Journal of Cancer Research and Clinical Oncology (2025) 151:268
https://doi.org/10.1007/s00432-025-06314-7

REVIEW

Prediction models for different types of leukemia: a systematic review and critical appraisal

Yingzi Yang[1,2] · Ayizhati Tuerxun[1,2] · Xinqi Cai[1,2] · Xinyu Chen[1,2] · Zhuoya Zhao[1,2] · Yang Zhao[1,2] · Zinuo Lin[1,2] · Shengfeng Wang[1,2]

Received: 13 August 2025 / Accepted: 4 September 2025

Abstract

Objectives To systematically review and evaluate the methodological quality and risk of bias (ROB) of leukemia prediction models essential for clinical decision-making.

Methods We reviewed 148 prediction models published before August 2023 from PubMed, Embase, Cochrane Library, and Web of science databases. Two reviewers independently screened articles and extracted data using CHARMS criteria. ROB was assessed using PROBAST. Models were categorized by leukemia subtype and analyzed for methodological characteristics.

Results A total of 61 acute myeloid leukemia (AML) models primarily predicted survival (82.0%), diagnosis (4.9%), or death (4.9%) using predictors including age, cytogenetic risk, and white blood cell count. Among the 22 chronic myeloid leukemia (CML) models, the focus was on survival (72.7%) and time to treatment (19.0%), utilizing blast percentage, age, and platelet count. A total of 21 chronic lymphocytic leukemia (CLL) models primarily predicted survival (71.4%) using IGHV status, Rai stage, and age. The methodological shortcomings including incomplete reporting, methodological limitations, and high ROB were consistent across different leukemia subtypes. Traditional statistical methods predominated (Cox regression 72.9%, logistic regression 12.2%), with only nine machine learning models. Critical methodological limitations included lack of internal validation (52.0%) and external validation (57.4%). Only 43.2% reported discrimination metrics (AUC 0.60–0.99), with 28.0% achieving AUC > 0.7. Calibration was reported in only 23.0% of models. High ROB affected 93.9% of studies, primarily due to inadequate data handling and validation.

Conclusions Existing leukemia prediction models have limited clinical utility due to methodological shortcomings and high ROB. Future research should prioritize transparent reporting, rigorous validation, and external validation to enhance clinical applicability and generalizability.

Keywords Leukemia · Prediction model · Systematic review · Risk of bias

Abbreviations

ROB	Risk of bias
CHARMS	Checklist for critical appraisal and data extraction for systematic reviews of prediction modelling studies
PROBAST	Prediction model risk of bias assessment tool
ALL	Acute lymphoblastic leukemia
AML	Acute myeloid leukemia
CLL	Chronic lymphocytic leukemia
CML	Chronic myeloid leukemia
TRIPOD	Transparent reporting of multivariable prediction models for individual prognosis or diagnosis
PICOTS	Population, intervention, comparator, outcome, timing, and setting
SD	Standard deviation
IQR	Interquartile ranges
PRISMA	Preferred reporting items for systematic

Yingzi Yang, Ayizhati Tuerxun and Xinqi Cai contributed equally to this work.

✉ Shengfeng Wang
shengfeng1984@126.com

[1] Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing, China

[2] Key Laboratory of Epidemiology of Major Diseases (Peking University), Ministry of Education, Beijing, China

Published online: 27 September 2025

	reviews and meta analyses
TTFT	Time to first treatment
WBC	White blood cell
FLT3	FMS-like tyrosine kinase 3
LDH	Lactate dehydrogenase
IGHV	Immunoglobulin heavy-chain variable region gene
AUC	Area under receiver operating characteristic curve
APL	Acute promyelocytic leukemia
AL	Acute leukemia
NPM1	Nucleophosmin 1
EPV	Events per variable
TLS	Tumor lysis syndrome

Introduction

Leukemia represents a class of malignant hematologic disorders, arising from the dysregulated proliferation and aberrant differentiation of hematopoietic stem cells within the bone marrow (DiNardo et al. 2023). The World Health Organization delineates leukemia into distinct subtypes such as acute lymphoblastic leukemia (ALL) and acute myeloid leukemia (AML), chronic lymphocytic leukemia (CLL) and chronic myeloid leukemia (CML) (Arber et al. 2016). Leukemia manifests clinically as anemia, hemorrhagic episodes, and susceptibility to infections (DiNardo et al. 2023), which is not only fraught with severe adverse effects but also cost-prohibitive. There were 461,000 incident cases, 320,000 deaths, and 11.0 million disability-adjusted life years due to leukemia globally in 2021. Guided by evidence-based clinical guidelines (Brown et al. 2021; Döhner et al. 2022; Pollyea et al. 2023), a stratified approach to leukemia patient care is essential, enabling targeted interventions for high-risk populations during diagnosis and treatment (DiNardo et al. 2023; El Chaer et al. 2023; Della Porta et al. 2024). The precise prevention, early diagnosis and precise treatment are urgent worldwide.

Clinical prediction models are, in essence, intricate mathematical equations designed to gauge the risk of an individual suffering from a disease in the present or having a medical outcome in the future (Odrobina 2024). Numerous prediction models for leukemia have been developed, yet few are actually utilized in clinical practice or incorporated into guidelines. Similar to the prediction models of childhood asthma (Smit et al. 2015) or contrast induced nephropathy (Silver et al. 2015), the poor methodological quality and high risk of bias (ROB) were the key restrictions for their using in clinical practice. Therefore, a systematic appraisal for these models is essential to assess their utility and limitations in leukemia. Though previous studies had reviewed prediction models of AML (Walter and Estey 2020), CML (Patnaik and Tefferi 2022) or CLL (Molica and Giannarelli 2021; Kreuzberger et al. 2020; Elhadary et al. 2023; González-Gascón et al. 2021; Yun et al. 2020), they often lacked systematic literature searches and were based on a limited number of selected models. Additionally, previous reviews have frequently identified deficiencies in the statistical analysis of prediction models (Smit et al. 2015; Silver et al. 2015; Damen et al. 2016; Feng et al. 2023), while many reviews of prediction models for leukemia ignored detailed information for statistical analysis such as candidate predictor handling, continuous predictor management, and missing data (Kreuzberger et al. 2020; Elhadary et al. 2023; Yun et al. 2020).

