Application of machine learning in the management of lymphoma: Current practice and future prospects

Junyun Yuan[1], Ya Zhang[1,2,3] and Xin Wang[1,2,3,4,5] (iD)

Abstract

In the past decade, digitization of medical records and multiomics data analysis in lymphoma has led to the accessibility of high-dimensional records. The digitization of medical records, the visualization of extensive volume data extracted from medical images, and the integration of multiomics methods into clinical decision-making have produced many datasets. As a promising auxiliary tool, machine learning (ML) intends to extract homologous features in large-scale data sets and encode them into various patterns to complete complicated tasks. At present, artificial intelligence and digital mining have shown promising prospects in the field of lymphoma pathological image analysis. The paradigm shift from qualitative analysis to quantitative analysis makes the pathological diagnosis more intelligent and the results more accurate and objective. ML can promote accurate lymphoma diagnosis and provide patients with prognostic information and more individualized treatment options. Based on the above, this comprehensive review of the general workflow of ML highlights recent advances in ML techniques in the diagnosis, treatment, and prognosis of lymphoma, and clarifies the boundedness and future orientation of the ML technique in the clinical practice of lymphoma.

Keywords

lymphoma, artificial intelligence, machine learning, diagnosis, prognosis, treatment

Submission date: 4 December 2023; Acceptance date: 28 March 2024

Introduction

Lymphoma is a set of biologically and clinically heterogeneous diseases that has complex types and is typically characterized by painless lymphadenopathy with systemic symptoms such as fever, night sweats, and weight loss. According to the pathological manifestations, it is mainly divided into Hodgkin's lymphoma (HL) and non-Hodgkin's lymphoma (NHL). NHL includes mantle cell lymphoma, follicular lymphoma (FL), marginal zone lymphoma, and the most common diffuse large B-cell lymphoma (DLBCL). Although the survival of lymphoma patients has been dramatically improved with the increasing progress of diagnostic technology and the promotion of standardized treatment, the disease burden of NHL is still higher than that of HL.[1]

The digitization of data has caused the emergence of high-dimensional datasets. The advent of next-generation

[1]Department of Hematology, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, Shandong, China
[2]Department of Hematology, Shandong Provincial Hospital, Shandong University, Jinan, Shandong, China
[3]Taishan Scholars Program of Shandong Province, Jinan, Shandong, China
[4]Branch of National Clinical Research Center for Hematologic Diseases, Jinan, Shandong, China
[5]National Clinical Research Center for Hematologic Diseases, Hospital of Soochow University, Suzhou, China

Corresponding authors:
Xin Wang, Department of Hematology, Shandong Provincial Hospital Affiliated to Shandong First Medical University, No. 324, Jingwu Road, Jinan, Shandong, 250021, China.
Email: xinw007@126.com

Ya Zhang, Department of Hematology, Shandong Provincial Hospital Affiliated to Shandong First Medical University, No. 324, Jingwu Road, Jinan, Shandong, 250021, China.
Email: maryzhangya@gmail.com

sequencing (NGS) technology has increased the perceived value of molecular genetic results in the diagnosis of lymphoma by speeding up the identification of new molecular changes. Under the World Health Organization classification guidelines, the diagnosis of lymphoma is in conjunction with relevant indicators such as cytomorphology, cytogenetics, immunophenotyping, and molecular genetics for a comprehensive evaluation.[2] Machine learning (ML) can manipulate large volumes of data that transcend human understanding.[3] It can integrate complex features of the disease and screen for people at high risk of developing cancer to mitigate unnecessary and costly diagnostic interventions.

The growing number of prognostic markers and risk stratification may guide patient management, facilitate their stratification into clinical practice, and pave the way for personalized medicine. For instance, in recent studies, CloneRetriever, a novel web-based model based on ML principles is as demonstrated to address the bioinformatics challenges associated with immune cell sequencing and clone detection. It is another promising application to distinguish clonal rearrangement in large-scale NGS data sets for auxiliary diagnosis of lymphoma.[4] The continuous efforts of ML in lymphoma in recent years may eventually affect the design and development of drugs so that patients face more targeted treatment options. ML and artificial intelligence (AI) assisted in diagnosing lymphoma in radiomics, genomics, proteomics, and histopathology. In addition, ML is also involved in lymphoma targeted therapy, and prognosis analysis (Figure 1). Thus, herein, we describe significant concepts in AI and ML for unfamiliar readers and summarize some studies of ML and AI in the diagnosis and prognosis of lymphoma. The purpose is to guide lymphoma patients to obtain more treatment selections and improve the survival rate.

Essentials of AI

AI has frequently appeared in healthcare for more than a decade with the rapid development of computer science.[5] However, the contemporary concepts of AI, ML, and deep learning (DL) are often equivocal and should be distinguished. AI is a computer program that enables the calculating machine to simulate certain human thinking processes and intelligent behaviors to assist us in finishing off complex problems. ML is a subdomain of AI and a promising auxiliary tool that aims at similar features in datasets. Through these characteristics, various models are established to complete complex tasks. As a thriving development direction, natural language processing (NLP), large language models (LLMs), explainable AI (XAI), and generative AI all represent specific branches or technologies in the field of AI. These branches are interrelated but each has unique characteristics and application areas (Figure 2).

On the basis of the relationship between computer program input data labels and output results, ML can be divided into three categories: supervised learning, unsupervised learning, and reinforcement learning. Supervised learning is to use labeled data to get a model through learning and training, and then uses this model to predict new samples to achieve the purpose of classification (e.g. classifying subtypes of lymphoma) or regression (e.g. estimating the length of survival). Unsupervised learning aims to automatically learn hidden patterns or structures from data without the need for predetermined labels or targets. Reinforcement learning also uses unmarked data to interact with the environment through behavior, and the environment obtains the best prediction behavior strategy in the continuous cycle through the reward function.

In this review, we present the general flow of supervised ML tasks. Supervised ML generally consists of two phases, first a set of image datasets with known classifications with corresponding class labels is used to generate a training prediction model through data cleaning, feature extraction, and other steps, followed by an application phase. This ML model is applied to the newly acquired medical images and eventually, is applied for the classification and diagnosis of lymphoma (Figure 3). This significantly improves the efficiency of diagnosing lymphoma in clinical practice. In summary, ML is traditionally defined as a process of self-improvement of a system. The machine represents the computer, and learning is a self-improvement process.[6] DL is an algorithm that attempts to abstract data using multiple processing layers composed of complex structures or multiple nonlinear transformations.[7] DL is mainly used in voice and image recognition. As a hot research field of ML, it mainly studies the internal law and representation level of sample data. The contemporary notion of DL originates from an artificial neural network（ANN）, which is an algorithm for pattern recognition. The traditional multilayer neural network has an only input, hidden, and output layer. There is no clear theoretical derivation to explain how many layers are appropriate. The steps of the multilayer neural network involve feature mapping to values, and feature selection is done manually. Compared with an ANN, it compresses the size of input data through the convolution layer and pooling layer, that is, the data is divided into many small data, and the maximum utilization is obtained by using fewer parameters. DL shows a high level of computer-aided diagnosis of NHL.[8] Modern DL algorithms comprise convolution neural network (CNN), cyclic neural network, recursive neural network, long-term and short-term memory, and deep belief network.[9] Deep convolution neural network has achieved good results in processing the task of medical image data classification.[10,11]

