher performance metrics when compared to univariate analysis. Both C18 colon and C16 stomach exhibited superior performance, confirming the findings of the heat map, which indicated higher correlation profiles for these tumors. Specifically, C18 colon achieved the highest AUC of 77% with to an excellent sensitivity (94%) and moderate specificity (52%). C16 stomach showed to be more precise (60%), with moderate recall (52%) and higher specificity (80%). Still, the f1-score of each tumor is comparable. C15 esophagus and C19+C20 rectum displayed lower performance metrics due to the lack of positive samples. Indeed, confusion matrix of the predicted test set failed to demonstrate sensitivity and f1-score due to the absence of true positive

predictions. C15 esophagus and C19+C20 rectum were less correlated with the studied blood parameters, and their AUC was 58% and 48%, respectively.

	AUC	Precision	Recall	Specificity	F1-score
C15 esophagus vs. rest	0.58	0.00	-	0.92	-
C16 stomach vs. rest	0.68	0.60	0.52	0.80	0.56
C18 colon vs. rest	0.77	0.43	0.94	0.52	0.59
C19+C20 rectum vs. rest	0.48	0.00	-	0.91	-

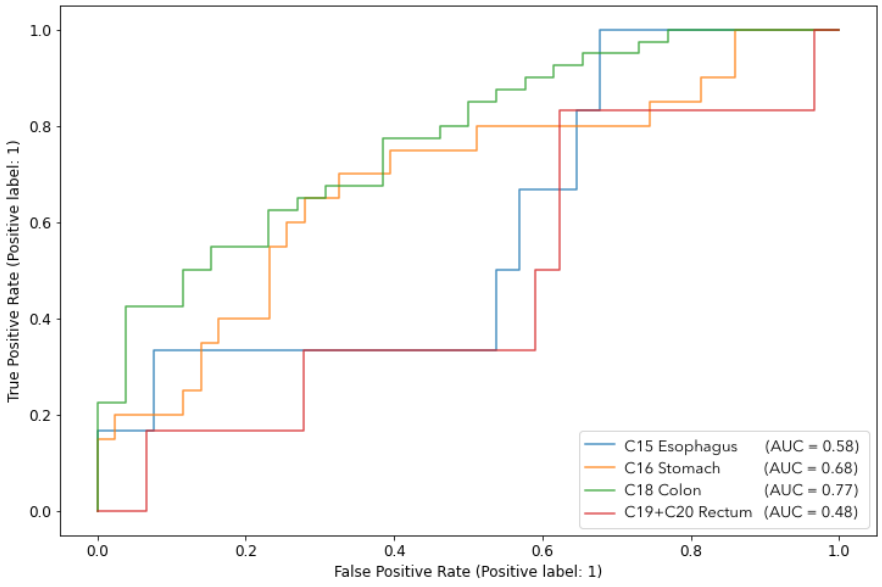


Figure 3: ROC curves and performance metrics of the LR prediction models for the classification of each digestive tract tumor.

DISCUSSION

In this study, we aimed to analyze if routine hematological parameters were able to perform profile distinction between the location of digestive tract tumors and whether they could contribute to distinguishing them at the time of diagnosis. We applied a cross-decomposition algorithm (PCA) to maximize the variance among the studied 184 blood tests and identify the latent variables that contributed to the model distribution.

Solely using the combination of the hemoglobin concentration, erythrocytes, hematocrit, and RDW count, the model clustered C16 stomach from the other tumors (C15 esophagus, C16 stomach, C18 colon, and C19 rectum) in a linear merge between high RDW count with low hemoglobin, erythrocytes, and hematocrit levels (PC1), explaining nearly 80% of the variance. The

remaining 20% belonged to the second principal component, which distinguished the C15 esophagus from the others due to its low erythrocytes and RDW count. Interestingly, no significant overlap was found between C18 colon and C19 rectum in the scores plot (**Figure 1**, first quadrant), neither across the hematological parameters compared in the statistical analysis of the two groups (**Table 1**), which is consistent with previous studies indicating similar patterns (of miRNA in the case) between C18 colon and C19 rectum due to the common hindgut region of tumor development[13]. Moreover, C15 esophagus and C16 stomach were clustered from the colorectal malignancies in the second (-PC1, +PC2) and fourth (+PC1, -PC2) quadrants of the PCA, keeping the different regions of tumor development separated.