Moreover, ROB assessment for prediction models is essential for identifying potential sources of bias, enhancing model reliability, and supporting informed clinical decision-making, which aids in the continuous improvement and updating of models, ultimately contributing to their transparency and quality in research and clinical practice (Wolff et al. 2019). However, many reviews often ignored this procedure (Walter and Estey 2020; Patnaik and Tefferi 2022; Molica and Giannarelli 2021; Elhadary et al. 2023; González-Gascón et al. 2021; Yun et al. 2020) possibly due to the limit adoption of ROB assessment tools (de Jong et al. 2021) in this field of leukemia.

In this study, we aimed to provide an overview of all currently available multivariable prediction models of leukemia, systematically assess the methodological quality and identify prediction models with external validation and low ROB, which will be crucial for improving future model development and their clinical application in the field.

Methods

This systematic review was reported in accordance with transparent reporting of multivariable prediction models for individual prognosis or diagnosis: checklist for systematic reviews and meta-analyses (TRIPOD-SRMA, Supplementary eTable 1). The study protocol was registered on at the PROSPERO International Prospective Register of Systematic Reviews Website (registration number CRD42024544227).

Study selection

This review included all primary studies that developed and/or validated multivariable prediction models, tools, indices, or scores for estimating the individual risk of leukemia and/or its precursor conditions. We presented a detailed description of the study population, intervention, comparator,

outcome, timing, and setting (PICOTS) for this review in supplementary etable 2. Studies published before August 7, 2023 in English were identified by searching in databases including PubMed, Embase, Cochrane Library, and Web Of Science. The search strategy consisted of (1) prediction models and (2) leukemia. A complete list of search terms can be found in supplementary etable 3. Two reviewers independently screened the search results by abstract and full-text screening. Any disagreement was resolved through discussion.

Eligibility criteria

Studies were eligible if they met the following criteria: (1) aimed to develop or validate a risk prediction model for the occurrence of leukemia and/or its precursor conditions; (2) included not less than two predictors in the prediction model; (3) an original research article in a peer-reviewed journal. The studies were excluded if they met any of the following criteria: (1) being systematic reviews, meta-analyses, study protocols, trial registers, meeting abstracts, letters and dissertations; (2) experimental studies for animals; (3) articles in a language other than English, or where the full text was not available.

Data extraction

Two authors (AYZT and CXQ) independently performed the literature search and study selection. Data were extracted using a standardized data extraction form designed according to the CHecklist for Critical Appraisal and Data Extraction for Systematic Reviews of Predictive Modelling Studies (CHARMS) (Moons et al. 2014). Pairs of two reviews (AYZT and CXQ, ZZY and LZN, CXQ and ZY) independently extracted data by Epidata and Excel software. Any disagreements in literature search, study selection and data extraction were resolved by discussion with a third researcher (YYZ).

For each eligible study, we extracted the following items: (i) characteristics of study (including year, first author, country, study setting, type of prediction model), (ii) study design (including data sources, study populations, study outcomes, sample size), (iii) predictors (including types and number of candidate predictors, finally included predictors, variable transformation, and imputation of missing data), (iv) statistical analysis (including internal validation methods, statistical models, variable selection), (v) model performance (including discrimination and validation measures).

Risk of bias assessment and critical appraisal

The Prediction model Risk Of Bias ASsessment Tool (PROBAST) was used to assess the ROB for eligible studies (Wolff et al. 2019). PROBAST comprises four domains (study population, predictors, outcomes, and analysis) with a total of 20 signaling questions for examine prediction model studies' performance in study design, conduct, and data analysis. ROB for each eligible study was assessed by pairs of two researchers (AYZT and CXQ, ZZY and LZN, CXQ and ZY) independently and any disagreements in quality assessment were resolved by a third researcher (YYZ). We did not assess the applicability of included prediction models, as this was beyond the objectives of our review.

Statistical analysis

We summarized the characteristics of all included models using mean and standard deviation (SD) or medians and interquartile ranges (IQR) when appropriate for continuous variables. The Kolmogorov–Smirnov test was performed to evaluate the normality of the distribution for continuous variables. Categorical variables were presented as number and percentages. All statistical analyses were conducted using SPSS software (Version 25).

Results

Study overview

The study flowchart according to the Preferred Reporting Items for Systematic reviews and Meta Analyses (PRISMA) statement is shown in Fig. 1. Our review included 148 eligible articles, comprising prediction models for the following subtypes: AML (n = 61), CML (n = 22), CLL (n = 21), and other leukemias (n = 44).

Study designs

The development of leukemia prediction models has evolved significantly over time. Early studies (pre-2010) were predominantly conducted in the United States and Germany, often utilizing single-center, retrospective cohort designs. A notable surge in model development occurred after 2018. From this period onwards, China emerged as a major contributor, accounting for 42.7% of all models developed globally. This recent wave has increasingly leveraged large-scale electronic medical records (76.4%) and registry data (20.3%), reflecting a shift towards bigger data sources. Despite this evolution, the overwhelming focus (99.3%) remains on inpatient settings throughout the entire timeline.