Applications of ML in lymphoma diagnosis

Over the past decade, in the field of healthcare, computer and electronic healthcare data (such as electronic medical records, laboratory test results, imaging examination, etc.)
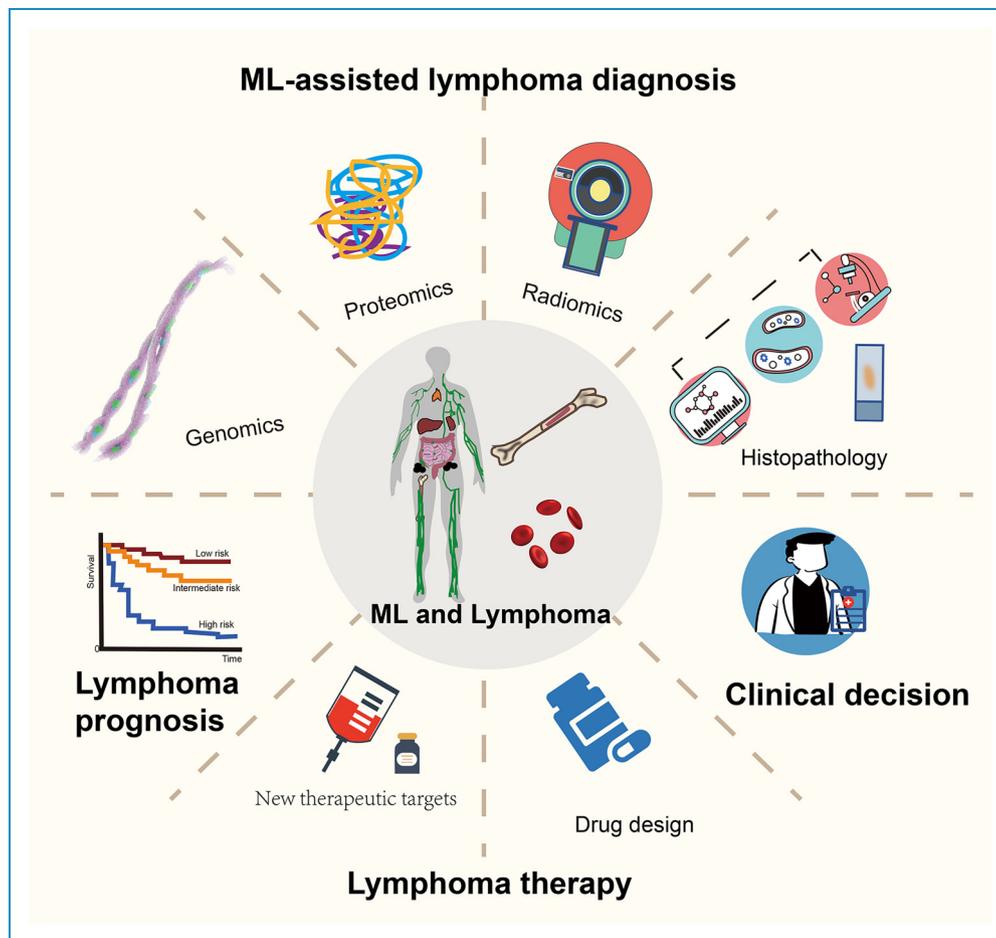
Figure 1. Schematic diagram of the applications of machine learning in lymphoma. It mainly introduces the applications of machine learning in the diagnosis of lymphoma in radiomics, genomics, histopathology, etc. In addition, Machine learning is also involved in the targeted therapy, drug design, prognosis analysis, and clinical decision of lymphoma.

of the fusion combined with ML tools improved the accuracy of lymphoma diagnosis and increased the possibility of personalized medicine.[12] Some applied research of AI in lymphoma are listed in Table 1.

ML and digital pathology

The advent of digital pathology has ushered in a plethora of possibilities, one of which involves leveraging AI for decision support. Employing AI in the interpretation of digital pathology data presents a promising avenue for assisting pathologists in the diagnosis of lymphoma. Pathology is an essential examination method for the diagnosis of malignant lymphoma.[13] Automatic image preprocessing reduces the overall workload of hematologists and pathologists. The initial attempt is to start with image preprocessing, then object recognition, feature extraction, and finally classify different cells. It is usually necessary to remove the diseased lymph nodes for histopathological examination to determine the subtype.[14] Therefore, pathologists who need to

survey pathological images to determine the subtype of lymphoma require a commitment of time and energy.[15] It is difficult to draw a correct conclusion quickly only by traditional naked-eye observation and subjective judgment. This challenges the experience and ability of technicians to quickly and accurately detect abnormal cells. In addition, the number of pathologists is decreasing, and not all hematopathologists can rely on rich experience and expensive equipment for molecular analysis.[16,17] It is gratifying that AI can establish a fast and accurate classifier to help pathologists analyze and calculate pathological images, obtain correct relevant data, and help clinicians make a clinical diagnosis. An ML model with texture analysis has been discovered in a study for the automatic classification of histopathological images depicting ocular adnexal mucosa-associated lymphoid tissue lymphoma from two distinct origins.[18]

