



Advancing Laboratory Medicine Practice With Machine Learning: Swift yet Exact

Jiwon You , M.S.¹, Hyeon Seok Seok , B.S.E.², Sollip Kim , M.D., Ph.D.³, and Hangsik Shin , Ph.D.¹

¹Department of Digital Medicine, Brain Korea 21 Project, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea; ²Department of Biomedical Engineering, Graduate School, Chonnam National University, Yeosu, Korea; ³Department of Laboratory Medicine, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea

Machine learning (ML) is currently being widely studied and applied in data analysis and prediction in various fields, including laboratory medicine. To comprehensively evaluate the application of ML in laboratory medicine, we reviewed the literature on ML applications in laboratory medicine published between February 2014 and March 2024. A PubMed search using a search string yielded 779 articles on the topic, among which 144 articles were selected for this review. These articles were analyzed to extract and categorize related fields within laboratory medicine, research objectives, specimen types, data types, ML models, evaluation metrics, and sample sizes. Sankey diagrams and pie charts were used to illustrate the relationships between categories and the proportions within each category. We found that most studies involving the application of ML in laboratory medicine were designed to improve efficiency through automation or expand the roles of clinical laboratories. The most common ML models used are convolutional neural networks, multilayer perceptrons, and tree-based models, which are primarily selected based on the type of input data. Our findings suggest that, as the technology evolves, ML will rise in prominence in laboratory medicine as a tool for expanding research activities. Nonetheless, expertise in ML applications should be improved to effectively utilize this technology.

Key Words: Artificial intelligence, Clinical laboratory tests, Laboratory medicine, Machine learning

Received: July 8, 2024

Revision received: September 1, 2024

Accepted: October 25, 2024

Published online: November 26, 2024

Corresponding author:

Hangsik Shin, Ph.D.
Department of Digital Medicine,
Brain Korea 21 Project, Asan Medical
Center, University of Ulsan College of
Medicine, 88 Olympic-ro 43-gil, Songpa-gu,
Seoul 05505, Korea
E-mail: hangsik.shin@amc.seoul.kr



© Korean Society for Laboratory Medicine

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

In recent decades, machine learning (ML) has significantly advanced in terms of analytical and predictive capabilities, establishing itself as a vital tool across various fields. Developments in big data and high-performance computing have significantly improved the performance of ML algorithms, thereby enabling more effective methods for addressing complex challenges. The ability of ML to analyze large datasets and identify patterns can assist clinicians in diagnosis and prediction of clinical outcomes. ML applications have been investigated in various areas, including medical-image analysis, patient prognosis, and personalized

treatment planning. A few models have been approved by the Food and Drug Administration, commercialized, and implemented in clinical practice [1, 2].

Additionally, ML has been investigated in laboratory medicine [3-5] to reduce errors and enhance the accuracy and reliability of test results. ML processes or analyzes large datasets, which facilitates the extraction of meaningful information that would otherwise require extensive manual effort. For example, ML has improved the efficiency of repetitive or manually-intensive tasks, such as validating general chemistry test results or analyzing blood cells and urine cultures [6, 7]. Owing to its inference and big data analytical capabilities, ML can substantially enhance

laboratory medicine by effectively managing diverse data types frequently analyzed in healthcare.

In this review, we comprehensively assessed the current state of ML applications in laboratory medicine. We explored the major uses of ML, the data types processed, the results obtained, and the characteristics and considerations for implementing major ML models. Based on these findings, we also examined existing research challenges and identified potential future developmental trends.

RESEARCH METHODOLOGY

Literature search and screening

We searched PubMed for original articles that utilized ML in laboratory medicine and were published between February 2014 and March 2024. The search string was generated by combining words related to laboratory medicine with keywords related to ML and excluding unrelated topics (e.g., coronavirus disease 2019 [COVID-19], genome, magnetic resonance imaging, computed tomography, ultrasound, electrocardiography, and electroencephalography). The search strategy is detailed in the Supplemental Data. We initially retrieved 779 articles. A clinical pathologist first excluded articles outside the scope of laboratory medicine based on their title and abstract. Articles meeting the primary screening were subjected to a full-text review. The exclusion criteria in the secondary screening included: (i) data were not used in a clinical laboratory test process or did not originate from a laboratory test; (ii) data were unrelated to the primary duties of the laboratory; (iii) laboratory results served solely for disease prediction; (iv) the ML model used was unspecified; (v) the full text was unavailable; (vi) the article was not written in English; and (vii) the article failed to present original research. When results were ambiguous during the secondary screening, a clinical pathologist reviewed the full text to ascertain its eligibility. Finally, 144 articles were selected for our review.

Literature analysis

The selected articles were categorized into laboratory medicine subspecialties based on criteria specified in a laboratory medicine textbook [8]: diagnostic hematology, clinical chemistry, clinical microbiology, molecular diagnostics, transfusion medicine, and diagnostic immunology. The full text of each article was analyzed, and the research objectives, specimen types, data types, ML models, evaluation metrics, and sample sizes were summarized. Research objectives were categorized as “recognition” for identifying specific entities or performing binary classifications,

“classification” for categorizing into three or more groups, and “counting” for quantifying elements such as cell counts.

“Specimen type” refers to the type of specimen used as ML input data and was classified by referring to the specimen type list in the Logical Observation Identifiers Names and Codes [9]. “Data type” refers to the type of material associated with the input data and was categorized into image, table, sequence, and other types according to common classifications used in ML-based studies. We included all ML models used for analysis and comparative evaluation; however, customized models were described as base models. In such cases, popular models (e.g., You Only Look Once [YOLO]) [10] were described using their respective names.

“Evaluation metrics” included all metrics used to assess performance, excluding lesser-known metrics not typically used in ML studies. “Sample size” describes the total number of samples inputted into an ML model, regardless of sample type.

Sankey diagrams were used to analyze the application trends of ML models in laboratory medicine in terms of research objectives or data types. Where further comprehensive analysis of the proportion of each factor was necessary, we used pie charts, which display proportions intuitively, to aid in understanding the relative importance of each item. We created Sankey diagrams using a web-based visualization tool (SankeyMATIC [11]) and created pie charts using the Matplotlib package in Python (version 3.12.1) [12].

SUMMARY OF THE LITERATURE REVIEW RESULTS

Key characteristics for categorizing ML-based laboratory medicine research

Table 1 summarizes the key points of our literature review of ML in laboratory medicine, including the research objectives, specimen types, data types, ML models, and evaluation metrics described in each article. The research objectives were categorized into 12 topics: autoverification, classification, clinical decision support (CDS) for laboratories, counting/enumeration, disease screening, error detection, estimation/prediction, recognition, tools based on artificial intelligence (AI), data generation/process simulation, ML optimization, and preprocessing assistance. The number of input samples used for ML modeling varied widely among studies, i.e., 5–25 million, and was not clearly reported in some cases.

The main evaluation metrics used were accuracy, sensitivity, specificity, and area under the ROC curve (AUROC). Accuracy re-

Table 1. Categorical summary of key points from the literature review of studies applying ML in laboratory medicine from February 2014 to April 2024

Laboratory medicine field	Research objective	Specimen type	Data type	ML model	Evaluation metric and performance
<ul style="list-style-type: none"> Clinical chemistry Clinical microbiology Diagnostic hematology Diagnostic immunology Molecular diagnostics Transfusion medicine 	<ul style="list-style-type: none"> Autoverification Classification CDS for laboratories Counting/enumeration Disease screening Error detection Estimation/prediction Recognition Tools based on AI Others <ul style="list-style-type: none"> Data generation/process simulation Machine learning Optimization Preprocessing assistant 	<ul style="list-style-type: none"> Blood <ul style="list-style-type: none"> Blood image WBC image Blood cell image RBC image CGM data CBC data Bone marrow Plasma Urine <ul style="list-style-type: none"> Urine sample Urine micrograph image Urine culture image Others <ul style="list-style-type: none"> Bacteria Antibiogram Sperm Stool 	<ul style="list-style-type: none"> Image Sequence Tabular <ul style="list-style-type: none"> Numeric Category Text 	<ul style="list-style-type: none"> CNN DNN DT MLP LR RF RNN SVM XGB Others <ul style="list-style-type: none"> CatBoost CNN+LSTM DBN Ensemble HCA KNN LLM PLS-DA UMAP 	<ul style="list-style-type: none"> AC AUROC SE SP PPV NPV F1 score FNR MSE MAE R² RMSE