The heat map analysis allowed an easier visualization of the hematological patterns and provided a better understanding of the relationship between blood tests and the malignancies analyzed. Notably, nodding (branching) across malignancies grouped profiles with low correlation values, such as C15 esophagus and C19 rectum. Furthermore, it aggregated C16 stomach and C18 colon, which substantially correlated with particular blood tests, such as RDW for C16 stomach and erythrocytes for C18 colon. Nevertheless, an extra node distinguished C16 stomach from the others (also influenced by hemoglobin and hematocrit, in accordance with PCA). Remarkably, these findings are coherent with recent studies that relate hematological parameters as predictors of diagnosis and prognosis of digestive tract malignancies[14]. Moreover, Pietrzyk et al. found that RDW alone could discriminate patients with gastric cancer from healthy individuals[15]. Yazici et al. described RDW as a prognostic gastric cancer biomarker with elevated values associated with short-term mortality[16]. In colorectal carcinoma, Kinar et al. used hemoglobin, hematocrit, RDW, MCH, MCHC, and MCV to diagnose, with sensitivity stability between 480 and 240 days before diagnosis (AUC 0.81 in an external evaluation set of 5000+ patients)[7].

The ability to associate routine blood tests to distinguish digestive tract tumors at the time of diagnosis was herein evaluated through the computation of univariate and multivariate analysis. While the univariate analysis confirmed the feature importance described in the heat map, the multivariate analysis computed on a binomial LR with the four predictors enhanced the discriminatory ability for each tumor. C18 colon had the most significant AUC of 0.77, with 94% sensitivity and 52% specificity, followed by C16 stomach, which had an AUC of 0.68, with 52% sensitivity and 80% specificity. C15 esophagus and C19 rectum were less predictable, with an AUC of 0.58 and 0.48, respectively. Interestingly, although C18 colon and C19 rectum were similarly patterned (PCA), their discrimination from the others was considerably different, demonstrating

that comparable patterns are not accurately anticipated until feature correlation maintains higher correlation values (heat map and ROC analysis).

This study presents some limitations, which we next highlight: the small sample size in each tumor category (particularly in C15 esophagus and C19 rectum) and the research's retrospective, single-center nature. Yet, because this strategy focused on recognizing disease profiles, blood tests from the ICD-10 codification date were valued more. Nonetheless, the given performance metrics are virtuous and promising, especially because comparisons were made between diseased patients only (without healthy volunteers) and predicted using routine blood parameters, which implies an additional potential of this methodological approach. We believe that the novelty generated by this study will trigger further multicentric studies to further validate the current findings. Furthermore, the ability to evaluate profile changes over time and correlate them with labeled stages of the disease may enable the development of a point-of-care follow-up map, extracting additional value from routine blood assessments.

CONCLUSIONS

Patterns of prevalent digestive tract tumors were recognized and categorized, considering the hematological results of the hemoglobin concentration, erythrocytes, and hematocrit counts from the date of ICD-10 codification. Tumor profiles were decomposed in agreement with the anatomical location of tumor development, separating the C18 colon and C19 rectum from the C15 esophagus and those from the C16 stomach, confirming the ability of hematological parameters to perform profile distinction in digestive tract tumors.

Both the heat map analysis and the multivariate binary logistic regression confirmed the importance of higher erythrocyte count in distinguishing C18 colon from other malignancies (AUC=0.77, 94% sensitivity, and 52% specificity) and the importance of both a high number of RDW with low levels of hemoglobin, hematocrit, and erythrocytes in distinguishing C16 stomach from the remaining tumors (AUC=0.68, 52% sensitivity, 80% specificity). Although C15 esophagus and C19 rectum were less predicted, this study demonstrates that routine blood tests have the potential predictive capacity to distinguish digestive tract tumors at diagnosis.

