Blood Reviews 62 (2023) 101134

Contents lists available at ScienceDirect

Blood Reviews

journal homepage: www.elsevier.com/locate/issn/0268960X

Review

Revolutionizing chronic lymphocytic leukemia diagnosis: A deep dive into the diverse applications of machine learning

Mohamed Elhadary [a,*], Amgad Mohamed Elshoeibi [a], Ahmed Badr [a], Basel Elsayed [a], Omar Metwally [a], Ahmed Mohamed Elshoeibi [b], Mervat Mattar [c], Khalil Alfarsi [d], Salem AlShammari [e], Awni Alshurafa [f], Mohamed Yassin [f,*]

[a] College of Medicine, QU Health, Qatar University, Doha, Qatar
[b] School of Medicine, Newgiza University, Giza, Egypt
[c] Internal Medicine and Clinical Hematology, Cairo University, Cairo, Egypt
[d] Department of Hematology, College of Medicine and Health Sciences, Sultan Qaboos University, Muscat, Oman
[e] Department of Medicine, Faculty of Medicine, Kuwait University, Kuwait, Kuwait
[f] Hematology Section, Medical Oncology, National Center for Cancer Care and Research (NCCCR), Hamad Medical Corporation, Doha, Qatar

ARTICLE INFO

Keywords:
Artificial intelligence
Chronic lymphocytic leukemia
Diagnosis
Machine learning

ABSTRACT

Chronic lymphocytic leukemia (CLL) is a B cell neoplasm characterized by the accumulation of aberrant monoclonal B lymphocytes. CLL is the predominant type of leukemia in Western countries, accounting for 25% of cases. Although many patients remain asymptomatic, a subset may exhibit typical lymphoma symptoms, acquired immunodeficiency disorders, or autoimmune complications. Diagnosis involves blood tests showing increased lymphocytes and further examination using peripheral blood smear and flow cytometry to confirm the disease. With the significant advancements in machine learning (ML) and artificial intelligence (AI) in recent years, numerous models and algorithms have been proposed to support the diagnosis and classification of CLL. In this review, we discuss the benefits and drawbacks of recent applications of ML algorithms in the diagnosis and evaluation of patients diagnosed with CLL.

1. Introduction

Chronic lymphocytic leukemia (CLL) and small lymphocytic lymphoma (SLL) belong to the category of mature B cell neoplasms, characterized by the progressive accumulation of monoclonal B lymphocytes. Pathological and immunological features are identical between the two, where CLL is the designation for the blood involvement, while SLL refers to primary lymph node involvement [1,2]. CLL is the most common leukemia in the western countries, accounting for about 25% of new cases, with >100,000 incident cases reported globally in 2019 [3,4]. Patients with CLL/SLL are typically asymptomatic or have painless lymph nodes swelling. A subset of patients may exhibit characteristic lymphoma symptoms, including fever, fatigue, night sweats, and unexplained weight loss. Furthermore, some individuals may manifest signs suggestive of immune dysfunction, such as acquired immunodeficiency disorders characterized by recurrent infections or autoimmune complications like hemolytic anemia, red cell aplasia, thrombocytopenia, and hypogammaglobulinemia. CLL is suspected when routine blood tests show absolute lymphocytosis. Further investigation with peripheral blood smear can demonstrate lymphocytosis of small mature lymphocytes with dark nuclei. Additionally, flow cytometry is utilized to conduct immunophenotypic analysis using a panel of antibodies specific to CLL [1,5,6]. CLL is considered a heterogeneous disease with regards to prognosis, that is why clinicians use tools such as Rai and Binet staging systems that have proven to be strongly associated with clinical outcomes to stratify patients with CLL and treat them accordingly [7–9]. Treatment is not initiated for all patients who receive a CLL diagnosis, as those with asymptomatic early-stage disease are actively monitored without initiating treatment. Indications for commencement of CLL therapy include progressive bone marrow failure evidenced by anemia and/or thrombocytopenia, progressive or symptomatic splenomegaly, constitutional symptoms, autoimmune hemolytic anemia and/or immune thrombocytopenia irresponsive to corticosteroids, and progressive lymphocytosis with rapid lymphocyte

* Corresponding authors.
E-mail addresses: me1902913@qu.edu.qa (M. Elhadary), yassin@hamad.qa (M. Yassin).

https://doi.org/10.1016/j.blre.2023.101134

Available online 22 September 2023

doubling time. Drugs that inhibit enzymes in the B-cell receptor (BCR) signaling pathway, such as Bruton tyrosine kinase (BTK) and inhibit proteins that regulate apoptosis such as B-cell lymphoma 2 (BCL2), are the first-line agents for treating CLL [1,5,6].

For most medicinal purposes, the terms "artificial intelligence" (AI) and "machine learning" (ML) are used interchangeably, yet it is crucial to recognize their distinct roles within the field of computer science and technology. Artificial intelligence encompasses a diverse range of technologies and methodologies dedicated to creating systems that can emulate human-like intelligence and decision-making processes. On the other hand, machine learning is a specialized subset within the AI domain, concentrating on the development of models known as function approximators. These models are able to autonomously make informed decisions and draw conclusions by identifying patterns and extracting meaningful insights from raw data [10,11].

In hematology, machine learning has the potential to reshape clinical practices profoundly. It is actively utilized to construct intricate clinical models that aid in disease diagnosis, personalized treatment optimization, and comprehensive risk assessment across a wide spectrum of diseases and malignancies [12–15]. To gain a deeper understanding of machine learning in a medical context, it is essential to grasp the core concepts of supervised learning and unsupervised learning. Supervised learning thrives on labeled datasets, wherein each data point is tagged with a known outcome or classification. This supervised approach allows the algorithm to learn from historical data, making predictions or classifications on new, unseen data. In contrast, unsupervised learning comes into play when dealing with unlabeled data. It's analogous to explorative data analysis, where algorithms uncover hidden patterns, structures, or relationships within the data, often contributing to the discovery of novel insights. Central to these machine learning paradigms are function approximators, which serve as the mathematical engines enabling the modeling of intricate relationships between clinical parameters and disease outcomes. Together, these concepts provide researchers and clinicians with potent tools to harness the potential of AI and ML, driving advances in patient care and enhancing our comprehension of medical complexities [14,16].

Here, we provide an overview of the current state of ML applications in the classification and diagnosis of CLL. This review critically evaluates the performance and limitations encountered by the models documented in the existing literature. The aim of this review is to encourage further research to overcome these limitations and facilitate the integration of these models in clinical settings, ultimately enhancing the quality of patient care.

2. Material and methods

2.1. Literature search strategy

A literature search of all studies pertaining to ML implementations in CLL was conducted using the PubMed/MEDLINE and EMBASE databases on the 11th of April 2023. Terms pertaining to CLL (e.g., "chronic lymphatic leukemia", "chronic lymphocytic leukemia", "CLL") and machine learning (e.g., "AI", "machine learning", "neural network") were used in the search strategy and combined using Boolean operators 'AND' or 'OR'. After applying the search strategy, all of the identified studies were transferred to EndNote, where duplicates were eliminated. The resulting studies were then transferred to Rayyan to conduct further screening and remove any additional duplicates. In addition, the references of the identified studies, review articles, systematic reviews, and meta-analyses were manually screened to identify additional studies.

The collected data included several aspects including the type of study, publication year, assessed outcome, model creation methods, used model(s), and evaluation metrics for the model(s) such as sensitivity (SEN), specificity (SPE), accuracy (ACC), and area under the receiver operating curve (AUROC). An online confusion matrix calculator was used to obtain the evaluation metrics when these metrics were not explicitly reported in the reviewed articles' manuscripts. In cases where multiple models were used in a study, the metrics for the best-performing model were extracted. The collected data also encompassed the strengths and limitations of the studies.

2.2. Inclusion and exclusion criteria

The primary literature that discussed the use of ML algorithms in different CLL applications was considered for inclusion in this review. No date or language restrictions were considered. Research articles were included in the review if they met the following criteria: 1) The authors used a method that relies on the usage of ML to function, 2) The research reported conclusions regarding the reliability or accuracy of using such method, 3) The outcome of the research pertains to diagnosis and classification of CLL. Articles that were excluded from this review were non-English articles, animal studies, in vitro studies, abstracts, and review articles.

A total of 169 articles were identified through a search of PubMed and EMBASE databases. Duplicate articles were removed using Endnote® and Rayyan® software, resulting in 149 articles, which were further screened using Rayyan®. After screening, 14 studies met the inclusion criteria. Details of the screening process are provided in Fig. 1.