Abbreviations: AC, accuracy; AI, artificial intelligence; AUROC, area under the ROC curve; CBC, complete blood count; CDS, clinical decision support; CGM, continuous glucose monitoring; CNN, convolutional neural network; DBN, deep belief network; DNN, deep neural network; DT, decision tree; FNR, false-negative rate; HCA, hierarchical cluster analysis; KNN, *k*-nearest neighbor; LLM, large language model; LR, logistic regression; LSTM, long short-term memory; MAE, mean absolute error; ML, machine learning; MLP, multilayer perceptron; MSE, mean squared error; NPV, negative predictive value; PLS-DA, partial least squares-discriminant analysis; PPV, positive predictive value; R², coefficient of determination; RBC, red blood cell; RF, random forest; RMSE, root mean squared error; RNN, recurrent neural network; SE, sensitivity; SP, specificity; SVM, support vector machine; UMAP, uniform manifold approximation and projection; WBC, white blood cell; XGB, extreme gradient boosting.

fers to the percentage of correct predictions, whereas sensitivity (also known as recall) measures the ability of a model to identify true positives. Specificity measures a model's ability to identify true negatives, which is useful when correct classification of negative cases is necessary. The AUROC is commonly used to comprehensively evaluate classification model performance and represents the true and false positive rates at different thresholds. Mean squared error (MSE) is an important metric in regression analysis and refers to the average of the squared differences between predicted and actual values; a lower MSE indicates greater accuracy. The mean absolute error and coefficient of determination were also used to evaluate performance. Rarely used metrics included the relative distance error for contour-based measures and the mean structural similarity index for evaluating image similarities. Further details of our literature analysis are provided in Supplemental Data Table S1.

Applications of ML in laboratory medicine domains

The Sankey diagram in Fig. 1 illustrates relationships among representative laboratory medicine fields, the main objectives of

using ML, and the best-performing ML models. The recognition, classification, and counting/enumeration categories were classified under “detection” because they serve similar purposes. Before the classification, these categories accounted for 24.3%, 23.6%, and 4.7% of the overall objectives, respectively. Among the six laboratory medicine fields, diagnostic hematology was the most actively investigated area in terms of ML, representing 48.6% of all studies evaluated in this review. Clinical chemistry ranked second (28.5%), followed by clinical microbiology (15.3%). Molecular diagnostics, transfusion medicine, and diagnostic immunology each constituted <3% of the total number, indicating that ML utilization is low in these areas. ML was primarily used for detection in diagnostic hematology, constituting 70% of all applications. The next most common use of ML was disease screening, representing 15% of the studies. Although various other purposes were reported for molecular diagnostics, transfusion medicine, and diagnostic immunology studies, no reports documented the application of ML for error detection. Conversely, >50% of ML applications in clinical chemistry focused on error detection or estimation/prediction.

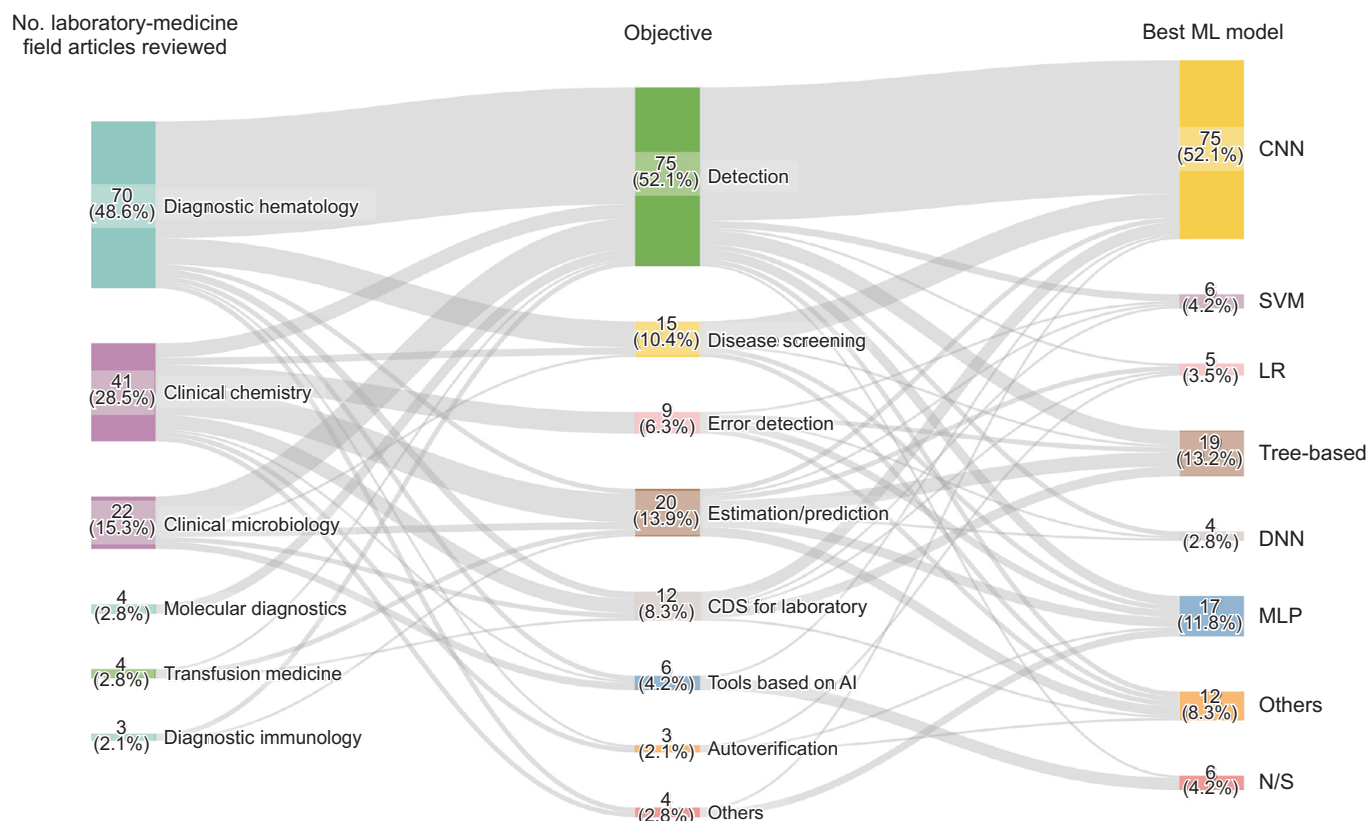


Fig. 1. Sankey diagram showing the relationships among representative laboratory medicine fields, their main objectives, and ML models identified through a literature review of studies applying ML in laboratory medicine from February 2014 to April 2024.

Abbreviations: AI, artificial intelligence; CNN, convolutional neural network; DNN, deep neural network; LR, logistic regression; ML, machine learning; MLP, multilayer perceptron; N/S, not specified; SVM, support vector machine.

Notably, all ML-based error detections were performed in the field of clinical chemistry. In clinical microbiology, detection accounted for >50% of the ML applications, but error detection was not included. In addition, estimation/prediction, CDS for laboratories, and tools based on AI were reported in these studies. In molecular diagnostics, ML was only applied for detection purposes. Conversely, in transfusion medicine and diagnostic immunology, ML was primarily applied for detection and estimation/prediction. Studies using ML in molecular diagnostics, transfusion medicine, and diagnostic immunology were uncommon (<5 studies each), making generalizations difficult.

Analysis of the ML models used for clinical laboratory testing showed that convolutional neural networks (CNNs), multilayer perceptrons (MLPs), and tree-based models were used in 77% of all studies. For detection, approximately 70% of the studies used CNNs, whereas the remainder used support vector machine (SVM), tree-based, MLP, or other models. Similarly, CNNs were prominently used in disease screening, i.e., in approximately 70% of the studies, owing to advantages in image analy-

sis, such as the ability to recognize and classify specific cell types (e.g., white blood cells [WBCs] and red blood cells [RBCs]) from specimen images.

Tree-based models were most frequently used for estimation/prediction, comprising approximately 30% of the total. The “tools based on AI” category refers to studies that primarily evaluated the performance of AI-based models and typically did not specify the model type. Fig. 2 depicts the percentages of ML models used in each laboratory medicine field. As shown in Fig. 2A, CNNs were most commonly used in diagnostic hematology, constituting 80% of all studies investigated, largely owing to their application in analyzing blood cells and images for recognition and classification.