REFERENCES

1. Torre, L. A. *et al.* Global cancer statistics, 2012. *CA. Cancer J. Clin.* **65**, 87–108 (2015).
2. Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA. Cancer J. Clin.* **68**, 394–424 (2018).
3. Pasechnikov, V., Chukov, S., Fedorov, E., Kikuste, I. & Leja, M. Gastric cancer: Prevention, screening and early diagnosis. *World J. Gastroenterol.* **20**, 13842–13862 (2014).
4. Kanth, P. & Inadomi, J. M. Screening and prevention of colorectal cancer. *BMJ* **374**, n1855 (2021).
5. Necula, L. *et al.* Recent advances in gastric cancer early diagnosis. *World J. Gastroenterol.* **25**, 2029–2044 (2019).
6. Hallworth, M. J. The '70% claim': What is the evidence base? *Ann. Clin. Biochem.* **48**, 487–488 (2011).
7. Matthew R. Pincus, N. Z. A. J. INTERPRETING LABORATORY RESULTS. in *Henry's Clinical Diagnosis and Management22th Edition.pdf* 92–94 (Elsevier, 2011).
8. Spell, D. W., Jones, D. V., Harper, W. F. & Bessman, J. D. The value of a complete blood count in predicting cancer of the colon. *Cancer Detect. Prev.* **28**, 37–42 (2004).
9. Han, F. *et al.* Diagnosis and survival values of neutrophil-lymphocyte ratio (NLR) and red blood cell distribution width (RDW) in esophageal cancer. *Clin. Chim. Acta* **488**, 150–158 (2019).
10. Wei, T. T. *et al.* Relationship between red blood cell distribution width, bilirubin, and clinical characteristics of patients with gastric cancer. *Int. J. Lab. Hematol.* **39**, 497–501 (2017).
11. Goldshtein, I., Neeman, U., Chodick, G. & Shalev, V. Variations in hemoglobin before colorectal cancer diagnosis. *Eur. J. Cancer Prev.* **19**, 342–344 (2010).
12. Kinar, Y. *et al.* 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. Informatics Assoc.* **23**, 879–890 (2016).
13. Hornbrook, M. C. *et al.* Early Colorectal Cancer Detected by Machine Learning Model Using Gender, Age, and Complete Blood Count Data. *Dig. Dis. Sci.* **62**, 2719–2727 (2017).
14. Soerensen, P. D. *et al.* 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. Lab. Med.* 1–12 (2021) doi:10.1515/cclm-2021-1015.
15. Mahmood, N. *et al.* Identification of significant risks in pediatric acute lymphoblastic leukemia (ALL) through machine learning (ML) approach. *Med. Biol. Eng. Comput.* **58**, 2631–2640 (2020).
16. Meiseles, A. *et al.* Explainable machine learning for chronic lymphocytic leukemia treatment prediction using only inexpensive tests. *Comput. Biol. Med.* **145**, 105490 (2022).
17. Sarker, I. H. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Comput. Sci.* **2**, (2021).
18. Bruckert, S., Finzel, B. & Schmid, U. The Next Generation of Medical Decision Support: A Roadmap Toward Transparent Expert Companions. *Front. Artif. Intell.* **3**, 1–13 (2020).
19. Gunčar, G., Kukar, M., Notar, M., Brvar, M. & Černelč, P. Application of machine learning for hematological diagnosis. 1–15.
20. Tang, S. *et al.* Stratification of Digestive Cancers with Different Pathological Features and Survival Outcomes by MicroRNA Expression. *Sci. Rep.* **6**, 1–10 (2016).
21. Jung, D. H. *et al.* Impact of perioperative hemoglobin levels on postoperative outcomes in gastric cancer surgery. *Gastric Cancer* **16**, 377–382 (2013).
22. Pietrzyk, L., Plewa, Z., Denisow-Pietrzyk, M., Zebrowski, R. & Torres, K. Diagnostic power of blood parameters as screening markers in gastric cancer patients. *Asian Pacific J. Cancer Prev.* **17**, 4433–4437 (2016).
23. Yazici, P., Demir, U., Bozkurt, E., Isil, G. R. & Mihmanli, M. The role of red cell distribution width in the prognosis of patients with gastric cancer. *Cancer Biomarkers* **18**, 19–25 (2017).