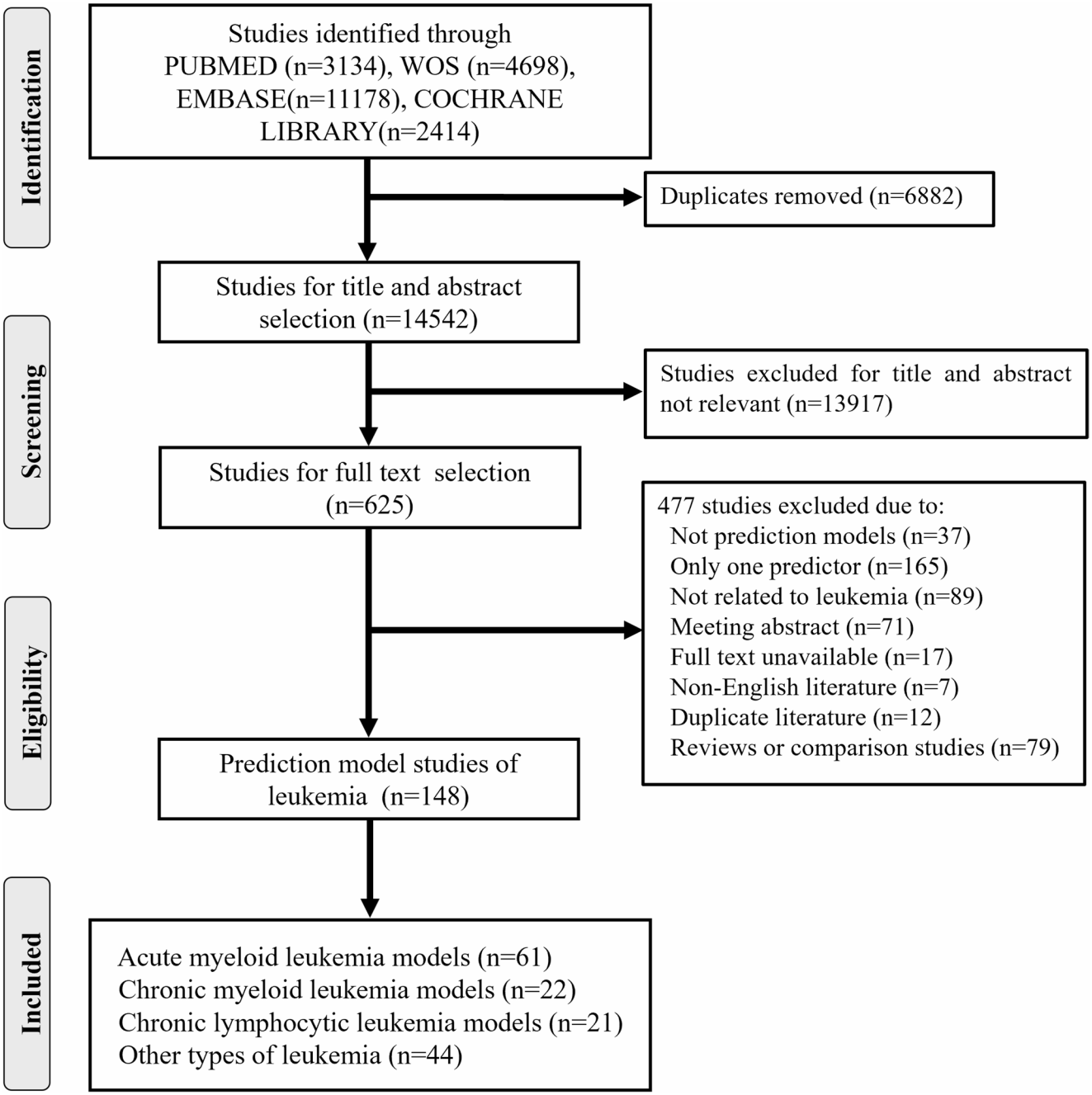

Fig. 1 Flowchart of study selection for leukemia prediction models

Most (n = 142, 95.9%) prediction models were derived from longitudinal cohort studies, five (3.4%) from case–control studies, and one (0.7%) from a randomized controlled trial. The sample sizes for model development and external validation were 412(213–909) and 306(131–680) (Supplementary eTable 4 and 5).

Study populations and outcome

The predominant target population was inpatients with AML (n = 61, 41.2%), followed by CML (n = 22, 14.9%), CLL (n = 21, 14.2%), and ALL (n = 18, 12.2%). Most models predicted survival (n = 100, 67.6%), diagnosis (n = 12, 8.1%) and death (n = 8, 5.4%). Specifically, Among the 61 AML models, survival prediction was the predominant outcome (n = 50, 82.0%), followed by diagnosis (n = 3, 4.9%) and death prediction (n = 3, 4.9%). For the 22 CML models,

the majority focused on survival (n = 16, 72.7%), while the remaining models addressed other outcomes (n = 6, 27.3%). Among the 21 CLL models, survival prediction was also the most common outcome (n = 15, 71.4%), with time to first treatment (TTFT) being the second most frequent focus (n = 4, 19.0%). The prediction horizon ranged from 3 days to overall survival, and it was not specified in 57 (38.5%) models (Supplementary eTable 6).

Predictors

The most common predictors were age, white blood cell (WBC) count, platelet count, cytogenetic risk group and blasts cell (Table 1). Except for these predictors, in 61 models of AML, FMS-like tyrosine kinase 3 (FLT3) status (16.4%) and lactate dehydrogenase (LDH) level (14.8%) were also used frequently. In 22 models of CML, spleen size (27.3%) and hemoglobin count (22.7%) were also commonly used. In 21 models of CLL, immunoglobulin heavy-chain variable region gene (IGHV) (38.1%), Rai stage (33.3%), LDH level (23.8%), and lymphocyte count (14.3%) were also included frequently (Table 1). The candidate predictors and their variable transformation methods of all included studies were presented in supplementary eetable 1. We also collected the 283, 80 and 108 candidate predictor lists for AML, CML and CLL (Supplementary eeTable 2–4).

Statistical analysis

Handling of missing data and continuous variables

Missing data proportions for predictors were reported in 68 models (45.9%). The percentage of missing data ranged from 0.01% to 93.6%. Only 38 (25.7%) models specified their approach to handling missing data. Of them, 14 (36.8%) models utilized complete case analysis, and 8 (21.1%) models using multiple imputation methods to impute missing data.