As DL invades the field of image recognition, the availability of digital images is also improved. We can rely on algorithms to distinguish artifacts and information
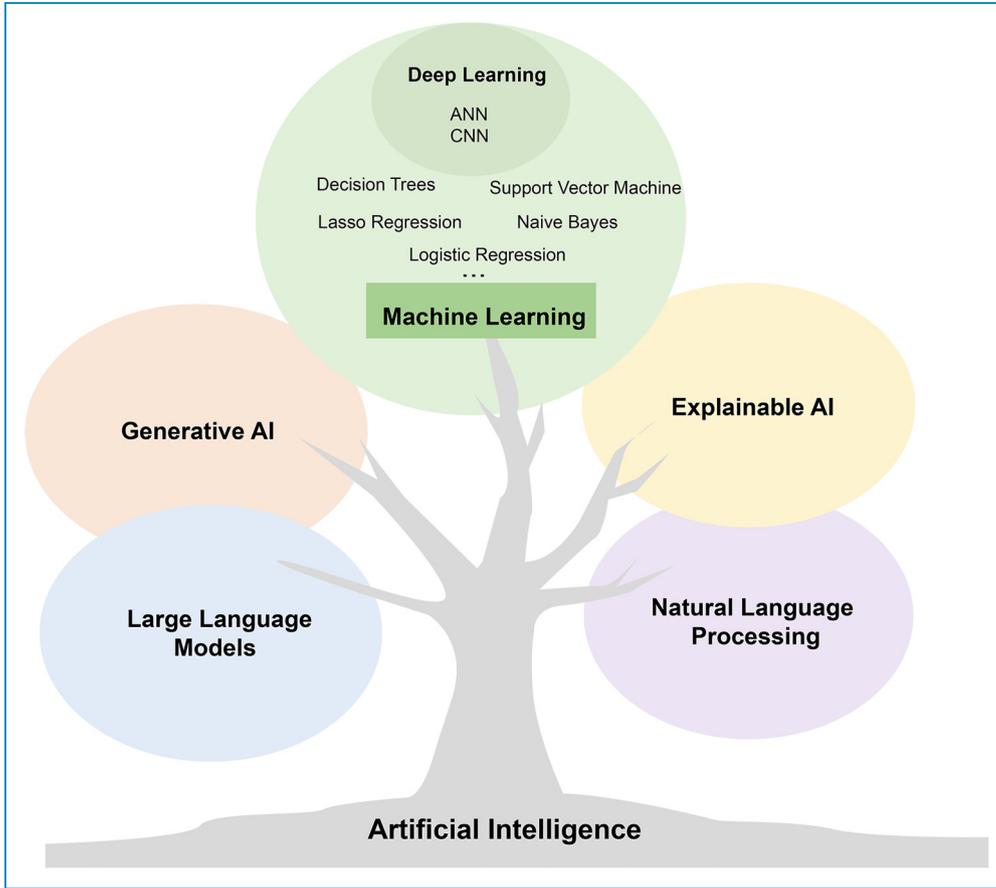
Figure 2. Branches and evolution of artificial intelligence technology. The diagram shows a tree representing artificial intelligence (AI), with five fruits on it. The largest one represents machine learning, while the remaining four represent explainable AI, natural language processing, large language models, and generative AI. These fruits symbolize different technological branches in the field of AI, showcasing the diversity and evolutionary process of AI technology. Note: These branches are interrelated but each has unique characteristics and application areas.
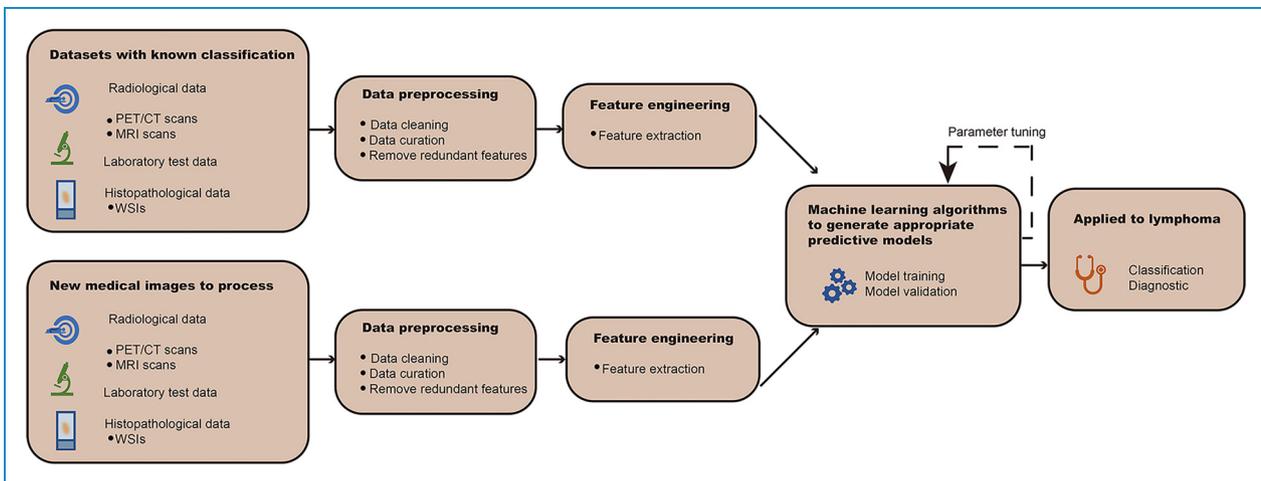

Figure 3. The general workflow of supervised machine learning in lymphoma. Supervised machine learning (ML) generally encompasses two steps, starting with a set of image datasets with known classifications and corresponding class labels, generating a training prediction model through steps such as data cleaning and feature extraction, followed by an application phase. This machine learning model is applied to newly acquired medical images and, eventually, is applied to the classification and diagnosis of lymphoma.

Table 1. Summary of studies utilizing machine learning applied to lymphoma.

Reference	Year	Task	Output	Input	Model
Steinbuss et al.[14]	2021	Diagnosis	Classification of non-Hodgkin lymphoma	Histopathological images	Convolutional neural network
Tagami et al.[18]	2023	Differential diagnosis	Distinguish between orbital and conjunctival mucosa-associated lymphoid tissue lymphomas	990 pathological images of 99 patients	Support vector machine
Can et al.[21]	2024	Diagnosis	Diagnosis of cervical lymph node enlargement	400 patients who underwent surgery for cervical lymph node enlargement between 2010 and 2022	Deep learning
Zhuang et al.[31]	2023	Validation of a prognostic model	Established a predictive model including seven key genes	The integrated genome of 1143 DLBCL samples	Support vector machine
Gross et al.[39]	2024	Differential diagnosis	Distinguish between tumor and nontumor B cells, and propose the correct classification of chronic lymphoproliferative diseases in over 90% of cases	Cases of chronic lymphoproliferative diseases with over 300 annotations	Random forest
Sun et al.[61]	2021	Validation of a prognostic model	Prognostic index of natural killer lymphoma	Clinical data from 250 extranodal natural killer/t-cell lymphoma patients treated with nonanthracyclines	Machine learning model
Zhao et al.[63]	2024	Validation of a prognostic model	Predicting the prognosis of DLBCL after combination chemotherapy based on the proportion of IgH detected in ctDNA samples tracked from organizational samples	55 DLBCL sequencing data cases	Decision Tree

DLBCL: diffuse large B-cell lymphoma.

materials, and efficiently extract relevant and meaningful features.[19] For example, the symbol-based ML method applied to the pixel classification of FL images shows that the classification ability is successful at least in the image preprocessing stage, and the accuracy of classification results is as high as 95%.[20] An ML model was developed in a recent study to facilitate pathological diagnosis for 400 patients who underwent surgery due to cervical lymph node enlargement, eliminating the necessity for radiologist interpretation and invasive examination.[21]

Besides, digital pathology is an interdisciplinary attempt, that is, using the whole slide image to train the deep neural network has become an essential tool for lymphoma diagnosis.[22–25] There are literature reports that the depth residual neural network model is used to preprocess the pathological images of lymphoma, and the data enhancement methods such as turnover and color transformation are used. The classification accuracy is 98.63%. The network model could offer objective evidence for clinicians

to diagnose the type of NHL.[15] In the research of Alferez et al., they trained a recognition system with an SVM algorithm. The system is an automatic analysis method based on an image. A group of images of 21 new patients were analyzed and parallel with the confirmed diagnosis. When the cells are divided into normal lymphocytes, abnormal lymphocytes and reactive lymphocytes, the overall recognition accuracy is 97.67%.[26]

ML and radiomics, genomics

The combination of radiometry and ML accelerates the development of histopathological images of lymphoma.[27–30] The CNN which is a promising tool can classify a large amount of B-cell lymphoma information obtained by multiparameter flow cytometry (MFC). However, we still have a long way to go in optimizing CNN's performance in analyzing images with different magnifications, different image attributes, and network structures.