The most diverse range of models used was found in clinical chemistry (Fig. 2B), reflecting a broader range of purposes than in other fields (Fig. 1). Tree-based models, such as random forest (RF), extreme gradient boosting (XGB), and CNN models were primarily used in clinical microbiology (Fig. 2C). Although the effectiveness of these models has not been confirmed

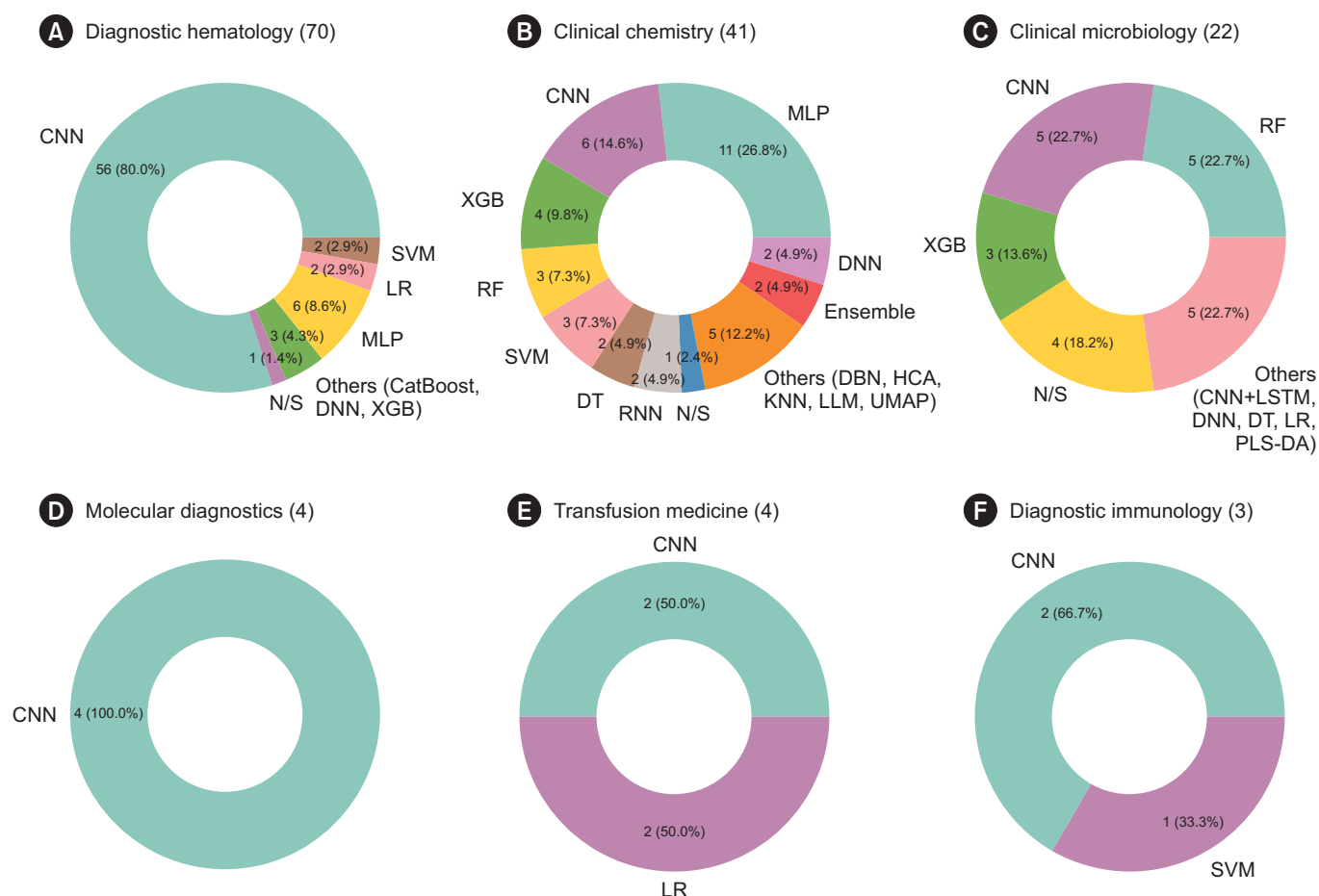


Fig. 2. Pie charts showing the proportions of ML models used in various laboratory medicine fields based on a literature review of laboratory medicine studies involving ML applications from February 2014 to April 2024. Numbers in parentheses indicate the number of published articles related to ML in each field. The frequencies at which various ML models were used in (A) diagnostic hematology, (B) clinical chemistry, (C) clinical microbiology, (D) molecular diagnostics, (E) transfusion medicine, and (F) diagnostic immunology are shown.

Abbreviations: CNN, convolutional neural network; DBN, deep belief network; DNN, deep neural network; HCA, hierarchical cluster analysis; LLM, large language model; LR, logistic regression; LSTM, long short-term memory; ML, machine learning; MLP, multilayer perceptron; N/S, not specified; PLS-DA, partial least squares-discriminant analysis; RF, random forest; RNN, recurrent neural network; SVM, support vector machine; UMAP, uniform manifold approximation and projection; XGB, extreme gradient boosting.

through multiple studies, researchers are actively investigating their potential. CNN, logistic regression (LR), and SVM models have been used in molecular diagnostics, transfusion medicine, and diagnostic immunology; however, significant trends could not be observed because of the limited number of relevant studies (Fig. 2D–F).

The following sections present representative use cases of ML in each laboratory medicine field.

Clinical chemistry

In clinical chemistry, ML has been applied to predict physiological and biochemical parameters, such as blood glucose levels [13, 14], clinical lipid concentrations [15], and urine culture re-

sults [16], with an emphasis on prediction and error detection. Blood glucose prediction studies using continuous glucose monitoring (CGM) data have demonstrated accuracy rates exceeding 90% in predicting type 1 diabetes using neural network models, such as CNNs, MLPs, and deep neural network (DNNs), as well as long short-term memory [13, 14]. ML has been used to detect errors in clinical laboratory test results, including wrong blood in tubes, sample labeling, and sample contamination, which may occur during clinical laboratory testing [17, 18]. Several studies have evaluated the feasibility of using ML models for the autoverification of clinical laboratory test results [19, 20], whereas other studies used ML for preprocessing and workflow improvement to increase the efficiency of clinical laboratory test-

ing [21, 22]. ML models used for validating clinical laboratory test errors include neural networks (e.g., CNNs, DNNs, and MLPs), tree-based algorithms (e.g., RFs and XGB), and statistical analysis-based techniques (e.g., SVMs and LR). Additionally, ML models have been applied to interpret thyroid function and urinary steroid profiles [23-25] and to recognize and classify specific cells and structures in medical images, urine, and blood samples [26-28].

Diagnostic hematology

In diagnostic hematology, ML has been primarily used to recognize or classify blood cells in blood images, with a focus on extracting characteristics from WBC images, diagnosing blood-related diseases, such as leukemia [29-31], or classifying different types of blood cells [32-35]. For example, ML has been utilized to recognize sickle-shaped RBCs [36-38] and count cells [39-41]. One study used a generative adversarial network to generate images of leukemia cells [42].

Transfusion medicine

In transfusion medicine, ML has been used to assess the appropriateness of blood for transfusions or to analyze the blood information required by testing the antigens present. For example, ML has been used to analyze hemoglobin and iron contents in blood to prevent iron overload during transfusions [43] or for ABO blood typing [44].

Clinical microbiology

Most clinical microbiology studies focused on bacteria and urine culture interpretations. ML models have been used to identify the main bacterial species causing urinary tract infections in urine samples to prevent delays in antibiotic treatment [45], classify bacterial strains [46], interpret antibiotic susceptibility test images [47], or classify colonies of bacterial species, such as *Escherichia coli* and *Staphylococcus aureus* [48]. Some studies have evaluated the accuracy of a commercialized AI-based urine culture interpretation system known as Automated Plate Assessment System (APAS; LBT Innovations, Adelaide, Australia) [49, 50]. The application of CNNs for image analysis in clinical microbiology has been investigated. LR, RF, and SVM models have been used to analyze urine culture results and clinical information of patients with urinary calculus for antibiotic dosing management [51].

Diagnostic immunology

In diagnostic immunology, most studies analyzed the patterns of

HEp-2 cells, which serve as a diagnostic biomarker for autoimmune diseases [52, 53]. An automatic immunofluorescence pattern classification framework that uses CNNs to detect HEp-2 cell features was proposed and demonstrated to be useful to reduce manual errors and efficiently classify large amounts of data.

Molecular diagnostics

In molecular diagnostics, ML has been used to study chromosomes in karyotyping, including chromosome detection and localization [54], diagnosing hematologic neoplastic cells by karyotyping cancer cell chromosomes [55], and detecting circulating tumor cells in blood samples [56]. CNNs were applied in most of these studies.