Of the 148 models, 94 (63.5%) models transformed continuous variables into categorical variables and 45 (30.4%) models used continuous variables without modification. There were 24(16.2%) models didn't report how continuous variables are handled and only 4 (2.7%) models reported using logarithmic isometric conversions for continuous

Table 1 The top twenty most frequent predictors in prediction models of different leukemia subtypes

Total(n = 148)		AML(n = 61)		CML(n = 22)		CLL(n = 21)	
Predictors	n (%)	Predictors	n (%)	Predictors	n (%)	Predictors	n (%)
Age	78(52.7)	Age	33(54.1)	Blasts cell	12(54.5)	IGHV	8(38.1)
WBC count	43(29.1)	Cytogenetic risk group	21(34.4)	Age	10(45.5)	Rai stage	7(33.3)
Platelet count	28(18.9)	WBC count	20(32.8)	Platelet count	9(40.9)	Age	5(23.8)
Blasts cell	27(18.2)	FLT3 status	10(16.4)	Spleen size	6(27.3)	LDH level	5(23.8)
Cytogenetic risk group	27(18.2)	Platelet count	9(14.8)	WBC count	5(22.7)	Lymphocyte Count	4(19.1)
Hemoglobin count	21(14.2)	LDH level	9(14.8)	Hemoglobin count	5(22.7)	WBC count	3(14.3)
LDH level	20(13.5)	Blasts cell	8(13.1)	Basophils	3(13.6)	Hemoglobin count	3(14.3)
Disease type	19(12.8)	NPM1 status	8(13.1)	ASXL1 mutations	2(9.1)	2-microglobulin level	3(14.3)
Sex	17(11.5)	Disease type	7(11.5)	Anemia	2(9.1)	TP53	3(14.3)
ECOG performance status	15(10.1)	ECOG performance status	7(11.5)	Blood Basophilia	2(9.1)	ALC	2(9.5)
Donor type	12(8.1)	CEBPA Mutation Status	4(6.6)	Cytogenetic risk group	2(9.1)	Binet stage	2(9.5)
FLT3 status	11(7.4)	ELN cytogenetic risk status	4(6.6)	LDH level	2(9.1)	Cytogenetic risk group	2(9.5)
MRD Status	10(6.8)	HCT-CI	4(6.6)	Lymphocyte count	2(9.1)	Del(11q)	2(9.5)
IGHV	9(6.1)	Sex	4(6.6)	Comorbidity(ies)	1(4.5)	Del(17p)	2(9.5)
NPM1 status	8(5.4)	MRD Status	4(6.6)	Eosinophils	1(4.5)	Disease Status	2(9.5)
Rai stage	7(4.7)	Creatinine	3(4.9)	Hepatomegaly	1(4.5)	Time from last therapy	2(9.5)
2-microglobulin level	7(4.7)	Hemoglobin count	3(4.9)	Monocyte count	1(4.5)	Unmutated IGHV	2(9.5)
Disease status at HSCT	7(4.7)	WT1	3(4.9)	Sex	1(4.5)	ApoA, g/L	1(4.8)
GVHD	6(4.1)	CBF fusion gene	2(3.3)	Splenomegaly	1(4.5)	ATM mutation	1(4.8)
Spleen size	6(4.1)	Donor type	2(3.3)	WHO classification	1(4.5)	CD38 expression	1(4.8)

AML, acute myeloid leukemia; CML, chronic myeloid leukemia; CLL, chronic lymphocytic leukemia; WBC, white blood cell; LDH, lactate dehydrogenase; ECOG, Eastern Cooperative Oncology Group; HCT-CI, Hematopoietic Cell Transplantation Comorbidity Index; HSCT, hematopoietic stem cell transplantation; GVHD, graft-versus-host disease; IGHV, immunoglobulin heavy-chain variable region; ELN, European Leukemia Net; MRD, minimal residual disease; CBF, core-binding factor; ALC, absolute lymphocyte count; TP53, Tumor Protein p53; NPM1, nucleophosmin 1; HCT-CI, Hematopoietic Cell Transplantation Comorbidity Index

variables. A total of 121 models (81.8%) models incorporated continuous predictors in their final models (Supplementary eeTable 1).

Modelling method

Prediction models selected predictors predominantly by stepwise method (n = 63, 42.6%) or multivariate analysis (n = 48, 32.4%), with an average of 10 (ranged 7 to 15) candidate predictors and 4 (ranged 3 to 6) included predictors per model. The number of events for each variable included in the final prediction model could be calculated for 66 (44.6%) models and ranged from 3.33 to 808.6, and for 10 (15.2%) models, their number of events for each variable was less than 10. The most common modeling approaches were the Cox proportional-hazards regression (n = 108, 73.0%), logistic regression (n = 18, 12.2%, Table 2).

Model performance and validation

A total of 77 (52.0%) models didn't report the method for internal validation and 85 (57.4%) models did not have external validation. The remaining models used random splitting (n = 31, 20.9%), bootstrapping (n = 21, 14.2%), or cross-validation (n = 10, 6.8%) for internal validation and used data from different setting (n = 32, 21.6%), district (n = 9, 6.1%) or period (n = 8, 5.4%) for external validation.

On discriminative performance, 84 (56.8%) models lacked information. Of the remaining 64 models, the area under receiver operating characteristic curve (AUC) or C-statistics ranged from 0.60 to 0.99. There were 47 (74.6%) models reporting an AUC above 0.70, and 16 (25.4%) models reported an AUC between 0.50 and 0.70. For calibration, 114 (77.0%) models didn't report any calibration measures or plot. Calibration plot or Hosmer–Lemeshow test were used in 26 (17.6%) and 9 (6.1%), respectively. All reported that the models were well-calibrated (Table 2). Detailed information on model performance measures of all included models was presented in supplementary eeTable 5.

Model presentation

Only 34 (23.0%) models reported the complete regression equation including regression coefficients and intercepts or baseline hazards. Over half of models (n = 81, 54.7%) were presented as sum scores, and 41 (27.7%) models were presented using score chart. There were 12 (8.1%) and 10 (6.8%) models presented as nomograms or automated calculation.

Risk of bias assessment

Of the 148 models, 139 (93.9%) were rated as high overall ROB (Supplementary eFig. 1). Most models were rated low ROB in the domains of participants (n = 132, 89.2%), predictors (n = 131, 88.5%), and outcomes (n = 127, 85.8%). However, 137 (92.5%) models had high ROB in the analysis domain, primarily due to issues with model performance measures (n = 91, 61.5%), unclear internal validation (n = 84, 56.8%), mishandling of continuous predictors (46.6%), and selection of predictors by univariate analysis (n = 63, 42.6%, Supplementary eeTable 6). The ROB results of domains and signaling questions were similar across models of different leukemia subtypes (Supplementary eFig. 2-1 to eFig. 3-4).