AUTHOR CONTRIBUTIONS M.A.S.-Silva, M. Majar and N. Sousa contributed to the conception and design of the study. M. Machado contributed to data extraction with J. Reis and M. Majar support in coordination and validation. M.A.S.-Silva, N. Sousa and J.C.Sousa contributed to data analysis and interpretation. M.A.S.-Silva drafted the manuscript with N. Sousa and J.C.Sousa feedback, proof reading, and content curation. All authors approved the manuscript. All authors guaranteed the integrity of the work.

ETHICS APPROVAL STATEMENT The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Ethics Committee of Hospital de Braga (protocol code 191_2022 and date of approval on 21 December 2022).

FUNDING This work has been funded by national funds, through the Foundation for Science and Technology (FCT)—project UIDB/50026/2020 and UIDP/50026/2020.

COMPETING INTERESTS The authors declare no competing interests.

CHAPTER 3

Discussion, Conclusion, and Future Perspectives

DISCUSSION

This dissertation was motivated by the implicit value of clinical blood analysis in determining pathological states and the possibility of leveraging its value using inferential statistics and artificial intelligence strategies, such as pattern recognition and machine learning. The aim was methodologically implemented in several stages, from inferential statistics (**Chapter 2.1**), progressing to a review of machine learning applications (**Chapter 2.2**), and lastly, exploring pattern recognition algorithms (**Chapter 2.3**). The two last stages differ from the first due to the substantial shift implied in AI data-driven studies, in which we addressed the relevance of AI to extract insights from broad datasets rather than collecting and analyzing data specific to the outcome of interest. In all three, we demonstrated the ability of AI strategies to provide real-world insights regarding clinical blood analysis and supporting clinical decisions in detailed contexts of patient care.

In the first study, described in **Chapter 2.1**, we confirmed the potential of statistical analysis in defining the clinical characteristics of patients admitted to an emergency department. Briefly, we demonstrated through frequency and correlation analysis that 36% of admitted patients exhibited both anemia and inflammation (reaching 67% in hospitalized patients) and that increased CRP levels explained 15% of Hb variations. While the frequency analyses agreed with previous literature (Guenter, Weiss. Lawrence T., 2005), especially in the anemia determined by hemoglobin measurements, the linear regression model developed was novel since it was the first to describe the relationship between both parameters in the context of emergency departments. Furthermore, it had not been considered for patient characterization regarding discharge (hospitalization, primary care, or home). In a preliminary assessment, both metabolites, Hb and CRP, extracted guidance in anemia and inflammation, respectively. In our approach, we verified a different distribution in Hb and CRP in comparing discharge groups, illustrated with a boxplot. Specifically, the ratio CRP/Hb was higher in hospitalized patients than in those discharged to primary care or home. We then applied the CRP/Hb ratio to a ROC analysis and accomplished an 80% sensitivity and a 68% specificity in predicting hospitalization versus discharge to primary care or home. This finding demonstrated that anemia and inflammation were not the only outcomes predicted by Hb and CRP, respectively. Indeed, we supervised the relationship between these features across a wide range of categories, including gender, age group, and diagnosis codification (using ICD-10 as a reference). While none of these categories were considered for the correlation model between Hb

and CRP, their specificity leveraged additional findings addressed to the groups of interest. For instance, when comparing disease codification, we confirmed that the correlation between both proteins was only significant for patients with nervous and respiratory diseases. While it has limitations, namely the small sample size in each group, the study demonstrates that routine blood protein measurements are of merit in creating metrics for other types of clinical inferences.

Future work should address this issue to improve the relevance of potential associations. For instance, in our study, we approached a cohort representative from an emergency department. We confirmed that hospitalized patients had a higher tendency to develop anemia due to higher levels of CRP. Studies in patients following hemodialysis treatment suggested a more robust correlation (r_s = -0.607, p = 0.001), associating this link (Hb and CRP levels) with increased mortality (Heidari et al., 2015). Of interest, this correlation is only marginal (rs = -0.025, $p < 0.001$) in healthy individuals (Ziv-Baran et al., 2020). Thus, this framework might achieve other associations with potential inferential insights depending on the outcome of interest.