Developing ML models for laboratory-medicine practice

Fig. 3 depicts the relationships among the publication year, best ML model, and input data used. ML models are increasingly being used (Fig. 3). MLPs were the first to be adopted in 2014 and remained the most frequently used models until 2016, after which their usage began to decline. Since their respective introductions in 2016 and 2018, the implementation of CNNs and tree-based models has increased. While they remain the most widely used models, the growing utilization of other models is indicative of concerted efforts to diversify the range of models as ML evolves.

Image data accounted for the largest portion, approximately 60%, of the data types used. Among ML studies that used image data, 85% employed CNNs because of their advantages in processing image data (Fig. 3). Various ML models except CNNs have been applied to analyze tabular data. To analyze sequence data, only DNN and tree-based models have been used. Fig. 4 illustrates the basic principles of commonly used ML models in laboratory medicine.

LR model

LR is typically used to solve binary classification problems [57-59]. LR uses logits to calculate the probability that a dependent variable belongs to a particular class, producing an output value between 0 and 1 (Fig. 4A). The prediction function for LR is presented in Eq. (1).

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}} \quad (1)$$

where p represents the probability that the dependent variable y equals 1 ($P(Y=1)$), β_1 is the regression coefficient for the inde-

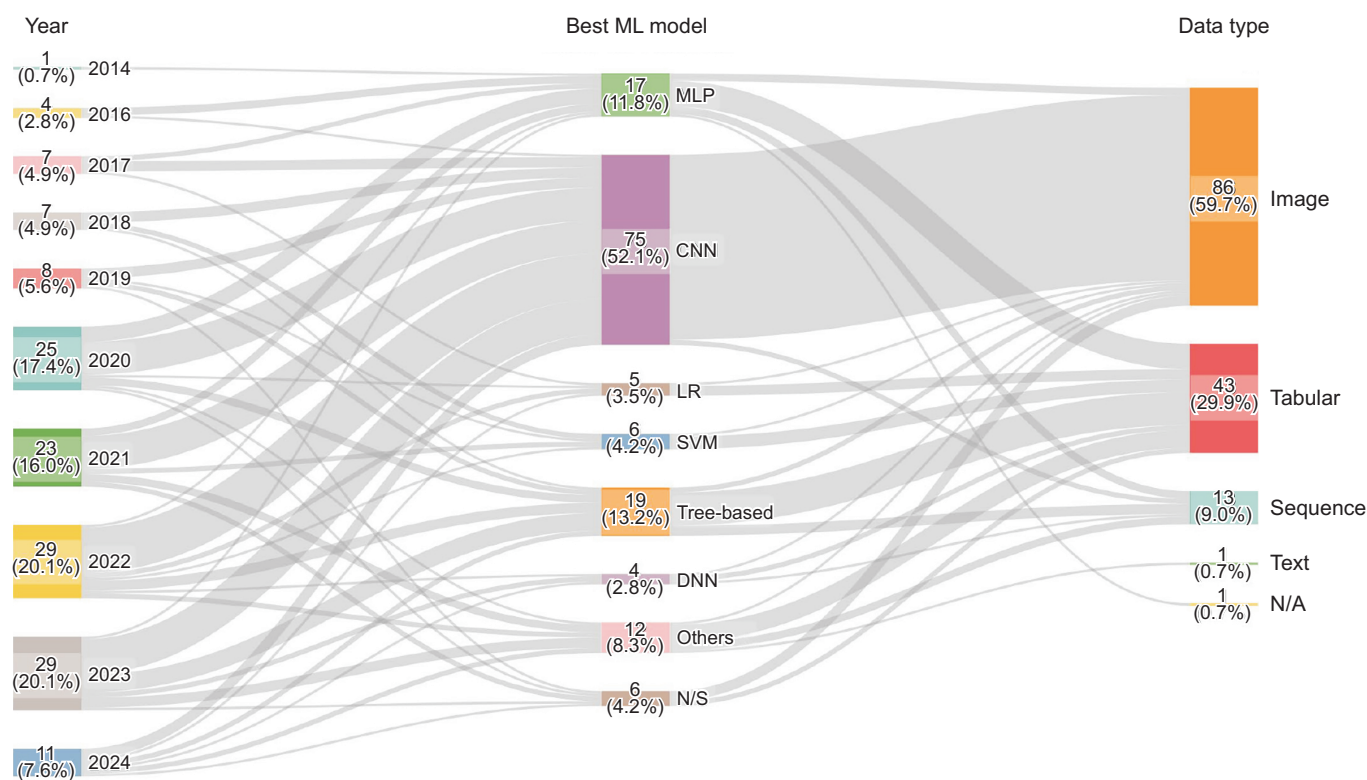


Fig. 3. Sankey diagram showing relationships among the year, best ML models, and data type based on a literature review of studies applying ML in laboratory medicine from February 2014 to April 2024.

Abbreviations: CNN, convolutional neural network; DNN, deep neural network; LR, logistic regression; ML, machine learning; MLP, multilayer perceptron; N/A, not applicable; N/S, not specified; SVM, support vector machine.

pendent variable x , and β_0 is the Y-intercept. LR not only predicts class labels in classification problems but also generates the probability that a dependent variable belongs to a particular class, providing confidence that a prediction can be expressed as a probability. However, LR cannot perform classifications easily with nonlinear data.

SVM model

SVMs represent conventional supervised learning models used for pattern recognition, data analysis, classification, and regression analysis. An SVM first selects a hyperplane (Fig. 4B) that maximizes the margin between classes [60–62]. Subsequently, for each data point x_i , the model identifies a weight w and bias b that satisfy the following condition: $y_i(w \cdot x_i + b) \geq 1$, where y_i is the class label (+1 or -1) of data point x_i . SVMs are widely used in classification problems and are robust against outliers, rendering them resistant to overfitting. However, their computational load increases exponentially with dimensionality, posing challenges in managing large datasets. Their decision boundaries are linearly constrained.

MLP model

MLPs, also known as feed-forward neural networks [62, 63], comprise one or more hidden layers that are fully connected between input and output layers (Fig. 4C). Each layer consists of multiple nodes that each perform computation by multiplying and summing the outputs of the previous layer with weights. Subsequently, an activation function is applied to these values to determine the final output value of the node. The output of node j , z_j , is derived using Eq. (2).

$$z_j = \sum_{i=1}^n w_{ij} \cdot a_i + b_j \quad (2)$$

where w_{ij} is the weight connecting node i in the previous layer to the current node j , a_i the output of node i in the previous layer, and b_j is the bias of node j . Although MLPs can solve complex nonlinear problems, the number of weight coefficients may increase exponentially with model complexity, leading to overfitting of the training data.