3. Role of ML in diagnosis and classification of CLL

To diagnose CLL, two criteria must be met: 1) an absolute B lymphocyte count of $\geq$5000/μL in the peripheral blood with predominantly mature small lymphocytes on smear examination, sustained for 3 months, and 2) evidence of immunoglobulin light chain restriction (kappa or lambda), low levels of surface membrane immunoglobulin (SmIg), and expression of B cell antigens (CD19, CD20, CD23) and CD5 on flow cytometric analysis of peripheral blood [5]. ML algorithms can aid this process by automating the interpretation of these tests and predicting the diagnosis, which can improve the efficiency of hematologists and decrease the processing time for these tests. A summary of the advantages, disadvantages, and outcome addressed in each study is provided in Table 1. Additionally, Table 2 outlines the performance for the best ML models developed in the studies reviewed here.

3.1. Using genetic data

The genomic characteristics of leukemias have important implications for diagnosis, risk stratification, and identification of therapeutic targets. Fluorescence in situ hybridization (FISH) and karyotyping are commonly employed techniques for detecting chromosomal abnormalities that distinguish CLL from other lymphoproliferative diseases. As such, an ML model was developed utilizing the targeted transcriptome of RNA samples from various neoplasms. The dataset included 167 samples of CLL, which were analyzed using next-generation sequencing with a targeted panel of 1408 cancer-associated genes. To address the risk of overfitting, the classification of the neoplasms was performed using the geometric mean Naïve Bayes algorithm. The model demonstrated excellent performance in classifying two diagnostic classes (e.g., neoplasm vs. normal or neoplasm vs. neoplasm), achieving an AUROC of 99.7% specifically in distinguishing CLL from normal samples [17]. By utilizing this approach, the study aimed to leverage genomic information to enhance the accuracy of leukemia classification. The integration of ML techniques with comprehensive genomic profiling holds promise for advancing our understanding of leukemia pathogenesis and prognostic factors and facilitating personalized treatment strategies. However, the study lacked data pertaining to molecular mutations or chromosomal abnormalities, which could have impacted the accuracy and reliability of the model's predictions. Furthermore, the model was not externally validated on a separate dataset. Consequently, additional evaluation of this model using a distinct sample is needed to consider the integration into clinical practice.

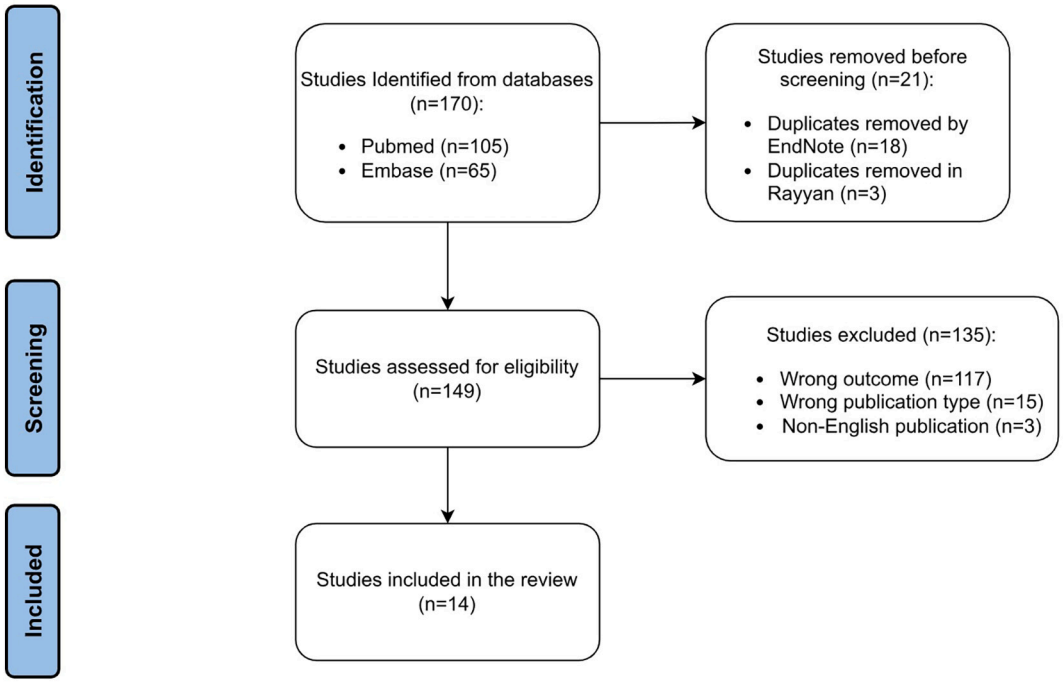


Fig. 1. Schematic representation of the screening process.

A study by Zhu et al. utilized the Gene Expression Omnibus (GEO) database to identify potential diagnostic biomarkers for CLL. The study employed differential gene expression analysis and weighted gene co-expression network analysis (WGCNA), resulting in the identification of 47 differentially expressed genes (DEGs) and 25 hub genes, respectively. To further refine the selection, LASSO (Least Absolute Shrinkage and Selection Operator) regression analysis was performed on the 14 genes that overlapped between the DEGs and hub genes, resulting in the identification of six final hub genes. Fig. 2 highlights the methods employed by the study for the identification and assessment of the candidate genes. The diagnostic performance of these genes was assessed using receiver operating characteristic (ROC) curves, which demonstrated AUROC values of up to 97.3% and 99.7% in the training and validation cohorts, respectively [18]. Although the identified genes exhibited excellent diagnostic performance, it is important to note that the study utilized multiple datasets obtained from the GEO database, which makes it prone to batch effects. Consequently, further prospective assessment of the diagnostic power of these genes is needed.

Similarly, Xia et al. utilized SVM algorithm to classify small B-cell lymphomas (SBCLs) based on DNA methylation patterns. A feature set of 26 probes was selected for the SVM model. The model was trained and evaluated using leave-one-out cross-validation (LOOCV) on 75 cases encompassing various SBCL subtypes (CLL/SLL, FL, MZL, MCL). In the 74 classifiable cases, the model achieved a 100% concordance rate with pathology diagnoses. Furthermore, an independent dataset evaluation yielded a 99% concordance rate, with CLL/SLL cases classified with 100% accuracy [19]. The study demonstrates the accurate discrimination of the four classes of SBCLs using methylation patterns. However, further studies exploring the clinical utility of this method are currently lacking.

3.2. *Using blood smear images*

Abhishek et al. employed deep transfer learning to develop multiple ML models for the classification of the four major subtypes of leukemia (ALL, AML, CLL, CML) using peripheral blood smear images. A dataset comprising 1250 microscopic blood smear images was created, including 250 images with a diagnosis of CLL. Two splits, namely 80/20 and 50/50, were utilized for training and validating the models. Convolutional neural network (CNN) models pre-trained on the ImageNet dataset [20] were employed for feature extraction. Subsequently, the extracted features were utilized to train support vector machine (SVM) and random forest (RF) models to classify the images into one of the leukemia subtypes or normal. Fig. 3 is a schematic representation of the proposed framework, the fine tuning of the VGG16 model, and the integration of extracted features to train the SVM and RF models. The performance of the classifiers was evaluated on a subject-dependent test dataset, and an additional subject-independent test dataset was used to assess the models' performance on unseen data. The RF model outperformed others in detecting CLL, achieving 100% sensitivity and 96% specificity when using the features extracted by the VGG16 model [21]. The model's ability to accurately detect CLL based on blood smear images offers substantial potential for enhancing the efficiency of hematologists in evaluating patients with lymphocytosis. However, it is important to consider that the feature extractors utilized in the study were trained on a distinct dataset from the one employed for the classifiers. This discrepancy introduces the possibility of negative transfer, potentially leading to a decline in classifier performance. Additionally, it is worth noting that the training dataset for the models consisted of CLL images obtained from only seven patients. To optimize performance, particularly with unseen data, further training and validation of the models with larger sample sizes are warranted.

In comparison, Dese et al. conducted a study utilizing SVM to train a model on blood smear images for the detection and classification of the four major subtypes of leukemia. The study employed various image preprocessing techniques to enhance image quality, and ML algorithms were utilized for feature extraction. The dataset was partitioned into training, testing, and validation sets with a distribution of 60:25:15, respectively. During the validation phase, the model achieved remarkable results, with 100% accuracy, sensitivity, and specificity in classifying CLL cases. Furthermore, the time required by the model to provide the diagnostic outcome was <1 min [22]. The developed system demonstrates promising potential in replacing manual methods of blood smear examination, thereby potentially expediting the leukemia diagnostic process. Nevertheless, it is crucial to validate this model on independent datasets before considering its implementation in a clinical

Table 1
Summary of the studies discussing the diagnosis and classification of CLL.