To present the most clinically relevant models for major leukemia subtypes, we selected prediction models based on predefined criteria—external validation or large development sample size (Supplementary eTable 7), including Li (2022) for AML, Zhang (2018) for acute promyelocytic leukemia (APL), Gao and Liu (2020) for ALL, Pfirrmann (2016) for CML, the International CLL-IPI working group (2016) for CLL, and Boldú (2021) for acute leukemia (AL). For AML, APL, and ALL, all selected models were externally validated but developed from relatively small samples and had unclear risk of bias due to limited methodological reporting. The CML, CLL, and AL models had larger sample sizes but lacked comprehensive performance data and were assessed as high risk of bias. Therefore, none of these models are recommended for direct clinical application at this time.

Discussion

This systematic review evaluated 148 prediction models, with a higher number of models targeting inpatients with AML, CML, and CLL, primarily focused on predicting survival of leukemia patients. The handling of predictors and missing data was incompletely reported in most models. Moreover, many models have methodological issues, mainly concerning the handling of continuous variables and model validation. Consequently, most models had high ROB, which diminished their clinical utility in this field. Based on these notable methodological limitations throughout their development, we address these issues sequentially from predictors, data processing, modeling methods, validation and performance, ROB assessment, to heterogeneity across studies.

Table 2 Statistical analysis and model performance for prediction models of leukemia

	Total (n = 148)	AML (n = 61)	CML (n = 22)	CLL (n = 21)
Source of data				
Electronic medical record	113(76.4)	43(70.5)	17(77.3)	18(85.7)
Registry data	30(20.3)	14(23.0)	5(22.7)	3(14.4)
Others[a]	5(3.3)	4(6.5)	0(0.0)	0(0.0)
Shrinkage methods[b]				
Stepwise	63(42.6)	28(45.9)	7(31.8)	9(42.9)
Based on multivariate analysis	48(32.4)	13(21.3)	7(31.8)	11(52.4)
LASSO selection	10(6.8)	8(13.1)	0(0.0)	0(0.0)
Bootstrap selection	7(4.7)	3(4.9)	2(9.1)	1(4.8)
Likelihood-ratio tests	5(3.4)	5(8.2)	0(0.0)	0(0.0)
Ridge regression	2(1.3)	2(3.3)	0(0.0)	0(0.0)
Unclear	16(10.8)	3(4.9)	7(31.8)	1(4.8)
Number of candidate predictors	10(7–15)	11(7–15)	10(5–13)	12(7–17)
Number of predictors	4(3–6)	5(3–6)	5(3–5)	4(4–5)
EPV < 10[c]	10(15.2)	3(15.8)	2(28.6)	0(0.0)
Modelling method				
Cox proportional-hazards regression	108(73.0)	49(80.3)	17(77.3)	19(90.5)
Logistic regression model	18(12.2)	5(8.2)	1(4.5)	1(4.8)
Neural network	2(1.3)	0(0.0)	0(0.0)	0(0.0)
Decision-tree model	2(1.3)	0(0.0)	0(0.0)	0(0.0)
Linear regression model	1(0.7)	0(0.0)	0(0.0)	0(0.0)
Generalized linear regression	1(0.7)	0(0.0)	0(0.0)	0(0.0)
Support Vector Machine	1(0.7)	0(0.0)	0(0.0)	0(0.0)
Random forest	1(0.7)	1(1.6)	0(0.0)	0(0.0)
Other[d]	3(2.0)	0(0.0)	1(4.5)	1(4.8)
Unclear	11(7.4)	6(9.8)	3(13.6)	0(0.0)
Internal validation				
Random split	31(20.9)	9(14.8)	4(18.2)	5(23.8)
Bootstrapping	21(14.2)	12(19.7)	3(13.6)	4(19.0)
Cross validation	10(6.8)	5(8.2)	1(4.5)	1(4.8)
Non-random split	9(6.1)	4(6.6)	1(4.5)	0(0.0)
Unclear	77(52.0)	31(50.7)	13(59.2)	11(52.4)
External validation				
Different setting	32(21.6)	18(29.5)	5(22.7)	6(28.6)
Geographical	9(6.1)	5(8.2)	2(9.1)	2(9.5)
Temporal	8(5.4)	5(8.2)	1(4.5)	0(0.0)
Other	14(9.5)	6(9.8)	1(4.5)	0(0.0)
No external validation	85(57.4)	27(44.3)	13(59.2)	13(61.9)
Discrimination measures[b]				
C-Statistic	27(18.2)	3(19.0)	2(9.1)	8(38.1)
ROC/AUC graph	46(31.1)	9(14.8)	3(13.6)	3(14.3)
Other[e]	6(4.1)	3(19.0)	0(0.0)	0(0.0)
None	84(56.8)	31(50.8)	18(81.8)	10(47.6)
Calibration measures[b]				
Calibration plot	26(17.6)	14(23.0)	5(22.7)	1(4.8)
Hosmer–Lemeshow test	9(6.1)	2(3.3)	0(0.0)	2(9.5)
None	114(77.0)	45(73.8)	17(77.3)	18(85.7)
Measures for survival analysis[b]				
Log-rank test	92(62.2)	38(62.3)	16(72.7)	16(76.2)
Risk group curves	120(81.1)	54(88.5)	17(77.3)	20(95.2)
Presentation of models[b]				
Sum score	81(54.7)	35(57.4)	15(68.2)	9(42.9)
Score chart	41(27.7)	15(24.6)	4(18.2)	12(57.1)
Full equations	34(23.0)	14(23.0)	8(36.4)	2(9.5)

Table 2 (continued)

	Total (n = 148)	AML (n = 61)	CML (n = 22)	CLL (n = 21)
App calculator	10(6.8)	4(6.6)	0(0.0)	1(4.8)
Nomograms	12(8.1)	6(9.8)	0(0.0)	2(9.5)

[a]Others included TCGA database(n = 2), TCGA and GTEx database(n = 1), Health Insurance Research Database(n = 1) and Datasets of BEATAML1.0, TARGET-AML, TCGA-LAML(n = 1)