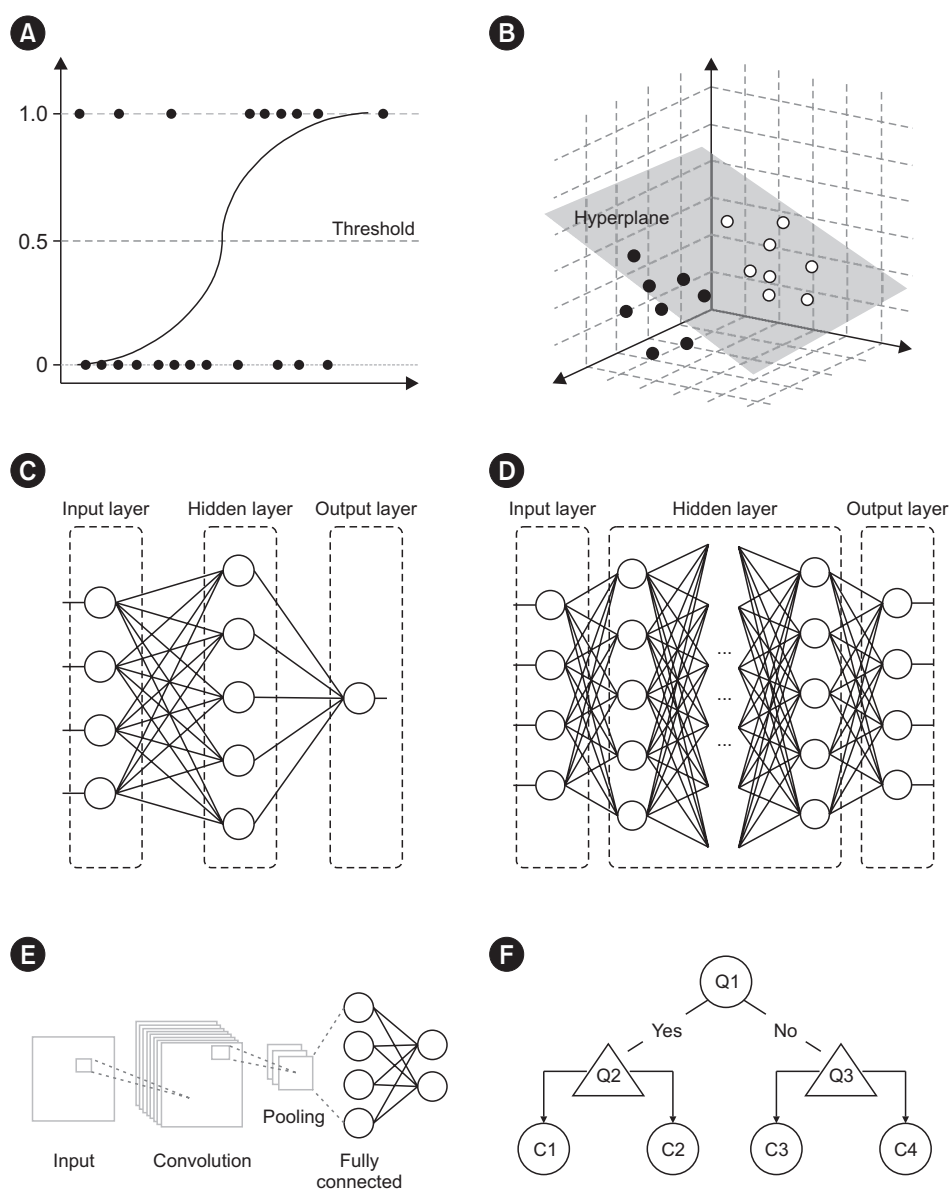


Fig. 4. Features of different ML models. (A) LR based on the sigmoid function, expressed as a probability value between 0 and 1, divided by the threshold. (B) An example of sample classification using a hyperplane and an SVM. (C) An MLP comprising an input layer, a hidden layer, and an output layer composed of connected perceptrons. (D) A DNN comprises more hidden layers than an MLP and is an extension of an MLP. (E) A CNN comprises convolution layers and is primarily used for image processing. (F) A DT-based model follows decision rules in a tree structure.

Abbreviations: C, class; CNN, convolutional neural network; DNN, deep neural network; DT, decision tree; LR, logistic regression; ML, machine learning; MLP, multilayer perceptron; SVM, support vector machine; Q, question.

DNN model

A DNN is an extension of the MLP that comprises multiple (typically three or more) hidden layers between the input and output layers (Fig. 4D) [63, 64]. The increase in the number of hidden layers enhances the efficiency of the model in learning patterns in data, enabling it to solve more complex nonlinear problems. A key advantage of DNNs is their ability to automatically extract features from data, eliminating the need for manual extraction. However, this capability requires substantial data and computational resources, and similar to MLPs, they are susceptible to overfitting.

CNN model

A CNN is an artificial neural network containing convolutional layers and is typically utilized for image or sequence data processing because it can capture spatiotemporal features [64-67]. CNNs generally comprise input and output layers, along with multiple hidden convolutional layers connected with pooled, fully connected layers to generate the output (Fig. 4E). The convolution operation is mathematically expressed in Eq. (3).

$$(X * K)(i, j) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} X(i + m, j + n) \cdot K(m, n) \quad (3)$$

where X is the two-dimensional input (e.g., image), K is the filter,

(i, j) is the two-dimensional output index, and M and N are the height and width of the filter, respectively. The filter (also known as the mask or kernel) is a matrix of numbers used in the convolution operation. CNNs have been used as the fundamental architecture of various models, including the ResNet, YOLO, and AlexNet models because they can preserve spatial information and process images through convolutional operations.

Tree-based models

Tree-based models are based on decision trees (DTs), among which RF and XGB (which are extensions of DTs) are prime examples. A DT model represents decision rules based on data features in a tree structure and is a supervised learning model used primarily for classification [62, 68]. DT models feature a hierarchical tree structure with multiple branches and nodes, that represent decision results and class labels, respectively (Fig. 4F). The decision process of a DT model can be interpreted easily; however, a single-tree model may not offer satisfactory predictions with complex datasets. RF models perform decision-making using multiple randomly generated DTs [62, 69]. RF models are not prone to overfitting and offer excellent predictions by combining multiple decision trees; however, their decision-making processes cannot be interpreted easily. XGB is a DT-based boosting method that learns by sequentially connecting DTs and compensating for their errors [62, 70]. XGB has some limitations, including challenges in parameter tuning, a high computational cost, and (similar to RFs) difficulty in interpreting the decision-making process.

Validation

We employed three primary validation methods. The two-way method trains the model on a training set and evaluates it on a test set. The three-way method uses a training set for training, a validation set for validation, and a test set for final evaluation. The k -fold cross-validation method splits the dataset into k subsets, training the model on $k-1$ subsets and testing it on the remaining subset. This process is repeated k times, each with a different test subset, to compute the average performance. The International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) endorses the k -fold method for ML research [71]. However, 94 of the 144 studies reviewed (65.3%) did not adopt k -fold cross-validation. These methods are detailed in Supplemental Data Table S1.

Optimization

Hyperparameters, determining aspects such as training speed,

batch size, and number of hidden layers, are pivotal for the performance of an ML model. Proper hyperparameter settings enhance predictive accuracy, stability, prevent overfitting, optimize resource use, and ensure reliable performance with new data. Formal optimization techniques (such as grid searching, random searching, and Bayesian optimization) were employed in only 12 cases (8.3%). In 16 studies (11.1%), optimization was performed arbitrarily, whereas in the remaining studies (80.6%, 116 cases), hyperparameter optimization was either not performed or was not comprehensively described. Details are summarized in Supplemental Data Table S1.

Evaluation

We assessed the efficacy of external and internal validations and noted issues in studies involving external validation. In seven studies [15, 17, 29, 52, 72-74], performance markedly declined with external validation data than with training and internal validation data. For instance, in a study designed to diagnose leukemia using a CNN and blood slide images [29], the accuracy was increased to 98.61% (compared with 92.79% in another study); however, the accuracy dropped to 70.24% during external validation, indicating potential overfitting. However, in some studies [22, 75-78], the performance variation was minimal or even showed higher AUROC values with external validation data [77, 78]. These outcomes were attributed to similar preprocessing of the training and external data or the use of appropriate regularization techniques to prevent overfitting. Two out of 15 studies [23, 79] involved the use of external validation data but did not provide precise results, complicating performance comparisons, highlighting the difficulty in evaluating performance when results from external validation are lacking.

CHALLENGES AND OPPORTUNITIES FOR ML APPLICATIONS IN LABORATORY MEDICINE

We analyzed the application of ML in laboratory medicine, the general trends observed when ML models were used, and the primary types of ML models used in research.

Current landscape

As discussed earlier, ML is primarily applied in diagnostic hematology, primarily because many laboratory medicine tests focus on blood analysis. Microscopy images contain several intricate details that cannot be identified easily with the naked eye, and results can vary between evaluators. Introducing ML should shift existing qualitative evaluations to quantitative evaluations and

reduce error margins, which may explain the growing number of studies involving ML applications [35, 36, 39, 80-83]. Conversely, fields with fewer ML applications, such as diagnostic immunology and transfusion medicine, have focused on well-established processes, such as blood-type determinations for transfusions and immune testing [53]. This trend is likely owing to the limited data inputs and the paramount need for accuracy in tests such as ABO typing [43, 44], which reduces the demand for complex ML methods.

The application of ML in laboratory tests has risen annually since 2014 (Fig. 3). Initially, ML utilization was limited to models such as MLPs and CNNs; however, as these models have advanced, the diversity of ML models used and their applications have expanded. Additionally, the estimation/prediction and disease screening collectively represent approximately 25% of all ML applications, signifying their role in augmenting the data provided by clinical laboratories. This evolution indicates that ML is increasingly instrumental in predicting disease occurrences or enhancing disease screening processes.

CNN, MLP, and tree-based ML models have been widely used in clinical laboratories. With advances in ML, various models categorized as “others” have been evaluated for their applicability; however, well-established models, such as CNNs, remain dominant. DNNs were developed much later than MLPs and therefore are not yet widely adopted. Stevenson, *et al.* [24] used large language models (LLMs) such as ChatGPT (v3.5) and Google Bard to interpret clinical test results and provide advice based on hypothetical inputs (similar to the roles of clinicians). Within the broader domain of laboratory medicine, extending beyond specialization, the latest evidence indicates that responses generated by chatbots such as ChatGPT consistently surpass the quality of those provided by clinicians [84-88]. As institutions such as the European Federation of Clinical Chemistry and Laboratory Medicine [89] are researching LLMs for laboratory medicine, this situation is likely to change in the future.