With this study, we demonstrated the potential of inferential statistics for patient characterization by using Hb and CRP as enablers of clinical information that was latent in clinical blood analysis rather than only categorizing for anemia or inflammation. As part of standard practice, this approach inserted in traditional clinical analysis systems should be continuously used to provide additional inferences related to other correlations and associations in specific clinical contexts. Those associations drive further investigations that lead to a holistic view of blood analysis, selecting features that may be important to measure over others in a predefined context.

In **Chapter 2.2**, we confirmed how AI and machine learning enhanced the value of routine blood parameters by using them to predict disease diagnosis and prognosis. We manually filtered the AI-driven studies related to diagnosis that used only routine blood parameters (specified according to routine testing in primary and point-of-care settings); we eliminated the remaining ones that were part of a broad section related to imaging and omics, which have been the growth engines of AI in medical applications. With this, we aimed to summarize the value of these metabolites for predicting pathologic states in the scope of AI and compare this additional value with the established domain knowledge in which traditional clinical decision support systems are applied. We compared studies on ten disease classes (using the ICD-10 as a reference), considering the outcome of interest, sample size, selected features, methods, and principal findings.

We found a significant gap between standard disease-associated features and those selected by machine learning models. We verified this gap by using a variety of supervised learning algorithms, including linear, bayesian, nearest neighbors, support vector machines, decision trees, neural networks, or ensembles that produced feasibility models for the disease classes studied, ranging from infections (or parasitic diseases), neoplasms, blood, endocrine (nutritional or metabolic), mental (behavioral or neurodevelopmental), circulatory, respiratory, digestive, genitourinary, and particular diseases (covid-19). Clinically, Hb is a biomarker of anemia, blood loss, and thalassemia for values below normality; it is a marker of heart disease, pulmonary fibrosis, smoking, polycythemia vera, kidney disorders, or myeloproliferative diseases in above-normality concentrations. Enhanced by AI, Hb is a valuable feature to model surgical site infection, colorectal cancer, pediatric acute lymphoblastic leukemia, women's iron deficiency anemia, thalassemia minor, β-thalassemia, diabetes polyneuropathy risk, ischemic stroke, smoking status in smokers, HBV-induced cirrhosis, chronic kidney disease, covid-19 diagnosis, and also relevant for the prognosis of lymphocytic leukemia treatment within two years, and of covid-19. In this review, AI-driven studies demonstrate their ability to predict other outcomes in addition to the potential predefined (i.e., already known in clinical practice) causes of Hb variations. Aside from the previously described enhancement of metabolites related to disease prediction, we found consistency in model development. Generally, authors computed independent datasets to implement AI algorithms and followed a similar framework built with feature selection prior to internal validation. External evaluation occurred in only 19% of the studied reports, which we identify as the main shortcoming of this approach in technology deployment.

Future research should address this issue as a critical aspect of the future of AI-driven applications for clinical diagnosis or prognosis.

It is important to highlight that deploying developed models could probably be easier with the alignment between data scientists and physicians to merge accessibility, interpretability, feedback, and the model's continuous testing in different settings. Nevertheless, we consider the panels studied as low risk regarding representativeness and bias since they are measured with highly standardized laboratory equipment and a tremendous amount of available data points since we only consider routine blood analysis. In this way, we valued studied models and argue that they could be enhanced by machine learning, with additional clinical value for the physician and the patient. Employing other learning strategies, such as deep learning, and the standardized guidance of machine learning frameworks, i.e., TRIPOD-ML, will create a deeper environment wherein the

model's accuracy must be balanced with clinical understanding. In the future, these models should be applied to longitudinal data, with a deeper understanding of patient profiles, particularly when used early in patients' diagnosis. Ultimately, as previously proposed, It may be feasible to develop a real-time blood augmentation diagnostic tool to provide a landscape view of each individual's profile and compare it to specific disease patterns (Rossi & Grifantini, 2018). A feasibility study of this approach was also developed in this dissertation, as is discussed next.