[b]Multiple-choice question

[c]Percentages here were calculated within 66 models, including 19 AML models, 7 CML models and 10 CLL models

[d]Includes non-linear least-squares fitting, Orthogonal SoftMax Layer, an ensemble algorithm composed of 28 machine learning algorithms

[e]Sensitive, specificity or accuracy

TCGA, the Cancer Genome Atlas; GTEx, Genotype-Tissue Expression; AUC, area under receiver operating characteristic curve. EPV, events per variable

Current study findings

For predictors, clear differences emerged across leukemia subtypes, reflecting both biological heterogeneity and methodological inconsistencies. In AML models (n = 61), cytogenetic risk groups and molecular markers such as FLT3 and nucleophosmin 1 (NPM1) were frequently included, underscoring the integration of molecular profiling into modern risk stratification. Traditional factors, including age and WBC count, remain fundamental, suggesting that robust AML models combine conventional clinical indicators with emerging genomic biomarkers. CML models (n = 22) displayed a distinct profile, relying primarily on traditional clinical and laboratory predictors—percentage of blast cells, spleen size, platelet count, and LDH levels—with limited incorporation of molecular genetics. The prominence of spleen size highlights its historical and clinical importance in tracking disease burden and treatment response, particularly in the pre-tyrosine kinase inhibitor era. By contrast, CLL models (n = 21) emphasized immunogenetic predictors such as IGHV mutation status (38.1%), Tumor Protein p53 (14.3%), and Rai stage (33.3%), together with serum biomarkers like LDH (23.8%) and β2-microglobulin (14.3%). These findings underline the central role of immunogenetics and tumor burden assessment in CLL prognosis.

Despite a general shift toward incorporating molecular and immunogenetic predictors, our finding that over 100 predictors appeared only once or twice across studies indicates continued fragmentation in predictor selection. The same challenge was also faced in prediction models of cardiovascular events (Damen et al. 2016). Similar difficulties in selecting predictors were observed in prognosis models for neuroglioma (Molinaro et al. 2016), likely due to the complex pathogenesis of these rare diseases, which involve multiple genetic abnormalities and alterations in cellular signaling pathways, and the critical factors in prediction models are not yet fully understood. To address this, we propose a more consistent approach to predictor selection within subtypes (Supplementary eeTable 2–4), informed by our comprehensive candidate lists for AML (n = 283), CML (n = 80), and CLL (n = 108). Machine learning methods may further aid in identifying optimal predictor sets, improving the robustness and applicability of future models.

Unlike the selection of predictors, prediction models for different leukemia subtypes presented similar problems in their statistical methodology. In data processing, only 68 (45.9%) of included leukemia prediction models reported the proportion of missing predictors, which was consistent with Kreuzberger, N et al.'s findings. They found less than 50% of adult CLL models reported the number of people with missing data (Kreuzberger et al. 2020). Furthermore, most studies that reported missing data did not handle it correctly. Some prediction models either deleted the data (Cai et al.2020; Pastore et al. 2014), risking sample loss and model accuracy, or used imputation without considering data distribution among different patients (Zhang et al. 2018; Krug et al. 2010; Kim et al. 2019). These findings underscores three critical areas for enhancing leukemia prediction models: adhering to TRIPOD guidelines (Collins et al. 2015) by reporting missing data rates in new models, selecting appropriate data handling methods as Li et al.'s recommendations (Li et al. 2024; Sperrin et al. 2020), and assessing existing models' quality by collecting and reporting missing data and handling methods in line with CHARMS and PROBAST checklists (Wolff et al. 2019; Moons et al. 2014). In fact, systematic reviews of prediction models for other diseases such as poor lung development, depression, perineal lacerations also paid attention to the assessment of missing data handling (Romijn et al. 2023; Tan et al. 2023; Hu et al. 2023). Therefore, enhancing missing data management in future leukemia prediction model studies will significantly elevate model quality.

In terms of variable transformation, no previous leukemia prediction model reviews systematically addressed the transformation of predictors. This review found that over 65% of models categorized continuous variables like age, WBC, LDH, Hemoglobin and Platelet Count. While this practice is not necessarily wrong, it may lead to loss of data

information and reduced statistical power, and is generally not recommended for developing prediction models. The same issue was also observed in depression and perineal laceration models (Tan et al. 2023; Hu et al. 2023), highlighting variable transformation in prediction models was a widespread problem that developers should address more diligently (Efthimiou et al. 2024). To handle continuous variables well, researchers should analyze the statistical distribution characteristics and the clinical thresholds of the included variables, and apply appropriate transformation methods such as logarithmic or square root transformations to meet the specific model requirements. Additionally, we found that there were multiple different definitions for outcomes, which made it difficult for us to compare and select existing prediction models. Future development and improvement of leukemia prediction models should pay more attention to the proper handling of continuous variables and the unified definition of outcomes. This review has compiled the handling methods of predictors in previous literature in supplementary eetable1 to provide valuable reference for researchers.

Additionally, prediction models of leukemia also had many issues during model development and validation. Firstly, regarding sample size, more than half of the models either fail to report or have less than 100 events, and only 19.4% of models had an events per variable (EPV) value above 10.The number of outcome events used for development or validation should not be less than 100 (Peduzzi et al. 1995), or the model's EPV should be above 10 (Wynants et al. 2015). Few outcome events may lead to overfitting, insufficient statistical power and inaccurate parameter estimation (Wynants et al. 2015). Most studies indicated the model could be practical, but they also acknowledged the need for further external validation. The issue was also common in prediction models of breast cancer (Phung et al. 2019), and esophageal and gastric cancers(van den Boorn et al. 2018). Leukemia and many other rare diseases had more difficulties in gathering enough cases for modeling. However, researchers should also be aware of the sample size requirements for modeling and consider other methods, such as extending data collection periods, broadening locations, or tapping into real-world databases to meet these requirements.