Securing clinically validated performance

Accuracy and AUROC are the most common evaluation metrics; however, appropriate evaluation metrics must be considered when processing class-imbalanced data. Data imbalance occurs when the proportions of a test class (e.g., patients with a disease) and control class (e.g., patients without a disease) differ substantially. Any case involving unequal numbers of samples per class indicates data imbalance and significant imbalances are particularly problematic during training. Although a universal definition is lacking, researchers typically categorize a case as

severely imbalanced when the minority class constitutes $\leq 10\%$ of the total dataset [90, 91]. This imbalance can skew learning toward the majority class, which can result in unsatisfactory prediction performance for the minority class, which typically represents the disease group. To address this issue, metrics such as the F1 score should be used [76]; however, some studies did not consider this aspect [74, 81]. In future studies, data imbalance must be addressed when evaluating ML model performance, and the appropriate evaluation metrics must be employed.

Some guidelines exist for the effective use of ML in clinical practice. The IFCC [71] proposed 15 recommendations for applying ML in clinical research, covering (1) stakeholders, (2) objectives, (3) clinical scenarios, (4) data descriptions, (5) statistical analysis of training and validation data, (6) steps to ensure proper data preparation, (7) dataset diversity, (8) ethical design, (9) validation methods, (10) the use of test sets, (11) performance metrics, (12) external validation, (13) interpretability, (14) code availability, and (15) generalizability.

Most studies examined in this review adhered to IFCC recommendations 2, 4, 7, and 10, although the level of detail provided varied. For instance, some studies only briefly discussed the data collection process, whereas others provided detailed explanations of all data collection and processing steps. This variation may reflect differences in the type, complexity, and diversity of the data used but may also have resulted from a lack of detailed guidelines for describing data collection and processing procedures.

Recommendations related to verifying model reliability and performance validation, such as IFCC recommendations 9, 11, 12, and 15, were observed in few studies. Despite being crucial for evaluating not only model performance but also robustness, generalizability, and clinical utility, these recommendations are not yet widely adopted, suggesting that researchers may not be aware of the importance of proper validation methods and procedures in ensuring the reliability of ML-based research findings. IFCC recommendations that are less directly related to performance (e.g., 1, 3, 5, 6, 8, 13, and 14) were not commonly addressed in most studies, suggesting that ML is still in the early stages of adoption in the field of laboratory medicine and that confirming its potential for high performance should be the primary focus. Notably, recommendation 14 to make data and code publicly available is frequently followed in general ML research but is generally restricted in the medical field because of patient privacy and data security concerns.

In conclusion, several ML-based studies may have been per-

formed without sufficient mechanisms to ensure the reliability of the results. Establishing standardized methods and guidelines is crucial for facilitating robust ML research and the generation of comparable results, and enhancing the reproducibility and credibility of such studies. A thorough consideration of research ethics and the broader ecosystem, which are currently underrepresented in ML research, will be essential as ML is integrated into clinical practice. Practical strategies for incorporating these aspects as fundamental research components are urgently required.

Limitations

Although we intended a comprehensive review, analyzing certain laboratory fields presents limitations. First, we could not include all specific keywords used in laboratory medicine or ML when constructing our search string; therefore, some publications may have been missed. Additionally, during the search and screening, we excluded literature we deemed less relevant to laboratory practice. Consequently, studies related to anatomical pathology and disease prediction (including sepsis) [92, 93] were omitted. We excluded COVID-19-related studies to avoid bias, as they represent a temporary epidemic rather than a universal application of ML.

Second, although gene or exome analyses are performed in clinical laboratories for patient care, most genome-related papers examined in this review were intended for research rather than clinical practice. Although these topics were not included in this review, they are important issues within laboratory medicine and warrant a separate review.

Third, some aspects must be considered when interpreting the results presented herein. For example, the Sankey diagram shows only the “best ML model,” which may not accurately represent diversity as it does not enumerate all models currently in use. In simplifying the model categories, DNNs are represented as typical networks with fully connected layers; therefore, accurately capturing the tendencies of DNNs over a broader range, including deep CNNs, is challenging.

Fourth, assessing AI model reproducibility and variability is crucial in evaluating model performance. To accurately assess these factors, certain conditions must be controlled (e.g., identical analytical goals and datasets) and consistent evaluation metrics must be used. However, in the reviewed studies, the analytical goals, datasets used, and performance evaluation metrics varied substantially, rendering it impossible to directly compare and analyze the performance of different models on the same basis. To analyze reproducibility and variability to some

degree, we analyzed the performance changes based on external validation research. Although this approach differs from repeatedly analyzing the same data to assess the stability and reproducibility of model outputs, comparing external and internal validation results can provide valuable insights into model generalizability, which is key for evaluating overall model performance.

The categorization of ML use cases can be ambiguous and open to interpretation. We categorized such cases based on the final output of the ML model. Additionally, the criteria for evaluating the appropriate number of ML input samples may vary depending on the laboratory medicine field, specimen type, and ML model used. Hence, an analysis using a more precise search string and clearer categorization criteria and considering the evaluation criteria for different sample sizes would be useful.

Finally, although not considered herein, ensuring data quality is a prerequisite for ML research [94-96]. To present a more robust and practical blueprint for ML utilization, the entire process (from data acquisition and preprocessing to analysis with various models) should be considered.

CONCLUSIONS

ML utilization in laboratory medicine is poised for continued growth and diversification. To date, CNN, MLP, and tree-based models have dominated the landscape, with the data type being the primary factor that influences model selection. However, as ML technology evolves, the introduction of new models is likely. We have identified several technical challenges associated with ML applications, primarily concerning data imbalances, missing hyperparameter optimization, inadequate evaluation metrics, and insufficient external validation. These findings emphasize the necessity for more sophisticated ML study designs and expert involvement. Considering the rapid advancement of ML and its established relevance in laboratory medicine, we anticipate that enhancing long-term education and fostering collaboration among domain specialists will optimize the use of ML in this field.

SUPPLEMENTARY MATERIALS

Supplementary materials can be found via <https://doi.org/10.3343/alm.2024.0354>

ACKNOWLEDGEMENTS

None.

AUTHOR CONTRIBUTIONS

Shin H designed and supervised the study; You J collected and verified the data; all authors screened the data; You J, Shin H, and Seok HS summarized the data; You J visualized the results; You J, Seok HS and Shin H wrote the original draft of the manuscript; You J, Kim S and Shin H revised and finalized the manuscript; and Kim S and Shin H secured the funding. All authors read and approved the final manuscript.

CONFLICTS OF INTEREST

None declared.

RESEARCH FUNDING

This study was supported by the National Research Foundation of Korea (NRF), a grant funded by the Korean government (MSIT) (grant No. RS-2024-00335644), and a grant from the Asan Institute for Life Sciences, Asan Medical Center, Seoul, Korea (grant No. 2024IP0001).