In the third study, described in **Chapter 2.3**, we modeled digestive tract tumors by applying a cross-decomposition algorithm (principal components analysis) in a dataset containing data on the Hb concentration, erythrocyte count, hematocrit levels, and red blood cell distribution width (RDW). From a methodological point-of-view, this study contrasts with the first by incorporating the algorithm into the data, driving findings without pre-established assumptions or labels related to pathological states (Sarker, 2022). Indeed, the algorithm computed the normalized values of the studied metabolites and maximized the variance between disease groups through the linear merging of the features, defining new latent variables (principal components) that explained scores distribution. Tumors were clustered according to their localization in the digestive tract, separating C15 esophagus from C16 stomach and those from C18 colon and C19 rectum, consistent with similar pattern recognition studies that evaluated the miRNA profiles of digestive carcinomas (Tang et al., 2016). Furthermore, the strong association between erythrocytes, Hb, and hematocrit with the C18 colon was consistent with the second study's findings, indicating that hematological parameters can predict the diagnosis of digestive tract malignancies.

We used a multivariate analysis based on a binomial logistic regression to supplement this methodological approach. The resulting models discriminated C18 colon from the rest with an AUC of 0.77, followed by C16 stomach with an AUC of 0.68. Despite having similar patterns in the PCA, the C18 colon and C19 rectum's discrimination performance was noticeably different, proving that similar patterns cannot be predicted until feature correlation maintains higher correlation values with the outcome. Even though these models are still far from absolute tumor site differentiation, either due to a lack of sample size, a single-center nature, or only referring to patients analyzed in the codification date, they support previous studies in which hematological blood parameters contain unappreciated information that could be used to model pathological states in predefined timeframes and spatial contexts.

With this dissertation, we aimed to describe a methodology that could endorse these findings and eventually lead to comparative clinical trials to achieve real-world evidence. However, because this level of testing was not feasible, the described methodologies were used to contextualize data mining strategies according to the available clinical data. Outcomes such as screening, diagnosis, or prognosis are served from these tools to provide better care for general populations. Our emphasis on routine blood analysis may have limited study results regarding performance measures; still, it supported the direction of future research that can offer deeper insights without the need for centralized laboratory facilities. This viewpoint is backed by the advancement of deep learning algorithms, data availability, and, most significantly, the medical community's desire to succeed with these tools (Bienefeld et al., 2023). We also anticipate that adding longitudinal profiles, intrinsically tailored for each patient, would endorse model's performance. As we studied tumor profiles by analyzing nearly 180 patients only once, longitudinal profiles will be able to describe the progression of each patient, comparing data with his past values and with the overall cohort, allowing us to model disease status more robustly. Importantly, algorithms will be able to learn the 'normal' values for each individual and highlight their numerical evolution, both for each metabolite or together as a pattern (Gunčar et al., n.d.). Ultimately, available databases containing this information will allow for a richer understanding of individuals' states at the point-of-care level thanks to the constrained input data necessary.

Future studies should aggregate data to construct more robust prediction models that can be evaluated in preclinical settings. Future work should investigate these methodologies to extract more clinical value for each routine blood analysis and the usage of particular blood panels for tracking disease outcomes. The integration of artificial intelligence with routine blood analysis allows direct assessment of data patterns which can be furtherly evaluated in terms of statistical analysis, as described in the first experimental work. The models developed could be integrated with pilot systems for preclinical and clinical evaluations, especially in point-of-care.

Herein, we demonstrated different approaches to how routine blood analysis could display additional clinical value, markedly improved by integrating inferential statistics or artificial intelligence technologies. We describe them as non-mutually exclusive technologies, as they complement each other depending on the data set, structure, information, and outcome to be analyzed. We must continue to collect data for training with more complex AI algorithms for studying human disease and progression at the point-of-care or at home.