Secondly, for model selection, traditional models such as cox and logistic regression have been widely used in prediction models of leukemia and many other diseases for their solid theory and clear interpretability, and they help to clearly understand the impact of various factors on the diagnosis or prognosis. However, they often face limitations due to strict assumptions. For instance, some well-established prognostic variables in leukemia, such as the status of specific genetic markers (e.g., FLT3 internal tandem duplication in AML), often violate the proportional hazards assumption required by Cox regression models, as their effect on hazard (risk of death or relapse) may diminish or change over time rather than remain constant. In this review, only nine articles used machine learning methods such as neural networks and decision tree. Compared to traditional statistical models, machine learning can handle complex data and variable interactions with greater flexibility but are less interpretable, require large amounts of data, and demand high-quality annotated data (Ley et al. 2022; Bzdok et al. 2018). In the future, it may be worthy to develop hybrid models, including using machine learning to mine features firstly and then using traditional models for prediction and interpretation; or employing ensemble learning methods to integrate the predictive results of multiple models through weighting and other techniques.

For model validation and performance, most leukemia models lacked external validation or essential metrics of model discrimination or calibration, which were consistent with Walter and Estey's finding in AML models (Molica and Giannarelli 2021). The same issues were also common in prediction models of various diseases, and even for cardiovascular prediction models, no more than 40% of them had external validation or reported both on discrimination and calibration metrics (Damen et al. 2016). Models without external validation risk being overfitted, losing their real-world clinical value and making it hard to apply them broadly across different treatment settings and centers. In addition, although some included models have good discrimination, there is still room for improvement overall. The existing prediction models for leukemia are not ready for widespread implementation in clinical practice. To enhance the performance of leukemia prediction models, the leukemia research community could consider establishing a publicly accessible database to facilitate external validation of prediction models or prioritizing interdisciplinary collaboration to standardize modeling practices, enhance data quality, and increase sample sizes.

In light of these issues, we consistently observed a high or unclear ROB across numerous prediction models for AML, CML, and CLL, primarily stemming from the predictors and analysis domains. A pervasive issue was inadequate reporting, which often obscured crucial methodological details. Moreover, the ROB was also consistently distributed in most signaling questions across different leukemia subtypes, indicating that methodological deficiencies—rather than disease-specific characteristics—were the main cause of the high ROB. While the overall ROB was commonly high in leukemia models, we noted that model type had a substantial influence on the observed statistical methodological issues. Specifically, diagnostic models frequently exhibited problems in predictor handling and the

reporting of performance metrics. In contrast, prognostic models largely struggled with predictor selection and the rigor of model validation methods. Therefore, to substantially enhance the quality and reliability of leukemia prediction models, researchers should strengthen methodological rigor—particularly in predictor selection, statistical handling, and validation methods—throughout the entire model development and evaluation process.

Beyond the identified methodological limitations and biases above, there seems to be a broader challenge: the substantial heterogeneity across studies—stemming from differences in leukemia subtypes, population eligibility, predictor definitions, outcome measures, and analytical methods—significantly limits direct comparability between models and reduces the certainty of the synthesized evidence. However, this variability arises not only from methodological diversity but also reflects fundamental biological and clinical differences between subtypes such as AML, CML, and CLL, which inherently shape model development and applicability. Rather than solely a limitation, this heterogeneity accurately represents the current fragmented landscape of leukemia prediction research. By systematically mapping these sources of variation across subtypes, our review provides a necessary foundational overview that supports and informs subsequent subtype-specific systematic reviews and validation studies.

Implications for research

Despite the substantial number of models included in this review, the field of leukemia prediction modeling is commonly affected by an insufficient quantity of studies and suboptimal methodological quality. Specifically, many studies lack clear objectives and definitions, such as failing to specify the time horizon for outcomes and predictors. Furthermore, methodological details are generally inadequately reported, with severe deficiencies in documenting key aspects like model validation and variable handling. On the statistical front, not only is the application of advanced machine learning methods sparse, but the implementation of traditional approaches is also often flawed, with inconsistencies in predictor handling that directly impair model comparability. Therefore, we strongly recommend promoting methodological standardization across this research domain. This includes establishing uniform reporting standards to enhance transparency, reinforcing researchers' adherence to international guidelines such as CHARMS (Moons et al. 2014), TRIPOD (Collins et al. 2015), and PROBAST (Wolff et al. 2019), studying methodological literature of prediction model studies, like the step-by-step guide for developing prediction models (Efthimiou et al. 2024), the guidelines for clinical prediction models using regression or machine learning methods in developing, internally validating, and evaluating (Collins et al. 2024; Moons et al. 2012, 2015), and encouraging the broader application of advanced techniques like machine learning to handle complex clinical data in model development.

For different leukemia subtypes, we propose that future research directions should be more targeted. Subsequent research priorities should be determined by the number and quality of existing models for a specific subtype-outcome pair: development should be prioritized where models are scarce; validation, updating, or systematic reviews should be the focus when a sufficient number of models exist; and models with robust performance can be further evaluated for their clinical application and health-economic value. Based on this framework, we propose the following specific recommendations: For all leukemia subtypes, external validating, updating, and systematically reviewing existing survival prediction models should be prioritized. But the development of new prediction models should be strategically directed towards distinct, under-researched clinical outcomes specific to each subtype. Specifically, for AML, development should target models predicting diagnosis, relapse, hemorrhage, and tumor lysis syndrome (TLS), while for ALL, key unaddressed outcomes include diagnosis, relapse, infection, and bone-related event. For CML, new models should target outcomes including diagnosis, relapse, TTFT, cardiovascular events, and molecular response. For CLL, efforts should focus on predicting diagnosis, TTFT and severe infections.