Reference	Outcome	Advantages	Disadvantages
Zhang, Qureshi et al. (2023)	Diagnosis and classification of tumors using targeted RNA expression profiling	- Lower chance of overfitting - Can give information about cancer biology, prognosis, and therapeutic targets	- NGS is not routinely ordered for CLL workup - Data did not include mutations and chromosomal abnormalities - The model was not externally validated
Zhu, Gan et al. (2022)	Identification of diagnostic biomarkers for CLL using GEO database	- Combination of bioinformatic analyses and ML - Validation of the identified genes	- High risk of batch effects due to usage of multiple datasets from the GEO database
Xia, Leon et al. (2021)	Diagnosis and classification of SBCLs using DNA methylation profiling	- Able to classify CLL/SLL, MCL, MZL, FL with high accuracy - The model is available online for research use - The model was internally and externally validated	- Unsupervised analysis was not possible due to significant batch effects
Abhishek, Jha et al. (2023)	Diagnosis and classification of leukemia using images of blood smears	- Able to classify ALL, AML, CLL, CML, and normal samples - Uses only images of peripheral blood smear - CNN models were used as feature extractors to optimize the performance of the classifier - The models were internally and externally validated	- Time taken for classification was not mentioned - The entire sample of CLL images was obtained from 7 patients only - Dataset used to train feature extractors was different from the one used to train the classifier (negative transfer)
Dese, Raj et al. (2021)	Diagnosis and classification of leukemia using images of blood smears	- Able to classify ALL, AML, CLL, CML, and normal samples - Uses only images of peripheral blood smear - Rapid time to diagnosis (<1 min) - ML algorithms used in feature extraction to optimize performance of the classifier	- The model was not externally validated
Mohammed, Mohamed et al. (2017)	Diagnosis of CLL using images of blood smears	- SVM was used for lymphocyte segmentation - Multiple ML models were evaluated for the classification task - majority voting fusion method was used to improve classification performance - Can be used as a quick and cheap screening tool for CLL	- The image acquisition method used in the study differs from other clinical settings - A search technique for cells is needed to obtain images similar to the ones in the study
Simonson, Lee et al. (2022)	Predict whether additional antibody panel should be ordered to distinguish CLL from MCL	- Use of ensemble learning - Models utilize flow cytometry data - Models were evaluated prospectively on new cases - Models were internally and externally validated	- Dataset lacked information on previous diagnoses of leukemia/lymphoma - The model was developed using data from a single laboratory
Ng and Zuromski (2021)	Diagnosis and classification of B-cell malignancies using flow cytometry	- Able to classify BNHL, B-ALL/LBL, CLL, DLBCL, and others accurately - Able to detect BNHL and B-ALL/LBL cases that require confirmatory studies - Relatively large sample - Rapid time to diagnosis (~35 s) - Use of UMAP for dimensionality reduction	- Difficulty of troubleshooting the misclassifications
Zhao, Mallesh et al. (2020)	Diagnosis and classification mature B-cell neoplasms using flow cytometry	- Able to classify seven subtypes of mature B-cell neoplasms (CLL, MCL, PL, LPL, MZL, FL, HCL), MBL, and normal samples - Large sample - Use of SOM for dimensionality reduction	- Specific performance metrics for CLL classification were not reported - Time taken for classification was not mentioned
Haider, Ujjan et al. (2022)	Early diagnosis and classification of leukemia using CBC	- Able to classify ALL, AML, APML, CLL, CML, and others - Able to accurately detect and subtype leukemia using only CBC items and CPD	- The model was not externally validated - Time taken for classification was not mentioned
Steinbuss, Kriegsmann et al. (2021)	Diagnosis and classification of NHL using LNs histopathological images	- Relatively large sample - Use of quality control limits to improve accuracy	- The model can only classify two disease entities - Low sensitivity in detecting CLL/SLL
do Nascimento, Martins et al. (2018)	Diagnosis and classification of NHL using LNs histopathological images	- Able to classify CLL, MCL, and FL - Multiple classifiers were evaluated in the study - Multiple statistical methods were applied to optimize feature selection	- Small sample - The model was not externally validated - The proposed algorithm requires long processing time
Zhang, Cui et al. (2020)	Classification of NHL subtypes using histopathological images	- Able to classify CLL, FL, and MCL - TL and PCA were used for fine-tuning and feature extraction, with a neural network model used for classification	- The model was not externally validated - Time taken for classification was not mentioned
Féré, Gobinet et al. (2020)	Diagnosis of CLL using Raman data	- The use of rdCV to reduce overfitting - The use of adaptive decision thresholds to adapt the model to different clinical scenarios - The use of consensus label strategy to improve model stability - The use of four tests to ensure the quality of Raman data - Rapid time to diagnosis (13 s) - Can be used to study biochemical changes in CLL	- Raman spectroscopy is not routinely ordered for CLL workup

NGS, next-generation sequencing; ALL, acute lymphoblastic leukemia; AML, acute myeloblastic leukemia; APML, acute promyelocytic leukemia; CML, chronic myelocytic leukemia; CNN, convolutional neural network; MCL, mantle cell lymphoma; SLL, small lymphocytic lymphoma; FL, follicular lymphoma; CBC, complete blood count; CPD, cell population data; GEO, gene expression omnibus; NHL, non-hodgkin lymphoma; LN, lymph node; BNHL, B-cell non-Hodgkin lymphoma; B-ALL/LBL, B-lymphoblastic leukemia/lymphoma; UMAP, uniform manifold approximation and projection; MCL, mantle cell lymphoma; MZL, marginal zone lymphoma; PL, prolymphocytic leukemia; HCL, hairy cell leukemia; LPL, lymphoplasmacytic lymphoma; MBL, monoclonal B-cell lymphocytosis; SOM, self-organizing map; TL, transfer learning; PCA, principal component analysis; rdCV, repeated double cross-validation; SVM, support vector machine.

Table 2
Performance of the best models for diagnosis and classification of CLL.

Reference	Outcome	Best model(s)	AUROC	ACC	SEN	SPE
Zhang, Qureshi et al. (2023)	Diagnosis and classification of tumors using targeted RNA expression profiling	GMNB	99.7% (CLL vs. normal) 98.6% (CLL vs. MCL) 98.4% (CLL vs. MZL)	NR	96.4% (CLL vs. normal) 94.6% (CLL vs. MCL) 98.7% (CLL vs. MZL)	98.8% (CLL vs. normal) 95.2% (CLL vs. MCL) 91% (CLL vs. MZL)
Zhu, Gan et al. (2022)	Identification of diagnostic biomarkers for CLL using GEO database	LASSO	NA	NA	NA	NA
Xia, Leon et al. (2021)	Diagnosis and classification of SBCLs using DNA methylation profiling	SVM	NR	100%	100%	99%
Abhishek, Jha et al. (2023)	Diagnosis and classification of leukemia using images of blood smears	VGG16 (feature extraction) – RF (classification)	NR	96.8%	100%	96%
Dese, Raj et al. (2021)	Diagnosis and classification of leukemia using images of blood smears	SVM	NR	100%	100%	100%
Mohammed, Mohamed et al. (2017)	Diagnosis of CLL using images of blood smears	MCS (SVM, KNN, and DT)	NR	85%	89%	87%
Simonson, Lee et al. (2022)	Predict whether additional antibody panel should be ordered to distinguish CLL from MCL	EnsembleCNN (CNN supplies predictions & RF integrates)	89%	94%	78%	95%
Ng and Zuromski (2021)	Diagnosis and classification of B-cell malignancies using flow cytometry	RF	96.9%	96.4%	86.9%	98.3%
Zhao, Mallesh et al. (2020)	Diagnosis and classification mature B-cell neoplasms using flow cytometry	CNN	NR	83%	NR	NR
Haider, Ujjan et al. (2022)	Early diagnosis and classification of leukemia using CBC	RBFN	90.5%	NR	NR	NR
Steinbuss, Kriegsmann et al. (2021)	Diagnosis and classification of NHL using LNs histopathological images	EfficientNetB3	NR	91.2%	62%	100%
do Nascimento, Martins et al. (2018)	Diagnosis and classification of NHL using LNs histopathological images	PL	100% (CLL vs. FL) 100% (CLL vs. MCL)	100% (CLL vs. FL) 100% (CLL vs. MCL)	100% (CLL vs. FL) 100% (CLL vs. MCL)	100% (CLL vs. FL) 100% (CLL vs. MCL)
Zhang, Cui et al. (2020)	Classification of NHL subtypes using histopathological images	VGG16 (feature extraction) – Neural network (classification)	NR	99.3%	98.4%	99.8%
Féré, Gobinet et al. (2020)	Diagnose CLL using Raman data	PLS-DA	NR	NR	95%	85%

GMNB, geometric mean naïve Bayes; MCL, mantle cell lymphoma; MZL, marginal zone lymphoma; VGG16, visual geometry group 16; SVM, support vector machine; RF, random forest; RBFN, radial basis function network; GEO, gene expression omnibus; LASSO, least absolute shrinkage and selection operator; CNN, convolutional neural network; PLS-DA, partial least squares–discriminant analysis; PL, polynomial; MCS, multiple classifier system; KNN, K-nearest neighbor; DT, decision tree; NR, not reported; NA, not applicable.

setting.