In addition to digitizing histopathological sections and analyzing medical images, ML tools are also used to help classify lymphoma by analyzing genomic data. Using ML, DLBCL subtypes are differentiated based on mRNA features abundant in immune infiltration and cell cycle.[31] Gene expression profile can divide DLBCL into germinal center B-cell-like, active B-cell-like, and unclassified subtypes. Different subtypes have different prognoses. The ML method also can detect eight specific markers such as BCL 6, MYBL1, LMO2, MME, NFKBIZ, IRF4, PDE4B, and SLA. Hierarchical management of patients will contribute to the personalized medical choice of DLBCL patients.[32] One study explored targeted RNA expression data generated by sequencing 1408 genes in DLBCL using a Bayesian approach in ML.[33] However, the application of MFC based on ML is still limited. Comprehensive analysis methods such as whole-genome sequencing and whole-transcriptome sequencing have great potential in the diagnosis of hematological malignancies, but most methods have not been explored. Because of its high dimensionality and multimodality, the genome dataset is an ideal choice for DL exploration. It can be used to improve diagnosis, further ameliorate classification systems and determine prognostic factors. However, the number of genes to be detected and interpreted is so large that this prognostic classification has not been used in clinical practice.

ML and laboratory examination

Accurate identification of cell morphology is the basis of the preliminary diagnosis of lymphoma. The combination of peripheral blood, bone marrow analysis, and ML algorithm gives the accurate diagnosis of lymphoma hit a spark of hope. Tang et al. developed an ML system Morphogo, which is based on an algorithm and uses digital imaging analysis to distinguish abnormal lymphocytes from normal lymphocytes.[34] ML algorithm can further classify bone marrow leukocytes by training CNN.[35] MFC is a mighty high-throughput technology, which could quickly quantify the markers on suspended cells. It is the cornerstone of lymphoma clinical decision-making. However, manual control of cell populations is time-consuming and not objective. In recent years, with the in-depth study of DL algorithms, we could accurately classify disease subtypes by identifying cell types in cytological images, and automatically classifying MFC data as diagnostic labels.[36] On this basis, Mallesh et al. further classify seven B-cell tumor subtypes and nine color panel healthy controls by using the method of migration learning. A workflow that extends the DL model to MFC panels is proposed to realize high-precision multi-tab classification between datasets. Therefore, some challenges of automatic flow cytometry classification are solved. This work makes the DL model a big step forward.[37]

Flow cytometry, traditionally reliant on a blood pathologist with specialized expertise to meticulously analyze digital data, serves as a pivotal clinical instrument for diagnosing numerous malignant tumors within the blood system. The advancements in ML hold the potential to be integrated into flow cytometry, potentially enhancing the efficiency and prioritization of cases while mitigating errors. Although flow cytometry has many advantages, its application is usually limited to the analysis of bone marrow or peripheral blood lymphoma. To solve this problem, Gaidano et al. combined ML tools with flow cytometry to generate multiple clinically applicable prediction systems, which could be utilized for each biomaterial involved in lymphoproliferative disorders, from blood to lymph nodes or pleural effusion.[38] ML-assisted flow cytometry can distinguish between tumor and nontumor B cells in all cases, and propose the correct classification of chronic lymphoproliferative diseases in over 90% of cases.[39] An ML logistic regression model was utilized by Zijtregtop et al. to create a diagnostic scoring model for pediatric lymphoma. This model facilitates the prompt referral of suspected lymphoma patients to oncologists and has the potential to decrease the number of referrals and unnecessary invasive procedures for patients with benign lymph node enlargement.[40]

ML and prognosis of lymphoma

Prognosis is a well-known and arduous task. Among the widely used prognostic tools, there are also significant differences in risk levels.[41] This can be done in advance throughout the treatment process. If found in time, many major diseases can be improved. On the contrary, the deferment in the prognosis inference could provoke the onset of severe problems. The prediction model is established with the help of patient history, current status, doctor records, and genetic research details.[42] The prediction model is not only a model but also a mathematical function. It considers at least two predictors to simultaneously predict the probability of an event in a single patient within a specific future time range.[43–45] Therefore, predicting the risk of disease or treatment will help to develop better prevention plans.

Developing a prognostic model of lymphoma could predict the disease progression, the related risks, and the impact of treatment on patients, and help doctors make more accurate treatment decisions. Many indicators can predict the prognosis of lymphoma, such as total metabolic tumor volume (TMTV) which evaluates the metabolically active volume of a tumor by measuring the body FDG-PET/CT. As an independent prognostic factor, TMTV shows strong prognostic value in both HL and NHL.[46] The prognosis of patients with DLBCL was evaluated according to the degree of lesion diffusion of baseline PET/CT features.[47] Nevertheless, in mantle cell lymphoma,

although TMTV is an essential prognostic instrument to improve risk stratification in untreated mantle cell lymphoma patients, the International Prognostic Index for mantle cell lymphoma and KI-67 remain the strongest predictors of outcome.[48] Capobianco et al. demonstrated that the DL method could fully automatically estimate TMTV, and the results are consistent with the expert measurements of a large number of DLBCL subjects. Significant prognostic value of progression-free survival and overall survival was shown in subjects.[49] Biccler et al. integrated clinical data to establish a new prognostic model based on ML techniques, which was validated to be remarkably better than existing DLBCL prognostic indexes, such as the International Prognostic Index.[50]

Using ML, texture analysis of central nervous system lymphoma and other neoplasms can be carried out in advanced magnetic resonance imaging.[51,52] ML is also applied to structured electronic health record (EHR) to forecast individual survival at the beginning of first-line treatment.[53] FL needs to observe the progress of 24 months two years after the first-line treatment, which is the most precise prognostic method of FL, ML provides a tool based on EHR with more accurate prediction ability.[54] Carreras et al. used ML to analyze the gene expression of FL and identified the overall survival genes related to patient prognosis and microenvironment.[55] Research by Irshaid et al. indicates that ML can solve the problem of large cell transformation probability for patients diagnosed with FL and undergoing bone marrow biopsy.[56]