REFERENCES

1. Benjamins S, Dhunoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *NPJ Digit Med* 2020;3:118.
2. Cui M and Zhang DY. Artificial intelligence and computational pathology. *Lab Invest* 2021;101:412-22.
3. Rabbani N, Kim GYE, Suarez CJ, Chen JH. Applications of machine learning in routine laboratory medicine: current state and future directions. *Clin Biochem* 2022;103:1-7.
4. Khan AI, Khan M, Khan R. Artificial intelligence in point-of-care testing. *Ann Lab Med* 2023;43:401-7.
5. Duan X, Zhang M, Liu Y, Zheng W, Lim CY, Kim S, et al. Next-generation patient-based real-time quality control models. *Ann Lab Med* 2024;44:385-91.
6. Pillay TS. Artificial intelligence in pathology and laboratory medicine. *J Clin Pathol* 2021;74:407-8.
7. Niazi MKK, Parwani AV, Gurcan MN. Digital pathology and artificial intelligence. *Lancet Oncol* 2019;20:e253-61.
8. Korean Society for Laboratory Medicine. Laboratory medicine. 6th ed. Seoul: Panmun Education, 2021.
9. Logical Observation Identifiers Names and Codes Committee. LOINC specimen type <https://loinc.org/66746-9>.
10. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2016:779-88.
11. Bogart S. SankeyMATIC. <https://sankeymatic.com/>.
12. Hunter JD. Matplotlib: a 2D graphics environment. *Comput Sci Eng* 2007;9:90-5.
13. Xie J and Wang Q. Benchmarking machine learning algorithms on blood glucose prediction for type I diabetes in comparison with classical time-series models. *IEEE Trans Biomed Eng* 2020;67:3101-24.
14. Zhu T, Li K, Herrero P, Georgiou P. Personalized blood glucose prediction for type 1 diabetes using evidential deep learning and meta-learning. *IEEE Trans Biomed Eng* 2023;70:193-204.
15. Snowden SG, Korosi A, de Rooij SR, Koulman A. Combining lipidomics and machine learning to measure clinical lipids in dried blood spots. *Metabolomics* 2020;16:83.
16. Müller M, Sägeser N, Keller PM, Arampatzis S, Steffens B, Ehrhard S, et al. Urine flow cytometry parameter cannot safely predict contamination of urine—a cohort study of a Swiss emergency department using machine learning techniques. *Diagnostics (Basel)* 2022;12:1008.
17. Zhou R, Liang YF, Cheng HL, Wang W, Huang DW, Wang Z, et al. A highly accurate delta check method using deep learning for detection of sample mix-up in the clinical laboratory. *Clin Chem Lab Med* 2022;60:1984-92.
18. Seok HS, Choi Y, Yu S, Shin KH, Kim S, Shin H. Machine learning-based delta check method for detecting misidentification errors in tumor marker tests. *Clin Chem Lab Med* 2024;62:1421-32.
19. Wang H, Wang H, Zhang J, Li X, Sun C, Zhang Y. Using machine learning to develop an autoverification system in a clinical biochemistry laboratory. *Clin Chem Lab Med* 2021;59:883-91.
20. Demirci F, Akan P, Kume T, Sisman AR, Erbayraktar Z, Sevinc S. Artificial neural network approach in laboratory test reporting: learning algorithms. *Am J Clin Pathol* 2016;146:227-37.
21. Ialongo C, Pieri M, Bernardini S. Artificial neural network for total laboratory automation to improve the management of sample dilution: smart automation for clinical laboratory timeliness. *SLAS Technol* 2017;22:44-9.
22. Dauwalder O, Michel A, Eymard C, Santos K, Chanel L, Luzzati A, et al. Use of artificial intelligence for tailored routine urine analyses. *Clin Microbiol Infect* 2021;27:1168.e1-6.
23. Wilkes EH, Emmett E, Beltran L, Woodward GM, Carling RS. A machine learning approach for the automated interpretation of plasma amino acid profiles. *Clin Chem* 2020;66:1210-8.
24. Stevenson E, Walsh C, Hibberd L. Can artificial intelligence replace biochemists? A study comparing interpretation of thyroid function test results by ChatGPT and Google Bard to practising biochemists. *Ann Clin Biochem* 2024;61:143-9.
25. Del Ben F, Da Col G, Cobârzan D, Turetta M, Rubin D, Buttazzi P, et al. A fully interpretable machine learning model for increasing the effectiveness of urine screening. *Am J Clin Pathol* 2023;160:620-32.
26. Lee KS, Lim HJ, Kim K, Park YG, Yoo JW, Yong D. Rapid bacterial detection in urine using laser scattering and deep learning analysis. *Microbiol Spectr* 2022;10:e0176921.
27. Wang HY, Hung CC, Chen CH, Lee TY, Huang KY, Ning HC, et al. Increase *Trichomonas vaginalis* detection based on urine routine analysis through a machine learning approach. *Sci Rep* 2019;9:11074.
28. Avci D, Leblebicioglu MK, Poyraz M, Dogantekin E. A new method based on adaptive discrete wavelet entropy energy and neural network classifier (ADWEENN) for recognition of urine cells from microscopic images independent of rotation and scaling. *J Med Syst* 2014;38:7.
29. Vogado L, Veras R, Aires K, Araújo F, Silva R, Ponti M, et al. Diagnosis of leukaemia in blood slides based on a fine-tuned and highly generalisable deep learning model. *Sensors (Basel)* 2021;21:2989.

30. Jha KK and Dutta HS. Mutual information based hybrid model and deep learning for acute lymphocytic leukemia detection in single cell blood smear images. *Comput Methods Programs Biomed* 2019;179:104987.
31. Sampathila N, Chadaga K, Goswami N, Chadaga RP, Pandya M, Prabhu S, et al. Customized deep learning classifier for detection of acute lymphoblastic leukemia using blood smear images. *Healthcare (Basel)* 2022;10:1812.
32. Wang D, Hwang M, Jiang WC, Ding K, Chang HC, Hwang KS. A deep learning method for counting white blood cells in bone marrow images. *BMC Bioinformatics* 2021;22(S5):94.
33. Cheuque C, Querales M, León R, Salas R, Torres R. An efficient multi-level convolutional neural network approach for white blood cells classification. *Diagnostics (Basel)* 2022;12:248.
34. Xing Y, Liu X, Dai J, Ge X, Wang Q, Hu Z, et al. Artificial intelligence of digital morphology analyzers improves the efficiency of manual leukocyte differentiation of peripheral blood. *BMC Med Inform Decis Mak* 2023; 23:50.
35. Tamang T, Baral S, Paing MP. Classification of white blood cells: a comprehensive study using transfer learning based on convolutional neural networks. *Diagnostics (Basel)* 2022;12:2903.
36. Kihm A, Kaestner L, Wagner C, Quint S. Classification of red blood cell shapes in flow using outlier tolerant machine learning. *PLoS Comput Biol* 2018;14:e1006278.
37. Douglass PM, O'Connor T, Javidi B. Automated sickle cell disease identification in human red blood cells using a lensless single random phase encoding biosensor and convolutional neural networks. *Opt Express* 2022;30:35965-77.
38. de Haan K, Ceylan Koydemir H, Rivenon Y, Tseng D, Van Dyne E, Bakic L, et al. Automated screening of sickle cells using a smartphone-based microscope and deep learning. *NPJ Digit Med* 2020;3:76.
39. Ahn D, Lee J, Moon S, Park T. Human-level blood cell counting on lens-free shadow images exploiting deep neural networks. *Analyst* 2018; 143:5380-7.
40. Huang X, Jiang Y, Liu X, Xu H, Han Z, Rong H, et al. Machine learning based single-frame super-resolution processing for lensless blood cell counting. *Sensors (Basel)* 2016;16:1836.
41. Alam MM and Islam MT. Machine learning approach of automatic identification and counting of blood cells. *Healthc Technol Lett* 2019;6:103-8.
42. Zini G, Mancini F, Rossi E, Landucci S, d'Onofrio G. Artificial intelligence and the blood film: performance of the MC-80 digital morphology analyzer in samples with neoplastic and reactive cell types. *Int J Lab Hematol* 2023;45:881-9.
43. Epah J, Gülec I, Winter S, Dörr J, Geisen C, Haecker E, et al. From unit to dose: a machine learning approach for precise prediction of hemoglobin and iron content in individual packed red blood cell units. *Adv Sci (Weinh)* 2022;9:e2204077.
44. Larpant N, Niamsi W, Noiphung J, Chanakiat W, Sakuldamrongpanich T, Kittichai V, et al. Simultaneous phenotyping of five Rh red blood cell antigens on a paper-based analytical device combined with deep learning for rapid and accurate interpretation. *Anal Chim Acta* 2022;1207: 339807.
45. Roux-Dalvai F, Gotti C, Leclercq M, Hélie MC, Boissinot M, Arrey TN, et al. Fast and accurate bacterial species identification in urine specimens using LC-MS/MS mass spectrometry and machine learning. *Mol Cell Proteomics* 2019;18:2492-505.
46. Amano M, Mai DT, Sun G, Vu TN, Hoi LT, Hoa NT, et al. Deep learning approach for classifying bacteria types using morphology of bacterial colony. *Annu Int Conf IEEE Eng Med Biol Soc* 2022;2022:2165-8.
47. Rajaonison A, Le Page S, Maurin T, Chaudet H, Raoult D, Baron SA, et al. Antilogic, a new supervised machine learning software for the automatic interpretation of antibiotic susceptibility testing in clinical microbiology: proof-of-concept on three frequently isolated bacterial species. *Clin Microbiol Infect* 2022;28:1286.e1-8.
48. Huang L and Wu T. Novel neural network application for bacterial colony classification. *Theor Biol Med Model* 2018;15:22.
49. Brenton L, Waters MJ, Stanford T, Giglio S. Clinical evaluation of the APAS® Independence: automated imaging and interpretation of urine cultures using artificial intelligence with composite reference standard discrepant resolution. *J Microbiol Methods* 2020;177:106047.
50. Gammel N, Ross TL, Lewis S, Olson M, Henciak S, Harris R, et al. Comparison of an automated plate assessment system (APAS Independence) and artificial intelligence (AI) to manual plate reading of methicillin-resistant and methicillin-susceptible *Staphylococcus aureus* CHRO-Magar surveillance cultures. *J Clin Microbiol* 2021;59:e0097121.
51. He Y, Peng P, Ying W, Wang Q, Wang Y, Liu X, et al. Contrast between traditional and machine learning algorithms based on a urine culture predictive model: a multicenter retrospective study in patients with urinary calculi. *Transl Androl Urol* 2022;11:139-48.
52. Gao Z, Wang L, Zhou L, Zhang J. HEP-2 cell image classification with deep convolutional neural networks. *IEEE J Biomed Health Inform* 2017; 21:416-28.
53. Fang K, Li C, Wang J. An automatic immunofluorescence pattern classification framework for HEP-2 image based on supervised learning. *Brief Bioinform* 2023;24:bbad144.
54. Xiao L, Luo C, Yu T, Luo Y, Wang M, Yu F, et al. DeepACEv2: automated chromosome enumeration in metaphase cell images using deep convolutional neural networks. *IEEE Trans Med Imaging* 2020;39:3920-32.
55. Vajen B, Hänselmann S, Lutterloh F, Käfer S, Espenkötter J, Beening A, et al. Classification of fluorescent R-band metaphase chromosomes using a convolutional neural network is precise and fast in generating karyograms of hematologic neoplastic cells. *Cancer Genet* 2022;260-261:23-9.
56. Gangadhar A, Sari-Sarraf H, Vanapalli SA. Deep learning assisted holography microscopy for in-flow enumeration of tumor cells in blood. *RSC Adv* 2023;13:4222-35.
57. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning. New York: Springer, 2013.
58. Cox DR. The regression analysis of binary sequences. *J R Stat Soc B* 1958;20:215-32.
59. Chung MK. Introduction to logistic regression. arXiv preprint arXiv:2008.13567 2020. <https://doi.org/10.48550/arXiv.2008.13567>.
60. Cortes C and Vapnik V. Support-vector networks. *Mach Learn* 1995;20: 273-97.
61. Mahesh B. Machine learning algorithms—a review. *Int J Sci Res.* 2020;9: 381-6.
62. Sarker IH. Machine learning: algorithms, real-world applications and research directions. *SN Comput Sci* 2021;2:160.
63. Goodfellow I, Bengio Y, Courville A. Deep learning. Cambridge: MIT Press, 2016.
64. Shrestha A and Mahmood A. Review of deep learning algorithms and architectures. *IEEE Access* 2019;7:53040-65.
65. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst* 2012;25.
66. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM* 2017;60:84-90.
67. Yamashita R, Nishio M, Do RKG, Togashi K. Convolutional neural networks: an overview and application in radiology. *Insights Imaging* 2018; 9:611-29.
68. Breiman L, Friedman J, et al. eds. Classification and regression trees.