Similarly, the study conducted by Mohammed et al. [23] presented a system capable of classifying WBCs as either CLL or normal based on blood smear images. A dataset comprising 1010 images (791 CLL and 219 normal) was utilized for training and validating the classifiers. Additionally, 5535 independent CLL images were employed to assess the agreement between the developed system and flow cytometry results. The authors evaluated two ML techniques, namely SVM and artificial neural network (ANN), for lymphocyte nucleus segmentation. The SVM method outperformed the ANN method in this task. Subsequently, five classifiers and different fusion combinations of these classifiers were evaluated. The fusion model consisting of SVM, k-nearest neighbors (KNN), and decision tree (DT) demonstrated the best performance, achieving an accuracy of 87%, sensitivity of 85%, and specificity of 89%. When evaluating the fusion model against the flow cytometer on 11 cases of CLL, concordance was observed between the two systems in 9 out of the 11 cases. This system offers the potential to assist hematopathologists by identifying the proportions of CLL and normal cells in a sample and providing diagnostic suggestions. Implementation of this system could enhance the efficiency of CLL screening by offering a quicker and more cost-effective alternative to advanced tests like flow cytometry.

3.3. Using flow cytometry

Flow cytometry plays a crucial role in diagnosing CLL, where a screening antibody panel comprising CD45, CD19, CD20, CD5, CD10, CD8, and immunoglobulin light chains is typically ordered for suspected B-cell malignancy. If necessary, the "CLL1 panel" (CD23, FMC-7, and CD200) is subsequently employed to differentiate CLL from MCL. In one study, ensemble learning techniques were employed to develop a model that predicts the requirement for ordering the additional CLL1 antibody panel based on flow cytometry data. A dataset of 9635 patient samples was used, with 887 cases (9.2%) requiring further testing with the CLL1 panel, divided in an 80:20 ratio for training and validation sets. Within the ensemble learning approach, the training dataset was further split into a 67:33 ratio for training the CNN models and integrating the RF model, respectively. Evaluation of the model was performed using a confusion matrix on the validation set, yielding an AUROC of 92%, ACC of 94%, SEN of 53%, and SPE of 58%. Furthermore, prospective evaluation was conducted by generating real-time predictions on 376 sequential cases where the screening antibody panel was ordered. The model produced predictions within 3 min of data upload from the flow cytometer, achieving an AUROC of 89%, ACC of 94%, SEN of 78%, and SPE of 95% [24]. Implementation of this model has the potential to enhance laboratory efficiency by identifying cases requiring additional testing, thereby reducing review time for pathologists and lab technicians, as well as minimizing turnaround time for sample processing. However, it should be noted that the model's performance is currently insufficient for clinical purposes, potentially due to the absence of information regarding previous diagnoses of CLL/SLL or MCL. Additionally, since the task at hand is relatively straightforward, alternative ML approaches may offer sufficient accuracy in predicting the need for additional testing. It is crucial to acknowledge that the models in this study were trained and tested using data from a single laboratory, thus susceptible to interinstitutional variability due to different laboratory

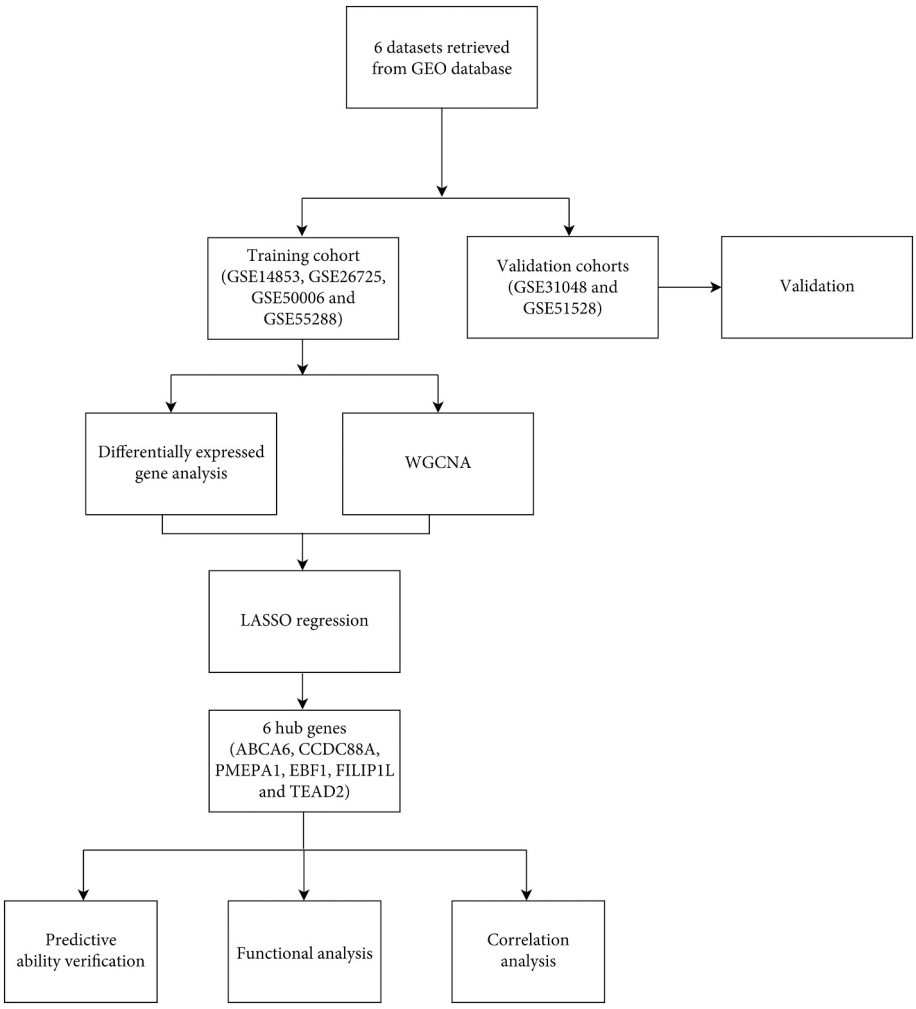


Fig. 2. Flowchart of the identification and assessment of candidate genes (Zhu et al., 2022) [18].

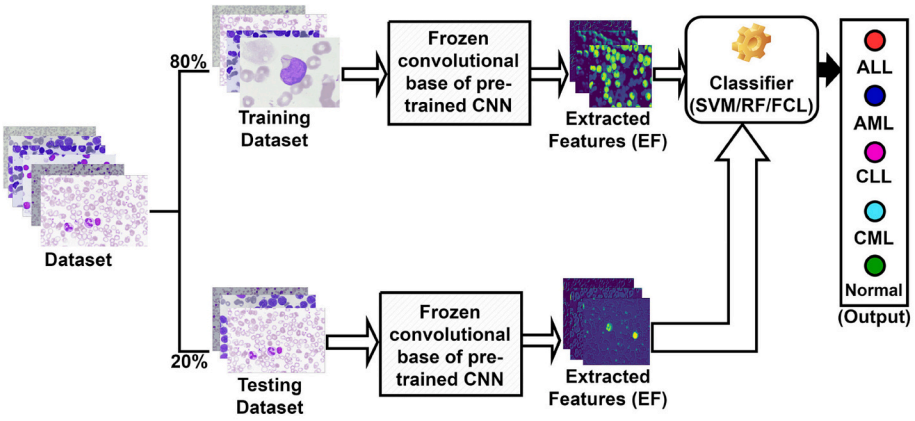
practices, equipment, and specific patient characteristics of the analyzed samples. Future research should address these issues to improve model's validity and consider implementation in clinical practice.

Another study focused on the classification of B-cell malignancies using flow cytometry. The dataset consisted of peripheral blood samples from 3417 cases encompassing different diagnostic classes (normal, BNHL, B-ALL/LBL, and CLL). The data was divided into training and validation sets in an 80:20 ratio. To visualize the data, UMAP was employed to generate two-dimensional projections, which were then converted to histograms and utilized for training a RF classifier. The RF classifier underwent ten-fold time-series cross-validation, demonstrating accuracy, sensitivity, and specificity values of 96.4%, 86.9%, and 98.3% respectively, for classifying CLL cases. Moreover, the model successfully identified cases requiring the ordering of mature B-cell and B-ALL/LBL add-on tubes to confirm the diagnosis, achieving accuracies of 96.5% and 100%, respectively [25]. The developed model shows significant potential in streamlining the workflow of hematopathologists and lab technicians. By assisting in the interpretation of flow cytometry data and flagging cases that may necessitate additional confirmatory tests, it can greatly facilitate their tasks. However, a drawback of this technique is the challenge of troubleshooting misclassifications made by the classifier, even with manual review of the flow data. Thus, further training and experimentation with such models is justified to address this issue and enhance the classifier's performance.