AI can stratify the risk of hematopoietic stem cells in patients with aGVHD, which has an impact on the immunosuppressive decision-making of these individuals.[57] Supervised learning prediction method can identify genes related to the prognosis of DLBCL, including some genes regulating B cell receptor signal transduction, key serine/threonine phosphorylation pathway, and apoptosis, to predict the prognosis of DLBCL and determine reasonable intervention targets.[58] Recently, researchers used AI to analyze a group of DLBCL. They used the segmentation method based on AI to quantify the immunohistochemical protein expression of tnfaip8 and confirmed it by the traditional figure quantification method. It was authenticated that the high protein expression of tnfaip8 was related to the low overall survival rate. This also proves that the single gene set of neural network analysis can be used for forecasting the prognosis of hematological neoplasms, including lymphoma.[59,60]

In addition, stem cell characteristics and tumor microenvironment defined by logistic regression, ML, and AI can also be used to predict the prognosis of DLBCL. The statistical pattern based on the Cox proportional hazard model can be used as the prognostic model for the optimal treatment of extranodal natural killer T-cell lymphoma nasal-type nonanthracycline drugs.[61] Surface-enhanced Raman spectroscopy (SERS) has become a latent medical tool. It

is reported that an ML algorithm is used to establish the prognostic model of the disease burden score of standard treatment patients. It has been proved that SERS can be used as a noninvasive and label-free tool for monitoring hematological malignancies.[62] By monitoring the rearrangement of immunoglobulin genes in circulating tumor DNA, the decision tree model forecasts the progression of DLBCL.[63] Many patients can now obtain a good prognosis through standard chemotherapy, but 30–50% of patients with DLBCL may recrudescence after remedy. It is reported that the DLBCL risk model based on the traditional machine model shows good differentiation and calibration ability, and may help clinicians make the best treatment decision to achieve accurate medicine.[64]

ML and treatment selection of lymphoma

Prediction of chemotherapy efficacy

ML algorithms are increasingly used to improve patients' treatment decisions and increase patients' treatment choices and we have been trying to develop targeted drugs for lymphoma. Clinically, for lymphoma patients, how to choose the best treatment scheme and how to minimize the side effects of chemotherapy and radiotherapy are the problems we need to solve. The endless emergence of targeted drugs depends on the progress of genomics and AI in recent years. More and more evidence shows that patients benefit from this paradigm shift, which is usually called precision medicine or personalized medicine.[65,66] Rosenberg and others used the NGS technology and electronic technology to design a research plan, which opened up a new field in evaluating the efficacy, safety, pharmacokinetics, and pharmacodynamics of drugs. It may also accelerate drug development and accelerate our goal of improving clinical outcomes.[67,68] PET/CT imaging has a high sensitivity rate for primary and metastatic lymphoma.[69,70] It can assess the lesion and treatment effect and guide the follow-up remedy of patients. It is reported that the probability-corrected ensemble method can be used to predict the mortality of patients with DLBCL, which assists us in judging the treatment effect and further improving the chemotherapy regimen.[71] AI affects lymphoma by influencing radiology. In addition to being used for diagnosis and anatomical measurement, it can also make more effective risk stratification for patients with HL, to affect treatment decision-making.[69] In lymphoma, the combination of NGS and AI is used to predict the response after chemotherapy.[72,73]

Drug discovery and response prediction

ML algorithms have been widely used in computer-aided drug discovery.[74,75] The probability of drug failure in clinical trials is high because pharmaceutical companies cannot

use traditional computing technology to predict the impact of latent drug compounds on targets. Under these circumstances, applying the ML prediction method could save a lot of resources. Recently developed AI methods, such as ML and related modeling research, provide a new solution for the safety evaluation of candidate drugs based on big data modeling and analysis.[76]

ML is used to identify toxic compounds that may cause side effects, which can save a large number of resources before drugs enter clinical trials. Adopting ML to automatically synthesize compounds, the drug development process can be improved and the time to market of drugs can be shortened. How to discover more effective drugs and how to predict the effects of these drugs on the body are the matters that researchers use ML tools to solve. Diverse algorithms and models reduce the laboratory workload involved in drug discovery. Datasets analysis enables researchers to analyze the results of various chemical compositions.[77] In a recent study, a universal pipeline has been devised which integrates ML with molecular dynamics simulation to uncover potential dual inhibitors of B-cell lymphoma from natural product databases.[78]

One prospect for in-depth study in the biomedical field is to predict therapeutic responses and conduct the development of new therapies. DL gives priority to new chemical targets in the early phase of drug discovery.[79] In addition, Julkunen et al. proposed ComboFM, an ML framework for predicting drug combination response in preclinical research, such as combined medication based on a cell line or patient-originated cells.[80] It is an exciting research field to use DL neural networks to clarify the pharmacological features of a variety of drugs under different biological systems and conditions on transcriptome data and realize drug repositioning.[81]

Current limitations and opportunities for the clinical implementation of ML methods

Data sources and data quality

The medical industry has remarkable "big data" characteristics. First, the amount of data is large, and second, the data types are confusing and complex. In the process of data collection, due to the lack of unified standards in different medical institutions, the data are scattered, nonstandard, incomplete, or missing, or it is difficult to directly collect enough electronic medical record information, resulting in the low-quality and unreliability of the collected data, and it is difficult to extract meaningful information from the data to predict and model. It is necessary to ensure the integrity and authenticity of the data as much as possible. The collected data still needs to be verified, preferably externally. For a modest sample size, virtually any ML model is predisposed to overfitting, which can lead to artificially inflated algorithmic accuracy and even to clinical

implications for decision-making. Feature selection and regularization have shown some effectiveness in preventing overfitting.[82] However, only a few studies usually verify the model externally for the sake of proving the precision and practicability of the model. Table 2 lists some applications and challenges of AI in lymphoma patient management.

Data security and counterattack prevention

Healthcare data has a critical role to play in the advancement of medicine. The issue of data security has been persistent and topical attention. Recent issues such as data breaches also illustrate the crucial need to ensure the privacy and security of healthcare data.[83] Recently Shakir Khan et al. explored an information entropy l-diversity model that minimizes implementation time and lessens the chance of critical data leakage while improving data accuracy and quality of service, all of which aids in the security of healthcare data interchange.[84] This methodology may potentially assist in alleviating future data-sharing concerns.

Clinical integration

In the realm of lymphoma clinical practice, the integration of ML heralds both novel opportunities and formidable challenges in diagnostic and therapeutic domains. LLM technology offers a sophisticated means to scrutinize patient imaging datasets, enabling physicians to swiftly and precisely discern lymphoma lesions, thereby steering the trajectory of treatment strategies. Concurrently, the incorporation of NLP augments physicians' capacity to assimilate intricate patient medical histories and clinical records, facilitating nuanced evaluations of disease progression and treatment efficacy. In navigating the complexities of treatment selection, XAI technology serves as an invaluable ally, furnishing clinicians with decision support mechanisms and elucidating the predictive outcomes of ML models,[85] thereby instilling a heightened sense of confidence in treatment modalities. In essence, a burgeoning wave of cutting-edge ML technologies is revolutionizing medical practice. By harnessing the full spectrum of these advancements, bespoke treatment paradigms can be meticulously crafted in clinical decision-making, fostering heightened precision and efficacy in treatment outcomes, while concurrently elevating the standards of clinical management and medical care afforded to lymphoma patients.