- Boca Raton: Routledge, 2017.
69. Breiman L. Random forests. *Mach Learn* 2001;45:5-32.
70. Chen T and Guestrin C. Xgboost: a scalable tree boosting system. *Proc 22nd ACM SIGKDD Intl Conf Knowl Discov Data Min.* 2016:785-94.
71. Master SR, Badrick TC, Bietenbeck A, Haymond S. Machine learning in laboratory medicine: recommendations of the IFCC working group. *Clin Chem* 2023;69:690-8.
72. Rodríguez-Temporal D, Herrera L, Alcaide F, Domingo D, Héry-Arnaud G, van Ingen J, et al. Identification of *Mycobacterium abscessus* subspecies by MALDI-TOF mass spectrometry and machine learning. *J Clin Microbiol* 2023;61:e0111022.
73. McFadden BR, Inglis TJJ, Reynolds M. Machine learning pipeline for blood culture outcome prediction using Sysmex XN-2000 blood sample results in Western Australia. *BMC Infect Dis* 2023;23:552.
74. Fang K, Dong Z, Chen X, Zhu J, Zhang B, You J, et al. Using machine learning to identify clotted specimens in coagulation testing. *Clin Chem Lab Med* 2021;59:1289-97.
75. Steinbach D, Ahrens PC, Schmidt M, Federbusch M, Heuft L, Lübbert C, et al. Applying machine learning to blood count data predicts sepsis with ICU admission. *Clin Chem* 2024;70:506-15.
76. Liao H, Xu Y, Meng Q, Mao Z, Qiao Y, Liu Y, et al. A convolutional neural network-based, quantitative complete blood count scattergram-mapping framework promptly screens acute promyelocytic leukemia with high sensitivity. *Cancer* 2023;129:2986-98.
77. Chang YH, Hsiao CT, Chang YC, Lai HY, Lin HH, Chen CC, et al. Machine learning of cell population data, complete blood count, and differential count parameters for early prediction of bacteremia among adult patients with suspected bacterial infections and blood culture sampling in emergency departments. *J Microbiol Immunol Infect* 2023;56:782-92.
78. Acevedo A, Merino A, Boldú L, Molina Á, Alférez S, Rodellar J. A new convolutional neural network predictive model for the automatic recognition of hypogranulated neutrophils in myelodysplastic syndromes. *Comput Biol Med* 2021;134:104479.
79. Ialongo C, Pieri M, Bernardini S. Smart management of sample dilution using an artificial neural network to achieve streamlined processes and saving resources: the automated nephelometric testing of serum free light chain as case study. *Clin Chem Lab Med* 2017;55:231-6.
80. Khan RU, Almakdi S, Alshehri M, Haq AU, Ullah A, Kumar R. An intelligent neural network model to detect red blood cells for various blood structure classification in microscopic medical images. *Heliyon* 2024;10:e26149.
81. Lin YH, Liao KYK, Sung KB. Automatic detection and characterization of quantitative phase images of thalassemic red blood cells using a mask region-based convolutional neural network. *J Biomed Opt* 2020;25:116502.
82. Sharma S, Gupta S, Gupta D, Juneja S, Gupta P, Dhiman G, et al. Deep learning model for the automatic classification of white blood cells. *Comput Intell Neurosci* 2022;2022:7384131.
83. Choi JW, Ku Y, Yoo BW, Kim JA, Lee DS, Chai YJ, et al. White blood cell differential count of maturation stages in bone marrow smear using dual-stage convolutional neural networks. *PLoS One* 2017;12:e0189259.
84. Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA Intern Med* 2023;183:589-96.
85. Kanjee Z, Crowe B, Rodman A. Accuracy of a generative artificial intelligence model in a complex diagnostic challenge. *JAMA* 2023;330:78-80.
86. Yang HS, Wang F, Greenblatt MB, Huang SX, Zhang Y. AI chatbots in clinical laboratory medicine: foundations and trends. *Clin Chem* 2023;69:1238-46.
87. Kurstjens S, Schipper A, Krabbe J, Kusters R. Predicting hemoglobinopathies using ChatGPT. *Clin Chem Lab Med* 2024;62:e59-61.
88. Wu AHB, Jaffe AS, Peacock WF, Kavsak P, Greene D, Christenson RH. The role of artificial intelligence for providing scientific content for laboratory medicine. *J Appl Lab Med* 2024;9:386-93.
89. Cadamuro J, Cabitza F, Debeljak Z, De Bruyne S, Frans G, Perez SM, et al. Potentials and pitfalls of ChatGPT and natural-language artificial intelligence models for the understanding of laboratory medicine test results. An assessment by the European Federation of Clinical chemistry and Laboratory Medicine (EFLM) Working Group on Artificial Intelligence (WG-AI). *Clin Chem Lab Med* 2023;61:1158-66.
90. He H and Garcia EA. Learning from imbalanced data. *IEEE Trans Knowl Data Eng* 2009;21:1263-84.
91. Drummond C and Holte RC. Severe class imbalance: why better algorithms aren't the answer. *European Conference on Machine Learning*. Berlin, Heidelberg: Springer. 2005;3720:539-46.
92. Carobene A, Milella F, Famiglini L, Cabitza F. How is test laboratory data used and characterised by machine learning models? A systematic review of diagnostic and prognostic models developed for COVID-19 patients using only laboratory data. *Clin Chem Lab Med* 2022;60:1887-901.
93. Agnello L, Vidali M, Padoan A, Lucis R, Mancini A, Guerranti R, et al. Machine learning algorithms in sepsis. *Clin Chim Acta* 2024;553:117738.
94. Kim S. Laboratory data quality evaluation in the big data era. *Ann Lab Med* 2023;43:399-400.
95. Cho EJ, Jeong TD, Kim S, Park HD, Yun YM, Chun S, et al. A new strategy for evaluating the quality of laboratory results for big data research: using external quality assessment survey data (2010–2020). *Ann Lab Med* 2023;43:425-33.
96. Kim S, Cho EJ, Jeong TD, Park HD, Yun YM, Lee K, et al. Proposed model for evaluating real-world laboratory results for big data research. *Ann Lab Med* 2023;43:104-7.