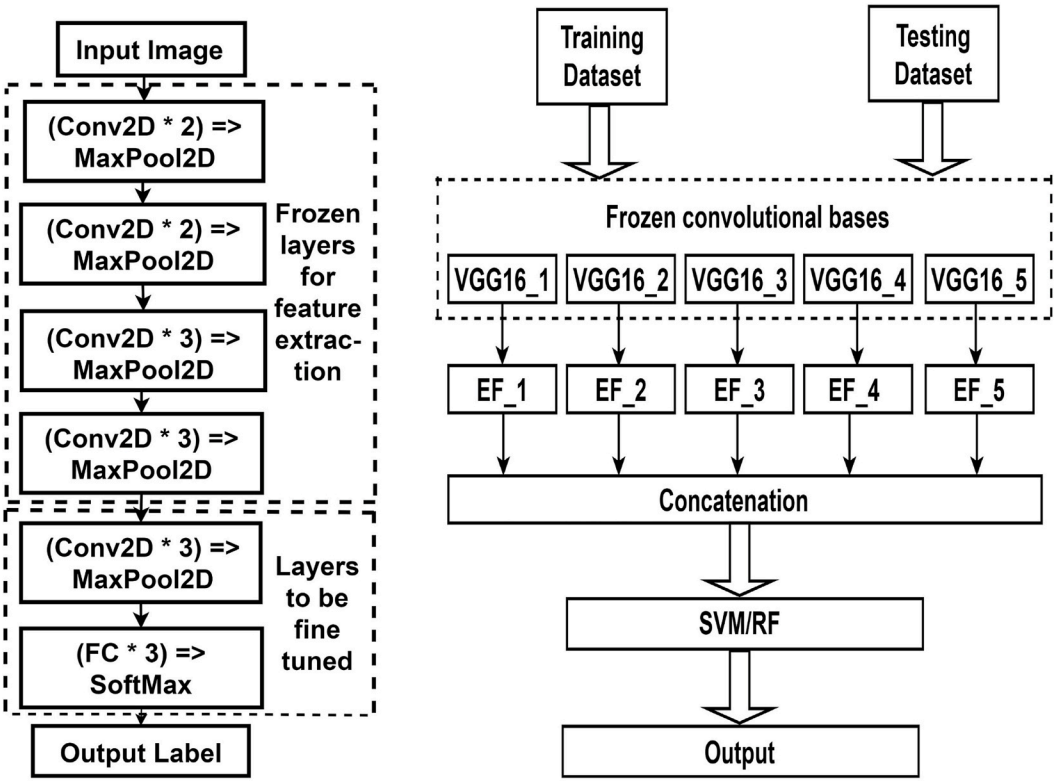
Zhao et al. conducted a study with the objective of automating the diagnosis and classification of B-cell neoplasms, employing a similar methodology to that described in [25]. The framework for the proposed classification system is outlined in Fig. 4. The study utilized flow cytometry data from a cohort of 20,622 patients, with 18,274 cases allocated for training and validation, and a hold-out set of 2348 cases for model development. The goal was to build a model capable of classifying nine diagnostic classes, namely CLL, MBL, MCL, MZL, PL, FL, HCL, LPL, and normal. To reduce the dimensionality of the data while maintaining the topological relationships between data points, a self-organizing map (SOM) algorithm was applied. The resulting SOMs were then utilized as input for a CNN classifier (Fig. 5) to generate predictions. Notably, the classifier achieved a weighted F1 score of 0.94 for the nine diagnostic classes against the hold-out set, with an 83% accuracy in classifying CLL cases [26].

3.4. Using hematological analyzers

The clinical significance of morphological characteristics of cells in various hematological diseases is well documented in the literature [27–31]. These characteristics alongside basic complete blood count (CBC) parameters can be quantitatively assessed using modern hematological analyzers. In this regard, Haider et al. took advantage of the cell population data (CPD) collected by advanced hematological analyzers to develop an ML model capable of predicting and subtyping leukemia. The dataset encompassed 1577 cases with various hematological diseases, including 153 cases of CLL. The hematological analysis of blood samples was done using Sysmex XN-Module. The study



(a)



(b) (c)

Fig. 3. (a) Outline of the proposed methodology, (b) fine tuning of VGG16, and (c) concatenation of extracted features to train and test SVM and RF "Reproduced with permission from Abhishek et al., Biomedical Signal Processing and Control; Published by Elsevier, 2023" [21].

employed radial basis function network (RBFN), an artificial neural network architecture, for predictive modeling of hematological malignancies within the dataset. The RBFN model achieved an AUROC of 90.5% in classification of CLL [32]. This study showcases the high accuracy of pre-microscopic prediction and classification of leukemia, suggesting the potential utility of the RBFN model as a screening tool. Its implementation could optimize the ordering process for relevant diagnostic tests and facilitate early referral and treatment for patients with leukemia. However, it is important to note that the model's validation was limited to internal validation. Therefore, to comprehensively assess its performance on novel data, further validation on an independent dataset is necessary.

3.5. *Using lymph nodes histopathological images*

Steinbuss et al. conducted a study to automate the classification of NHLs (non-Hodgkin lymphomas) using deep learning techniques. The study involved histopathological images of lymph nodes obtained from 629 patients, which were subsequently annotated by hematopathologists as diffuse large B-cell lymphoma (DLBCL), CLL/SLL, or normal. The EfficientNetB3, a convolutional neural network (CNN) architecture, was employed for training and evaluating the model. In the independent test set, the model demonstrated favorable accuracy and specificity in classifying nodal CLL/SLL. However, the model's sensitivity was inadequate, limiting its potential as a screening tool within routine clinical settings [33]. On the other hand, the model developed by do Nascimento