Conclusions

In summary, the application of ML in lymphoma management has demonstrated remarkable potential across various facets of clinical practice. In diagnosis, ML techniques have shown promise in analyzing imaging data and identifying lymphoma lesions with speed and accuracy,

Table 2. Applications and challenges of machine learning in the management of lymphoma.

	Immunophenotyping	*Cytomorphology*	*Molecular genetics*	*Radiology*
Technique	Multiparameter flow cytometry	Microscopy	Genomic analysis	^{18}F-FDG PET imaging
Objective	Minimal residual disease evaluation/identification and characterization of cell populations	Identification and characterization of proteins based on morphology	Identification of molecular map	For lymphoma staging, treatment response assessment, and relapse detection
Improvements needed	How to accurately represent the original data	How to correctly identify boundaries	The data matrix is sparse and the signal is easy to be lost	How to distinguish recurrence from inflammation/infection or active fibrosis after chemotherapy or radiotherapy
Challenges	• The number of available training samples is limited, and it is impossible to establish the depth model without overfitting.			
	• It poses potential security and data ownership challenges when applied to electronic medical record systems.			
	• The interpretability and acceptability of the constructed model are poor.			
	• Because the collected data is of low quality and unreliable, it is difficult to extract meaningful information from the data for prediction and modeling.			

thus facilitating timely intervention. Moreover, in prognosis assessment, ML models have proven valuable in predicting disease progression and treatment outcomes, aiding clinicians in formulating tailored management plans for individual patients. Personalized therapeutic options have been made possible in the field of therapy by ML-driven approaches, which maximize benefits while avoiding negative effects. It is crucial to recognize the inherent limits when incorporating ML into clinical practice. Robust precautions are required for safeguarding sensitive medical information since data security concerns, including patient privacy and confidentiality, are still of utmost importance. Furthermore, difficulties with ML models' interpretability and the moral implications of using them for policy-making emphasize how crucial it is to follow ethical standards and have open decision-making processes. Moving forward, concerted efforts are required to address these limitations and capitalize on the opportunities afforded by ML in lymphoma management. Future initiatives should prioritize the development of standardized protocols for data security and privacy protection, alongside the implementation of interpretability frameworks to enhance the transparency of ML algorithms. Furthermore, interdisciplinary collaboration and stakeholder engagement are essential to navigate ethical dilemmas and ensure the responsible integration of ML technologies into clinical workflows. Going forward, there is great potential for improving precision medicine in the treatment of lymphomas as a result of the convergence of ML with developments in genomics, imaging,

and therapeutic modalities. We can fully utilize ML to transform the treatment of lymphoma and enhance patient outcomes in the future by adopting a comprehensive strategy that takes into account technological, ethical, and societal factors.

Informed consent: As a review, the manuscript does not require informed consent.

ORCID iD: Xin Wang https://orcid.org/0000-0001-8051-1481

References

1. Liu W, Ji X, Song Y, et al. Improving survival of 3760 patients with lymphoma: experience of an academic center over two decades. *Cancer Med* 2020; 9: 3765–3774.
2. Choi SM, Malley O and P D. Diagnostically relevant updates to the 2017 WHO classification of lymphoid neoplasms. *Ann Diagn Pathol* 2018; 37: 67–74.
3. Ching T, Himmelstein DS, Beaulieu-Jones BK, et al. Opportunities and obstacles for deep learning in biology and medicine. *J R Soc Interface* 2018; 15: 20170387.
4. Halper-Stromberg E, McCall CM, Haley LM, et al. Cloneretriever: an automated algorithm to identify clonal B and T cell gene rearrangements by next-generation sequencing for the diagnosis of lymphoid malignancies. *Clin Chem* 2021; 67: 1524–1533.
5. Thomsen K, Christensen AL, Iversen L, et al. Deep learning for diagnostic binary classification of multiple-lesion skin diseases. *Front Med (Lausanne)* 2020; 7: 574329.
6. Radakovich N, Nagy M and Nazha A. Machine learning in haematological malignancies. *Lancet Haematol* 2020; 7: e541–ee50.
7. Chan S, Reddy V, Myers B, et al. Machine learning in dermatology: current applications, opportunities, and limitations. *Dermatol Ther (Heidelb)* 2020; 10: 365–386.
8. Miyoshi H, Sato K, Kabeya Y, et al. Deep learning shows the capability of high-level computer-aided diagnosis in malignant lymphoma. *Lab Invest* 2020; 100: 1300–1310.
9. Fan S, Zhao Z, Zhang Y, et al. Probability calibration-based prediction of recurrence rate in patients with diffuse large B-cell lymphoma. *BioData Min* 2021; 14: 38.
10. Greenspan H, van Ginneken B and Summers RM. Deep learning in medical imaging: overview and future promise of an exciting new technique. *IEEE Trans Med Imaging* 2016; 35: 1153–1159.
11. Shen DG, Wu GR and Suk HI. Deep learning in medical image analysis. In: Yarmush ML (eds) *Annual review of biomedical engineering. 19.* Palo Alto: Annual Reviews, 2017, pp.221–248.
12. Beam AL and Kohane IS. Big data and machine learning in health care. *JAMA-J Am Med Assoc* 2018; 319: 1317–1318.
13. Moran-Sanchez J, Santisteban-Espejo A, Martin-Piedra MA, et al. Translational applications of artificial intelligence and machine learning for diagnostic pathology in lymphoid neoplasms: a comprehensive and evolutive analysis. *Biomolecules* 2021; 11: 24.
14. Steinbuss G, Kriegsmann M, Zgorzelski C, et al. Deep learning for the classification of non-Hodgkin lymphoma on histopathological images. *Cancers (Basel)* 2021; 13: 2419.
15. Zhang XL, Zhang KX, Jiang M, et al. Research on the classification of lymphoma pathological images based on deep residual neural network. *Technol Health Care* 2021; 29: 335–344.
16. Adeyi OA. Pathology services in developing countries-the west African experience. *Arch Pathol Lab Med* 2011; 135: 183–186.
17. Adesina A, Chumba D, Nelson AM, et al. Improvement of pathology in sub-Saharan Africa. *Lancet Oncol* 2013; 14: E152–E1E7.
18. Tagami M, Nishio M, Katsuyama-Yoshikawa A, et al. Machine learning model with texture analysis for automatic classification of histopathological images of ocular adnexal mucosa-associated lymphoid tissue lymphoma of two different origins. *Curr Eye Res* 2023; 48: 1195–1202.
19. Ferjaoui R, Cherni MA, Boujnah S, et al. Machine learning for evolutive lymphoma and residual masses recognition in whole body diffusion weighted magnetic resonance images. *Comput Meth Programs Biomed* 2021; 209: 18.
20. Zorman M, de la Rosa JLS and Dinevski D. Classification of follicular lymphoma images: a holistic approach with symbol-based machine learning methods. *Wien Klin Wochen* 2011; 123: 700–709.
21. Can S, Türk Ö, Ayral M, et al. Can deep learning replace histopathological examinations in the differential diagnosis of cervical lymphadenopathy? *Eur Arch Otorhinolaryngol* 2024; 281: 359–367.
22. Dimitriou N, Arandjelovic O and Caie PD. Deep learning for whole slide image analysis: an overview. *Front Med* 2019; 6: 7.
23. El Achi H and Khoury JD. Artificial intelligence and digital microscopy applications in diagnostic hematopathology. *Cancers (Basel)* 2020; 12: 15.
24. Zhang JF, Cui WS, Guo XY, et al. Classification of digital pathological images of non-Hodgkin's lymphoma subtypes based on the fusion of transfer learning and principal component analysis. *Med Phys* 2020; 47: 4241–4253.
25. El Achi H, Belousova T, Chen L, et al. Automated diagnosis of lymphoma with digital pathology images using deep learning. *Ann Clin Lab Sci* 2019; 49: 153–160.
26. Alferez S, Merino A, Bigorra L, et al. Characterization and automatic screening of reactive and abnormal neoplastic B lymphoid cells from peripheral blood. *Int J Lab Hematol* 2016; 38: 209–219.
27. Bae S, Choi YS, Sohn B, et al. Squamous cell carcinoma and lymphoma of the oropharynx: differentiation using a radiomics approach. *Yonsei Med J* 2020; 61: 895–900.
28. Bathla G, Priya S, Liu YA, et al. Radiomics-based differentiation between glioblastoma and primary central nervous system lymphoma: a comparison of diagnostic performance across different MRI sequences and machine learning techniques. *Eur Radiol* 2021; 31: 8703–8713.
29. Pennig L, Hoyer UCI, Goertz L, et al. Primary central nervous system lymphoma: clinical evaluation of automated segmentation on multiparametricMRIUsing deep learning. *J Magn Reson Imaging* 2021; 53: 259–268.
30. Santiago R, Jimenez JO, Forghani R, et al. Prediction of high-risk group of primary refractory diffuse large B-cell lymphoma (DLBCL) patients using a CT-based radiomics model with machine learning. *Blood* 2019; 134: 4136–4136.
31. Zhuang X, Liu B, Long J, et al. Machine-learning-based classification of diffuse large B-cell lymphoma patients by a 7-mRNA signature enriched with immune infiltration and cell cycle. *Clin Transl Oncol* 2023; 26: 936–950.