CONCLUSION AND FUTURE PERSPECTIVES

In this dissertation, we studied routine blood tests in combination with three different data mining strategies to extract clinical information of interest to physicians in different stages of patient care. In the three studies, we analyzed data of acute patients in the emergency department, patients undergoing machine learning studies for diagnosis or prognosis, and patients' blood analysis at the time of diagnosis (codification date). Overall, we found merit in routine blood parameters for creating metrics for other clinical inferences. Inferential statistics described a prevalent correlation between CRP and Hb, and its ratio was higher in hospitalized patients, distinguishing those from the ones discharged to primary care or home, assessing also discharge rather than only inflammation or anemia.

AI-driven studies disclosed an inequality between traditional disease-associated features and those identified by ML for predicting diagnosis or prognosis. In particular, Hb is a valuable feature for modelling various diseases and pathologic conditions when enhanced by AI. We also confirmed distinct patterns built with erythrocytes, Hb, hematocrit, and RDW in digestive tract carcinomas, which clustered tumors according to the anatomical location of tumor development. Furthermore, in a separate test set examination, those blood tests acted as biomarkers for distinguishing C18 colon from the others, with promising accuracy metrics.

We consider that using both pattern recognition and machine learning has more potential to extract clinical insights in comprehensive datasets than finding restricted associations manually. However, as they are not mutually exclusive, AI-driven results could be evaluated by statistics (as described in the third study), allowing higher explainability and comprehension regarding developed models. Nonetheless, developments based on AI are becoming strategic for analyzing large amounts of multimodal data and establishing new guidance on the importance of blood analysis at specific moments of patient care, opposing rule-based systems wherein 'one-size-fits-all'.

Future directions should include the development of AI-driven models in longitudinal data collected at the point-of-care, where access to differentiated diagnosis is limited. Indeed, we believe that AI will remarkably impact the future of clinical decision support systems, which we envision to be a provider of better care by suggesting a 'second opinion', and an enabler of personalized, preventive, and accessible source of diagnosis.

REFERENCES

Bienefeld, N., Boss, J. M., & Lüthy, R. (2023). Solving the explainable AI conundrum by bridging clinicians ' needs and developers ' goals. *Npj Digital Medicine*, 1–7. https://doi.org/10.1038/s41746-023-00837-4

Guenter, Weiss. Lawrence T., G. (2005). Anemia of chronic disease. *The New England Journal of Medicine*, 1011–1023. https://doi.org/10.1002/9781119252863.ch12

Gunčar, G., Kukar, M., Notar, M., Brvar, M., & Černelč, P. (n.d.). *Application of machine learning for hematological diagnosis*. 1–15.

Heidari, B., Fazli, M. R., Misaeid, M. A. G., Heidari, P., Hakimi, N., & Zeraati, A. A. (2015). A linear relationship between serum high-sensitive C-reactive protein and hemoglobin in hemodialysis patients. *Clinical and Experimental Nephrology*, *19*(4), 725–731. https://doi.org/10.1007/s10157-014-1048-0

Rossi, R. L., & Grifantini, R. M. (2018). Big Data: Challenge and Opportunity for Translational and Industrial Research in Healthcare. *Frontiers in Digital Humanities*, *5*(May), 1–13. https://doi.org/10.3389/fdigh.2018.00013

Sarker, I. H. (2022). AI-Based Modeling: Techniques, Applications and Research Issues Towards Automation, Intelligent and Smart Systems. *SN Computer Science*, *3*(2), 1–20. https://doi.org/10.1007/s42979-022-01043-x

Tang, S., Wu, W. K. K., Li, X., Wong, S. H., Wong, N., Chan, M. T. V., Sung, J. J. Y., & Yu, J. (2016). Stratification of Digestive Cancers with Different Pathological Features and Survival Outcomes by MicroRNA Expression. *Scientific Reports*, *6*(March), 1–10. https://doi.org/10.1038/srep24466

Ziv-Baran, T., Wasserman, A., Goldiner, I., Stark, M., Shenhar-Tsarfaty, S., Shapira, I., Zeltser, D., Mailis, I., Berliner, S., & Rogowski, O. (2020). The association between C-reactive protein and common blood tests in apparently healthy individuals undergoing a routine health examination. *Clinica Chimica Acta*, *501*(November 2019), 33–41. https://doi.org/10.1016/j.cca.2019.12.002