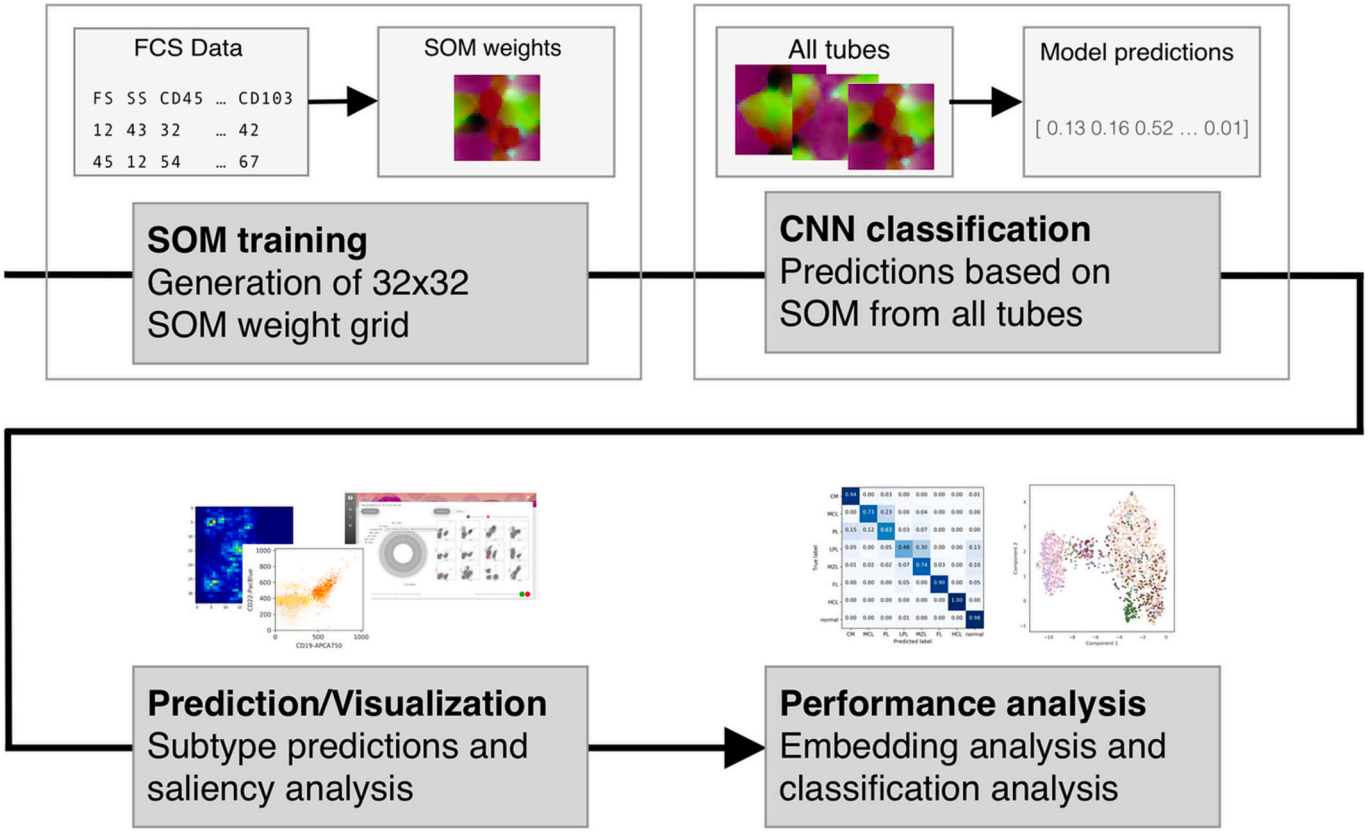


Fig. 4. Overview of the classification pipeline. Individual 2D SOMs are generated for each tube of a single case. The weights of the SOM nodes are used as input for a CNN that predicts lymphoma subtypes. The trustworthiness of a suggested diagnosis is computed and the cells contributing most to this decision are visualized in density plots and saliency maps. The overall performance of the classification process is benchmarked with a confusion matrix and the similarity of cases is visualized by a t-Distributed Stochastic Neighbor Embedding (t-SNE) plot "Reproduced with permission from Zhao et al., Cytometry Part A; Published by John Wiley and Sons, 2020" [26].

et al. [34] demonstrated outstanding performance in classifying lymph node images of NHL. Using 375 images (113 CLL, 140 FL, and 122 MCL), the model achieved perfect accuracy, sensitivity, specificity, and AUROC of 100%. This was achieved by employing a polynomial (PL) classifier with the Ansari-Bradley (AB) technique for feature selection, effectively capturing relevant features while eliminating redundant ones. Fig. 6 plots the AUROCs obtained from different classifiers using the AB technique. However, it is important to note that the dataset used in this study was limited to images extracted from only 10 patients. Consequently, the model trained on such small sample size is susceptible to misclassification when encountering new cases, which raises concerns about overfitting. To address this limitation and enhance the model's reliability, it is imperative to conduct further training and testing on larger and more diverse datasets. Only through such rigorous evaluation can the model be deemed suitable for integration into the healthcare system.

Deep learning (DL) has been extensively studied for diagnostic image classification [35–38]. However, many DL models suffer from poor classification performance due to inadequate data preprocessing and feature extraction methods. To address this, a study aimed to enhance the classification accuracy of a neural network model by incorporating image preprocessing techniques, transfer learning, and principal component analysis (PCA). The study utilized an online dataset consisting of 374 histopathology images of CLL, FL, and MCL [39]. The dataset was divided into training, validation, and test sets with a split ratio of 72:8:20, respectively. Fig. 7 represents the proposed framework for the classification of NHLs using histopathological images. Initially, the images underwent image segmentation for preprocessing, followed by transfer learning to fine-tune and extract features using four transfer models (VGG-16, VGG-19, ResNet-50, and DenseNet-121). Among these models, VGG-16 demonstrated superior performance and was chosen to extract features, which were subsequently mapped using PCA. The mapped features were then fed into a neural network model for training and evaluation. Notably, the neural network model achieved an accuracy of 99.3% in classifying CLL cases [40]. These findings highlight the potential of integrating advanced techniques, such as image preprocessing, transfer learning, and PCA, to significantly enhance the performance of DL models for diagnostic image classification. Future research efforts should focus on training and evaluating such models on larger datasets and to include other disease entities that can be diagnosed using this technique.

3.6. *Using Raman spectroscopy*

A unique diagnostic approach for chronic lymphocytic leukemia (CLL) was developed by Féré et al. [41], using Raman spectroscopy data obtained from blood smears. The study included Raman data from 140 patients, consisting of 61 healthy individuals and 79 diagnosed with CLL. Their objective was to create an ML model capable of accurately classifying patients as either healthy or having CLL. To achieve this, the data underwent preprocessing procedures to ensure its quality, followed by the division into two separate datasets for model training and evaluation. Dataset 1, which comprised 100 patients (41 healthy and 59 CLL cases), served as the training and validation set, while dataset 2 included 40 patients (20 healthy and 20 CLL cases) and functioned as an independent test set to assess the model's performance on previously unseen data. To extract the most discriminative features between CLL and healthy cells, canonical correlation analysis was employed.

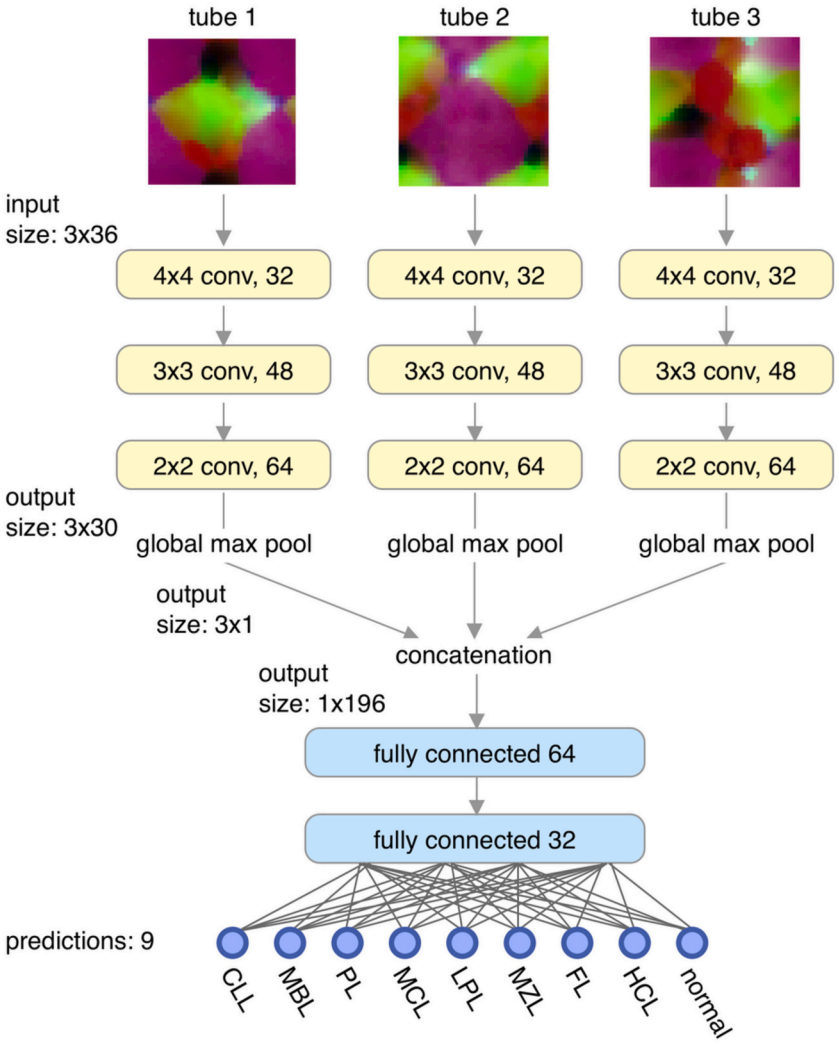


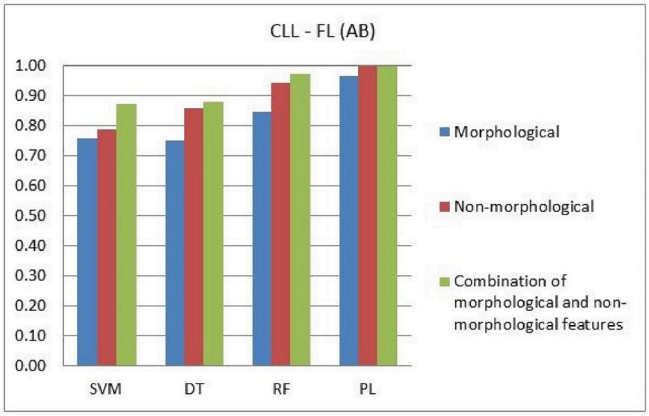
Fig. 5. Architecture of the CNN. First, the original 32 × 32 SOMs are toroidally wrapped by two pixels on each edge to produce a 36 × 36 input matrix, which is fed into convolutional layers with 32 4 × 4 filters. The input from each SOM is processed individually in a sequence of convolutional layers (conv), followed by a global max pooling and concatenation layer. This vector is further processed in two fully connected hidden layers and results in a softmax prediction layer "Reproduced with permission from Zhao et al., Cytometry Part A; Published by John Wiley and Sons, 2020" [26].