32. Zhao ST, Dong XL, Shen WZ, et al. Machine learning-based classification of diffuse large B-cell lymphoma patients by eight gene expression profiles. *Cancer Med* 2016; 5: 837–852.

33. Albitar M, Zhang H, Goy A, et al. Determining clinical course of diffuse large B-cell lymphoma using targeted transcriptome and machine learning algorithms. *Blood Cancer J* 2022; 12: 25.

34. Tang GS, Fu XY, Wang Z, et al. A machine learning tool using digital microscopy (morphogo) for the identification of abnormal lymphocytes in the bone marrow. *Acta Cytol* 2021; 65: 354–357.

35. Matek C, Krappe S, Munzenmayer C, et al. Highly accurate differentiation of bone marrow cell morphologies using deep neural networks on a large image data set. *Blood* 2021; 138: 1917–1927.

36. Matek C, Schwarz S, Spiekermann K, et al. Human-level recognition of blast cells in acute myeloid leukaemia with convolutional neural networks. *Nat Mach Intell* 2019; 1: 538–544.

37. Mallesh N, Zhao M, Meintker L, et al. Knowledge transfer to enhance the performance of deep learning models for automated classification of B-cell neoplasms. *Oncol Res Treat* 2021; 44: 96.

38. Gaidano V, Tenace V, Santoro N, et al. A clinically applicable approach to the classification of B-cell non-Hodgkin lymphomas with flow cytometry and machine learning. *Cancers (Basel)* 2020; 12: 18.

39. Gross Z, Veyrat-Masson R, Grange B, et al. Diagnosis of chronic B-cell lymphoproliferative disease in peripheral blood = how machine learning may help to the interpretation of flow cytometry data. *Hematol Oncol* 2024; 42: e3245.

40. Zijtregtop EAM, Winterswijk LA, Beishuizen TPA, et al. Machine learning logistic regression model for early decision making in referral of children with cervical lymphadenopathy suspected of lymphoma. *Cancers (Basel)* 2023; 15: 1178.

41. Patel SS, Sekeres MA and Nazha A. Prognostic models in predicting outcomes in myelodysplastic syndromes after hypomethylating agent failure. *Leuk Lymphoma* 2017; 58: 2532–2539.

42. Harkins RA, Chang A, Patel SP, et al. Remaining challenges in predicting patient outcomes for diffuse large B-cell lymphoma. *Expert Rev Hematol* 2019; 12: 959–973.

43. Moons KGM, Royston P, Vergouwe Y, et al. Prognosis and prognostic research: what, why, and how? *BMJ-British Medical Journal* 2009; 338: b375–b375.

44. Steyerberg EW, Moons KG, van der Windt DA, et al. Prognosis research strategy (PROGRESS) 3: prognostic model research. *PLoS Med* 2013; 10: e1001381.

45. Alba AC, Agoritsas T, Walsh M, et al. Discrimination and calibration of clinical prediction models users' guides to the medical literature. *JAMA-J Am Med Assoc* 2017; 318: 1377–1384.

46. Blanc-Durand P, Jegou S, Kanoun S, et al. Fully automatic segmentation of diffuse large B cell lymphoma lesions on 3D FDG-PET/CT for total metabolic tumour volume prediction using a convolutional neural network. *Eur J Nucl Med Mol Imaging* 2021; 48: 1362–1370.

47. Cottereau AS, Meignan M, Nioche C, et al. New approaches in characterization of lesions dissemination in DLBCL patients on baseline PET/CT. *Cancers (Basel)* 2021; 13: 3998.

48. Vergote VKJ, Verhoef G, Janssens A, et al. [(18)F] FDG-PET/CT volumetric parameters can predict outcome in untreated mantle cell lymphoma. *Leuk Lymphoma* 2023; 64: 161–170.

49. Capobianco N, Meignan M, Cottereau AS, et al. Deep-Learning F-18-FDG uptake classification enables total metabolic tumor volume estimation in diffuse large B-cell lymphoma. *J Nucl Med* 2021; 62: 30–36.

50. Biccler JL, Eloranta S, de Nully Brown P, et al. Optimizing outcome prediction in diffuse large B-cell lymphoma by use of machine learning and nationwide lymphoma registries: a Nordic lymphoma group study. *JCO Clin Cancer Inform* 2018; 2: 1–13.

51. Lippi M, Gianotti S, Fama A, et al. Texture analysis and multiple-instance learning for the classification of malignant lymphomas. *Comput Meth Programs Biomed* 2020; 185: 105153.