Subsequently, they evaluated three supervised classification algorithms: partial least squares-discriminant analysis (PLS-DA), SVM, and RF. PLS-DA emerged as the top-performing model among them and was consequently selected for the classification task. Moreover, to address concerns of overfitting and optimize the model's parameters, repeated double cross-validation (rdCV) was implemented [42]. The rdCV method generated one hundred optimized models, and the final output was determined by majority voting based on the predictions of these models. The resulting algorithm attained a sensitivity of 95% and specificity of 85% in accurately classifying healthy and CLL patients. This method exhibits considerable accuracy in identifying cases of CLL, and it can be tuned to optimize sensitivity or achieve a balance between sensitivity and specificity, depending on the clinical context. Moreover, leveraging Raman data allows for the investigation of metabolic and biochemical alterations within CLL cells through the correlation of spectral bands with specific proteins and nucleic acids. This contributes to advancing our comprehension of the disease's pathophysiology and facilitates the development of personalized treatments. However, before considering the integration of this method into clinical settings, a comprehensive cost-benefit analysis is essential, considering that Raman spectroscopy is not yet a routine practice in the evaluation of CLL patients.

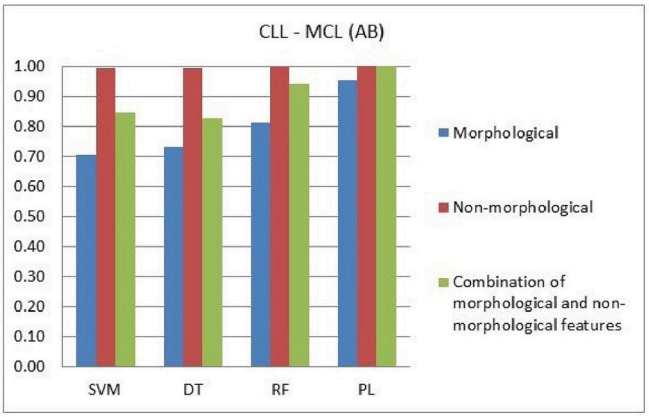
4. Conclusion and future considerations

This review explored various approaches to enhance the diagnosis of CLL through the implementation of ML algorithms. A critical evaluation is conducted on multiple applications of ML models in CLL diagnosis using blood smears, flow cytometry, histopathological images, genetic data, and others. The current evidence suggests that AI can accurately predict CLL diagnosis, aid in CLL screening, identify potential biomarkers for diagnosis, and explore the underlying biochemical and molecular mechanisms in CLL. The majority of the ML models assessed in this review exhibit adequate performance in predicting CLL diagnosis. Specifically, the leading model in discriminating CLL from healthy cases demonstrated a sensitivity of 96.4% and specificity of 98.8%, along with an AUROC of 99.7%. Moreover, one model achieved 100% accuracy in distinguishing CLL from FL and MCL, while another model achieved the same level of performance in distinguishing CLL from CML, ALL, and AML.

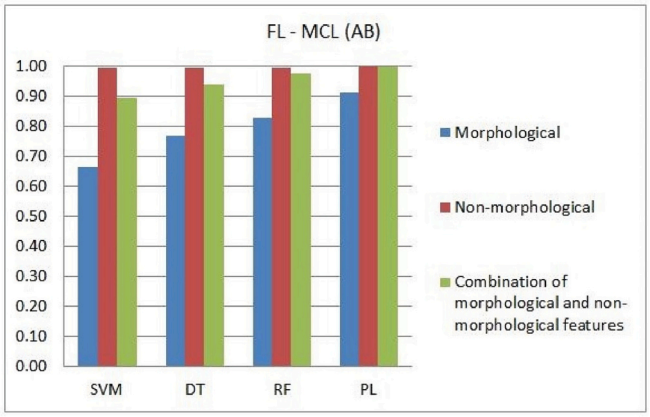
The application of AI and ML in the field of hematology offers a wide range of benefits. These technologies have the potential to enhance the efficiency and effectiveness of hematologists by automating various steps involved in patient workup, risk assessment, and treatment. Through the automation of these processes, hematologists can reallocate



(a)



(b)



(c)

Fig. 6. The AUC metric obtained with the AB method and the classifiers with the investigated lesions groups: (a) CLL-FL; (b) CLL-MCL and (c) FL-MCL "Reproduced with permission from do Nascimento et al., Computer Methods and Programs in Biomedicine; Published by Elsevier, 2018" [34].

their resources toward other critical aspects of clinical practice requiring human judgment, intuition, and empathy, such as patient care and research.

In the realm of hematological malignancies, ML algorithms hold promise for improving patient care through activities such as screening, early diagnosis, risk stratification, treatment recommendations, and prognosis prediction [43–45]. Specifically, our review highlights the potential integration of ML algorithms into clinical practice for the diagnosis and workup of patients with CLL.

Despite the promising performance and potential advantages of employing ML models in CLL diagnosis, several important considerations need to be acknowledged [44,46–48]. Several reviewed models had limited sample sizes derived from a single center or laboratory, limiting their generalizability to other populations. Additionally, some studies only validated their models internally, thereby increasing the risk of overfitting and rendering their performance on unseen data unknown. To address this limitation, it is imperative to develop models with enhanced generalizability by employing large, homogeneous datasets obtained from multiple centers and laboratories.

In addition, most studies retrospectively evaluated the predictive abilities of ML models, with few prospective studies available. Moreover, there is a lack of research evaluating the influence of these models on patient outcomes. Future investigations should focus on prospective assessment of the effect of ML models on CLL diagnosis, patient prognosis, and ultimately, patient outcomes.

Finally, the integration of ML applications into direct patient care raises various ethical and medico-legal concerns. These issues encompass liability in case of medical errors, data privacy and security, the doctor-ML application interaction, comprehension of the capabilities and limitations of ML, as well as patient understanding of ML utilization in healthcare and its potential effects. To address these concerns, the development of an ethical framework specific to the clinical context of ML applications in healthcare is crucial. Furthermore, ML algorithms should be employed as aids to healthcare practitioners, complementing their role rather than replacing it. Doctors should undergo training and education on ML applications, including awareness of the variables considered and the sensitivity and specificity of the algorithms for specific tasks. Through the effective resolution of these issues, the successful integration of ML algorithms into the care of CLL patients can be accomplished.

Practice points

- ML can learn, distinguish patterns, and make decisions through analyzing input data. Patients with CLL can benefit from the applications of ML algorithms.
- Multiple ML applications have been developed to aid in the diagnosis and workup of CLL using images, genetic data, flow cytometry, hematological analyzers, and Raman spectroscopy.
- ML algorithms were incorporated in genetic analyses to delineate genetic biomarkers or mutation profiles that differentiate CLL from other neoplastic entities or normal samples.
- ML and DL algorithms were able to analyze patterns present in blood smear and lymph node histopathological images and achieved high performance in differentiating CLL images from non-CLL ones.
- Ensemble learning was employed to predict the need for ordering additional antibody panel to differentiate CLL from MCL using flow cytometry data.
- Despite their advantages, current AI-based methods are far from replacing hematologists' workup and assessment but can improve their efficiency and decision-making by analyzing data generated from different diagnostic modalities and providing a diagnostic outcome, or by performing tedious straightforward tasks.

Research agenda

- Development of large homogeneous datasets derived from multiple populations and settings to train models with high generalizability.
- Assessment of available models prospectively to evaluate their effect on patient outcomes.
- Investigation of ethical and medico-legal implications of using ML applications in patient care.

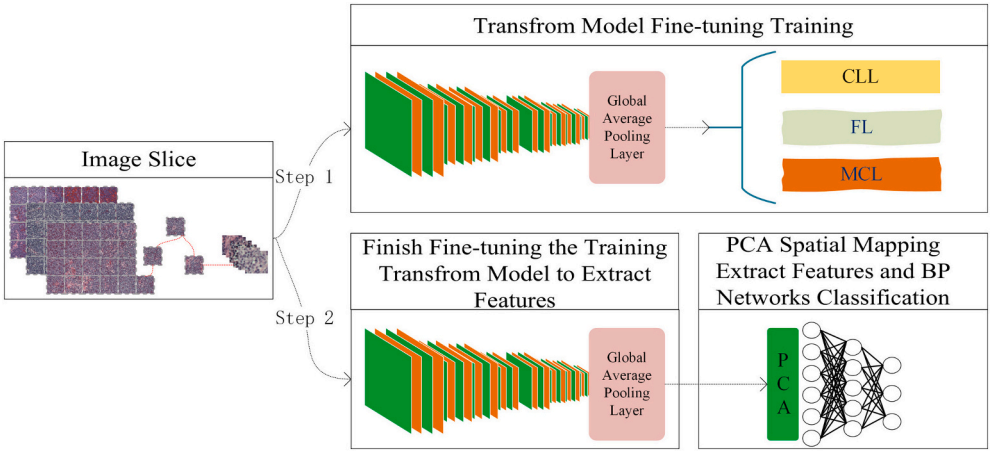


Fig. 7. Proposed framework for NHL classification based on the fusion of TL and PCA. The datasets were preprocessed first and then input to preselected transfer models for fine-tuning. The fine-tuned models were used to extract features, and the resulting features were processed by PCA for mapping. Finally, the mapped features were input to the neural network model for training, prediction, and evaluation "Reproduced with permission from Zhang et al., Medical Physics; Published by John Wiley and Sons, 2020" [40].

- Establishment of an ethical framework that governs the use of ML application in clinical settings.

Funding

Open Access funding provided by Qatar National Library.

Declaration of Competing Interest

All authors declare no conflict of interest.

References

[1] Strati P, Jain N, O'Brien S. Chronic lymphocytic Leukemia: diagnosis and treatment. Mayo Clin Proc 2018;93(5):651–64.
[2] Ghia P, Caligaris-Cappio F. Monoclonal B-cell lymphocytosis: right track or red herring? Blood 2012;119(19):4358–62.
[3] Siegel RL, et al. Cancer statistics, 2023. CA Cancer J Clin 2023;73(1):17–48.
[4] Yao Y, et al. The global burden and attributable risk factors of chronic lymphocytic leukemia in 204 countries and territories from 1990 to 2019: analysis based on the global burden of disease study 2019. Biomed Eng Online 2022;21(1):4.
[5] Hallek M, et al. iwCLL guidelines for diagnosis, indications for treatment, response assessment, and supportive management of CLL. Blood 2018;131(25):2745–60.
[6] Shadman M. Diagnosis and treatment of chronic lymphocytic leukemia: a review. JAMA 2023;329(11):918–32.
[7] Rai KR, et al. Clinical staging of chronic lymphocytic leukemia. Blood 1975;46(2): 219–34.
[8] Binet JL, et al. A new prognostic classification of chronic lymphocytic leukemia derived from a multivariate survival analysis. Cancer 1981;48(1):198–206.
[9] Shanafelt TD. Predicting clinical outcome in CLL: how and why. Hematology Am Soc Hematol Educ Program 2009:421–9.
[10] Walter W, et al. Artificial intelligence in hematological diagnostics: game changer or gadget? Blood Rev 2023;58:101019.
[11] Elsabagh A, et al. Artificial intelligence in sickle disease. Blood Rev 2023:101102.
[12] Ferih K, et al. Applications of artificial intelligence in Thalassemia: a comprehensive review. Diagnostics (Basel) 2023;13(9).
[13] Elhadary M, et al. Applications of machine learning in chronic myeloid Leukemia. Diagnostics (Basel) 2023;13(7).
[14] Shouval R, et al. Machine learning and artificial intelligence in haematology. Br J Haematol 2021;192(2):239–50.
[15] Elsayed B, et al. Applications of artificial intelligence in Philadelphia-Negative myeloproliferative neoplasms. Diagnostics (Basel) 2023;13(6).
[16] Elshoeibi AM, et al. Applications of artificial intelligence in Thrombocytopenia. Diagnostics (Basel) 2023;13(6).
[17] Zhang H, et al. Differential diagnosis of hematologic and solid tumors using targeted transcriptome and Artificial Intelligence 2023;193(1):51–9.
[18] Zhu Y, et al. Identification of Six Diagnostic Biomarkers for Chronic Lymphocytic Leukemia Based on Machine Learning Algorithms2022; 2022. p. 3652107.
[19] Xia D, et al. DNA methylation-based classification of small B-Cell Lymphomas: a proof-of-principle study 2021;23(12):1774–86.
[20] Iman M, Arabnia HR, Rasheed K. A review of deep transfer learning and recent advancements 2023;11(2):40.
[21] Abhishek A, et al. Automated detection and classification of leukemia on a subject-independent test dataset using deep transfer learning supported by Grad-CAM visualization83; 2023.
[22] Dese K, et al. Accurate machine-learning-based classification of leukemia from blood smear images. Clin Lymphoma Myeloma Leuk 2021;21(11):e903–14.
[23] Mohammed EA, et al. Toward leveraging big value from data: chronic lymphocytic leukemia cell classification 2017;6(1).
[24] Simonson PD, Lee AY, Wu D. Potential for process improvement of clinical flow cytometry by incorporating real-time automated screening of data to expedite addition of antibody panels 2022;157(3):443–50.
[25] Ng DP, Zuromski LM. Augmented human intelligence and automated diagnosis in flow cytometry for hematologic malignancies 2021;155(4):597–605.
[26] Zhao M, et al. Hematologist-level classification of mature B-cell neoplasm using deep learning on multiparameter flow cytometry data 2020;97(10):1073–80.
[27] Silva M, et al. Lymphocyte volume and conductivity indices of the haematology analyser Coulter® GEN.STM in lymphoproliferative disorders and viral diseases 2006;28(1):1–8.
[28] Park SH, et al. Sepsis affects most routine and cell population data (CPD) obtained using the Sysmex XN-2000 blood cell analyzer: neutrophil-related CPD NE-SFL and NE-WY provide useful information for detecting sepsis. Int J Lab Hematol 2015;37 (2):190–8.
[29] Jung YJ, et al. Evaluation of cell population data on the UniCel DxH 800 coulter cellular analysis system as a screening for viral infection in children. Int J Lab Hematol 2012;34(3):283–9.
[30] Furundarena JR, et al. The utility of the Sysmex XE-2100 analyzer's NEUT-X and NEUT-Y parameters for detecting neutrophil dysplasia in myelodysplastic syndromes. Int J Lab Hematol 2010;32(3):360–6.
[31] Haschke-Becher E, et al. A new high-throughput screening method for the detection of chronic lymphatic leukemia and myelodysplastic syndrome. Clin Chem Lab Med 2008;46(1):85–8.
[32] Haider RZ, et al. Beyond the in-practice CBC: the research CBC parameters-driven machine learning predictive modeling for early differentiation among leukemias 2022;12(1).
[33] Steinbuss G, et al. Deep learning for the classification of non-Hodgkin Lymphoma on histopathological images 2021;13(10).
[34] Do Nascimento MZ, et al. Lymphoma images analysis using morphological and non-morphological descriptors for classification163; 2018. p. 65–77.
[35] Akkus Z, et al. Deep learning for brain MRI segmentation: state of the art and future directions. J Digit Imaging 2017;30(4):449–59.
[36] Ehteshami Bejnordi B, et al. Using deep convolutional neural networks to identify and classify tumor-associated stroma in diagnostic breast biopsies. Mod Pathol 2018;31(10):1502–12.
[37] Wang X, et al. Searching for prostate cancer by fully automated magnetic resonance imaging classification: deep learning versus non-deep learning. Sci Rep 2017;7(1):15415.
[38] Shin HC, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans Med Imaging 2016;35(5):1285–98.
[39] Janowczyk A, Madabhushi A. Deep learning for digital pathology image analysis: a comprehensive tutorial with selected use cases. J Pathol Inform 2016;7:29.
[40] Zhang J, et al. Classification of digital pathological images of non-Hodgkin's lymphoma subtypes based on the fusion of transfer learning and principal component analysis 2020;47(9):4241–53.
[41] Féré M, et al. Implementation of a classification strategy of Raman data collected in different clinical conditions: application to the diagnosis of chronic lymphocytic leukemia 2020;412(4):949–62.

[42] Filzmoser P, Liebmann B, Varmuza K. Repeated double cross validation 2009;23 (4):160–71.
[43] Obstfeld AE. Hematology and machine learning. J Appl Lab Med 2023;8(1): 129–44.
[44] Muhsen IN, et al. Machine learning applications in the diagnosis of benign and malignant hematological diseases. Clin Hematol Int 2021;3(1):13–20.
[45] Radakovich N, Nagy M, Nazha A. Artificial intelligence in hematology: current challenges and opportunities. Curr Hematol Malig Rep 2020;15(3):203–10.
[46] Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. Lancet Oncol 2019;20(5):e262–73.
[47] Hedderich DM, et al. Artificial intelligence tools in clinical neuroradiology: essential medico-legal aspects. Neuroradiology 2023;65(7):1091–9.
[48] Eckardt J-N, et al. Application of machine learning in the management of acute myeloid leukemia: current practice and future prospects. Blood Adv 2020;4(23): 6077–85.