52. Xia W, Hu B, Li HQ, et al. Multiparametric-MRI-based radiomics model for differentiating primary central nervous system lymphoma from glioblastoma: development and cross-vendor validation. *J Magn Reson Imaging* 2021; 53: 242–250.

53. Liang XJ, Yao N, Liu D, et al. Stemness signature defined by logistic regression machine learning and artificial intelligence predicts outcomes in diffuse large B-cell lymphoma. *Blood* 2020; 136: 19–20.

54. Li CY, Patil V, Rasmussen KM, et al. Predicting survival in veterans with follicular lymphoma using structured electronic health record information and machine learning. *Int J Environ Res Public Health* 2021; 18: 19.

55. Carreras J, Kikuti YY, Miyaoka M, et al. Artificial intelligence analysis of the gene expression of follicular lymphoma predicted the overall survival and correlated with the immune microenvironment response signatures. *Machine Learning and Knowledge Extraction* 2020; 2: 647–671.

56. Irshaid L, Bleiberg J, Weinberger E, et al. Histopathologic and machine deep learning criteria to predict lymphoma transformation in bone marrow biopsies. *Arch Pathol Lab Med* 2022; 146: 182–193.

57. Arai Y, Kondo T, Fuse K, et al. Using a machine learning algorithm to predict acute graft-versus-host disease following allogeneic transplantation. *Blood Adv* 2019; 3: 3626–3634.

58. Shipp MA, Ross KN, Tamayo P, et al. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nat Med* 2002; 8: 68–74.

59. Carreras J, Kikuti YY, Miyaoka M, et al. A single gene expression set derived from artificial intelligence predicted the prognosis of several lymphoma subtypes; and high immunohistochemical expression of TNFAIP8 associated with poor prognosis in diffuse large B-cell lymphoma. *Ai* 2020; 1: 342–360.

60. Carreras J, Kikuti YY, Miyaoka M, et al. A combination of multilayer perceptron, radial basis function artificial neural networks and machine learning image segmentation for the dimension reduction and the prognosis assessment of diffuse large B-cell lymphoma. *Ai* 2021; 2: 106–134.

61. Sun JB, Ke XY, Zhang MZ, et al. New prognostic models for extranodal natural killer T-cell lymphoma, nasal-type using Cox regression and machine learning. *Transl Cancer Res* 2021; 10: 613–626.

62. Grieve S, Puvvada N, Phinyomark A, et al. Nanoparticle surface-enhanced Raman spectroscopy as a noninvasive, label-free tool to monitor hematological malignancy. *Nanomedicine* 2021; 16: 2175–2188.

63. Zhao K, Zheng X, Liu X, et al. Tissue-matched IgH gene rearrangement of circulating tumor DNA shows significant value in predicting the progression of diffuse large B cell lymphoma. *Oncologist* 2024; XX: 1–9.

64. Fan SL, Zhao ZQ, Zhang YB, et al. Probability calibration-based prediction of recurrence rate in patients with diffuse large B-cell lymphoma. *BioData Min* 2021; 14: 18.

65. Briguglio W, Moghaddam P, Yousef WA, et al. Machine learning in precision medicine to preserve privacy via encryption. *Pattern Recognit Lett* 2021; 151: 148–154.

66. Eraslan G, Avsec Z, Gagneur J, et al. Deep learning: new computational modelling techniques for genomics. *Nat Rev Genet* 2019; 20: 389–403.

67. Rosenberg L, Levaux H, Levine RL, et al. Streamlined operational approaches and use of e-technologies in clinical trials: beat acute myeloid leukemia master trial. *Ther Innov Regul Sci* 2021; 55: 926–935.

68. Cascione L, Rinaldi A, Bruscaggin A, et al. Novel insights into the genetics and epigenetics of MALT lymphoma unveiled by next generation sequencing analyses. *Haematologica* 2019; 104: E558–EE61.

69. Milgrom SA, Elhalawani H, Lee J, et al. A PET radiomics model to predict refractory mediastinal Hodgkin lymphoma. *Sci Rep* 2019; 9: 1322.

70. Yuan C, Zhang M, Huang XY, et al. Diffuse large B-cell lymphoma segmentation in PET-CT images via hybrid learning for feature fusion. *Med Phys* 2021; 48: 3665–3678.

71. Fan SL, Zhao ZQ, Yu HM, et al. Applying probability calibration to ensemble methods to predict 2-year mortality in patients with DLBCL. *BMC Med Inform Decis Mak* 2021; 21: 12.

72. Chang Y, Park H, Yang HJ, et al. Cancer drug response profile scan (CDRscan): a deep learning model that predicts drug effectiveness from cancer genomic signature. *Sci Rep* 2018; 8: 11.

73. Mani-Varnosfaderani A, Neiband MS and Benvidi A. Identification of molecular features necessary for selective inhibition of B cell lymphoma proteins using machine learning techniques. *Mol Diversity* 2019; 23: 55–73.

74. Vamathevan J, Clark D, Czodrowski P, et al. Applications of machine learning in drug discovery and development. *Nat Rev Drug Discov* 2019; 18: 463–477.

75. Boniolo F, Dorigatti E, Ohnmacht AJ, et al. Artificial intelligence in early drug discovery enabling precision medicine. *Expert Opin Drug Discov* 2021; 16: 991–1007.

76. Zhu H. Big data and artificial intelligence modeling for drug discovery. In: Insel PA (eds) *Annual review of pharmacology and toxicology, vol 60*. Palo Alto: Annual Reviews, 2020, pp.573–589.

77. Nayarisseri A, Khandelwal R, Tanwar P, et al. Artificial intelligence, big data and machine learning approaches in precision medicine & drug discovery. *Curr Drug Targets* 2021; 22: 631–655.

78. Liu L, Na R, Yang L, et al. A workflow combining machine learning with molecular simulations uncovers potential dual-target inhibitors against BTK and JAK3. *Molecules* 2023; 28: 7140.

79. Baskin II, Winkler D and Tetko IV. A renaissance of neural networks in drug discovery. *Expert Opin Drug Discov* 2016; 11: 785–795.

80. Julkunen H, Cichonska A, Gautam P, et al. Leveraging multiway interactions for systematic prediction of pre-clinical drug combination effects. *Nat Commun* 2020; 11: 11.

81. Aliper A, Plis S, Artemov A, et al. Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data. *Mol Pharm* 2016; 13: 2524–2530.

82. Hawkins DM. The problem of overfitting. *J Chem Inf Comput Sci* 2004; 44: 1–12.

83. Jiang D and Shi G. Research on data security and privacy protection of wearable equipment in healthcare. *J Healthc Eng* 2021; 2021: 1–7.

84. Khan S, Saravanan V, N GC, et al. Privacy protection of healthcare data over social networks using machine learning algorithms. *Comput Intell Neurosci* 2022; 2022: 1–8.

85. Rodríguez-Belenguer P, Piñana JL, Sánchez-Montañés M, et al. A machine learning approach to identify groups of patients with hematological malignant disorders. *Comput Methods Programs Biomed* 2024; 246: 108011.