Implications for practice

Systematic reviews inform evidence-based healthcare decisions; therefore, we attempted to identify the most clinically applicable models for different leukemia subtypes. However, we found that identifying ideal models that simultaneously satisfy "high performance, low risk of bias, and rigorous validation" was nearly impossible. Therefore, our pragmatic selection criterion was to identify models with the "greatest current potential for clinical translation" within each subtype, based primarily on the feasible metrics of "external validation" or "large development sample size" (Supplementary eTable 7). Nevertheless, we found that none of these models can currently be recommended for immediate clinical application. Key barriers include high risk of bias in CML, CLL, and AL models due to methodological shortcomings, and incomplete reporting leading to unclear ROB in AML, APL, and ALL models. Therefore, we have developed a set of modeling recommendations tailored to the specific shortcomings observed in the leukemia

Table 3 Methodological recommendations for developing a new leukemia prediction model according to key developing steps

Key developing steps	Methodological recommendations
Define research objective and model type	
Population	It is recommended to develop prediction models tailored to individual leukemia subtypes rather than employing a hybrid modeling approach across multiple subtypes
Outcome	The predicted outcome can cover the entire disease course of leukemia, including but not limited to diagnosis, triage, treatment selection, complication risk assessment, survival, and mortality
Predictors	Candidate predictors should be selected based on their significant clinical relevance to the disease and the predicted outcome. Refer to Supplementary eeTables 2–4 for suggested predictor lists pertaining to AML, CLL, and CML
Time horizon	The time horizon for the prediction model must be clearly defined, including the time point/period for predictors and outcome, and the time for model application
Prediction model type (Diagnostic or prognostic)	It is critical to explicitly distinguish between diagnostic models, which predict the current disease state, and prognostic models, which predict future events or the disease course, as they correspond to distinct clinical contexts and should not be used interchangeably
Data preparation	
Data collection	Given the low incidence of leukemia, it is recommended to collect data from multiple centers and diverse sources to ensure an adequate and representative sample size for robust model development
List of candidate predictors	A comprehensive list of candidate predictors should be explicitly provided during model development. Given the distinct characteristics of different leukemia subtypes, it is recommended to prioritize the inclusion of key emerging biomarkers, such as those derived from genomic, hematological, and proteomic analyses
Data cleaning	As leukemia prediction often involves a high-dimensional predictor set (e.g., genetic, proteomic, and hematologic variables), it is not recommended to convert continuous variables into categorical form, to avoid loss of information and statistical power
Sample size	It is recommended to perform a sample size calculation using events per variable (EPV) to ensure sufficient power for model development and validation
Internal validation and dataset splitting	If the sample size is relatively small, cross-validation and bootstrapping are recommended over simple random splitting to obtain more robust performance estimates
Model development and selection	
Determine modeling methods based on outcome type	For prognostic models, the Cox proportional hazards model is generally the preferred method, but its fundamental assumption of proportional hazards must be rigorously verified. In cases where this assumption is violated, alternative approaches such as time-dependent Cox models or parametric survival models should be considered. Given the prevalence of complex, high-dimensional, and non-linear high-throughput data (e.g., genomics, transcriptomics), machine learning classifiers (e.g., Random Forest, XGBoost, SVM) are recommended as they may offer higher predictive accuracy
Determine predictor selection method	Traditional methods relying on univariate/simple multivariate analyses or stepwise regression are not recommended due to their propensity for overfitting and unstable results. Penalized regression techniques, such as LASSO (Least Absolute Shrinkage and Selection Operator) or Ridge regression, are preferred for their ability to perform simultaneous variable selection and regularization, leading to more robust models
Model training and hyperparameter tuning	To optimize model performance, it is recommended to employ systematic strategies for hyperparameter tuning. Methods such as Grid Search or Random Search are valuable for efficiently exploring the hyperparameter space and identifying the optimal configuration
Model performance evaluation	
Determine the dataset for performance evaluation	It is strongly recommended to clearly and detailed report the internal and external validation methods used in models of leukemia. Multi-center, cross-regional, and cross-temporal collaborations should be actively pursued for effective external validation
Discrimination	It is strongly recommended to report clear, quantitative discrimination metrics. For example, the AUC or C-index and their 95% Confidence Interval. Simply presenting an ROC curve is insufficient
Calibration	Robust reporting of calibration metrics is strongly recommended. This includes visual assessment via calibration plots, complemented by the Brier score. For survival models, present time-dependent calibration plots to assess predictive accuracy across various time horizons
Clinical utility	Decision Curve Analysis (DCA) is an effective tool for evaluating clinical utility. It is recommended to include DCA to determine if the model is superior to "treat all" or "treat none" strategies across a range of intervention thresholds
Model selection	Select models based on a balance of model performance, number of predictors, and clinical relevance—not discrimination alone. A clear rationale must be provided, detailing the comprehensive evaluation of model performance (including discrimination and calibration) and simplicity
Determine model application format	
Model application scenario	The model's intended clinical application must be thoroughly defined, including the target patient population, predicted outcome, and detailed context of use, including the clinical setting (e.g., outpatient, inpatient), timing of prediction, intended end-user (e.g., hematologist, general practitioner), and mode of integration into clinical workflow

Table 3 (continued)

Key developing steps	Methodological recommendations
Model presentation	Models can be delivered in various formats, including mobile applications (APPs), nomograms, web-based calculators, and integration into Clinical Decision Support Systems (CDSS). Consider multimodal presentation, such as nomograms for visual interpretation alongside web or mobile calculators. Integration within a CDSS is also recommended
Clinical application methods	Clearly define risk strata (e.g., low, intermediate, high risk) with specific cut-off values based on statistically optimal thresholds (e.g., Youden's index) and clinical expert consensus. Propose specific, actionable clinical intervention recommendations for each risk group
Model reporting and quality assessment	
Reporting	The development and validation of leukemia prediction models should adhere to the TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) (Collins et al. 2015) guidelines to ensure transparent and standardized reporting
Bias and applicability assessment	The PROBAST (Prediction model Risk Of Bias ASsessment Tool) (Wolff et al. 2019) should be employed to evaluate the risk of bias and applicability of individual prediction models, not limited to their use in systematic reviews. In addition to overall and domain-level results, the results of 20 signaling questions (SQ) and their rationale should also be clearly documented

prediction model literature. These suggestions, structured around key stages of prediction model development, were summarized in Table 3 with the aim of assisting researchers in improving the rigor and applicability of future models in this field.

Although no current models are suitable for broad clinical implementation, we have identified the following key features that would enhance the practical applicability and value of future models, along with their rationale: (1) Inclusion of large, multicenter and diverse patient cohorts to ensure robust validation and generalizability; (2) Integration of disease-specific and prognostically significant biomarkers to enhance predictive accuracy; (3) Use of continuous predictors or advanced categorization methods to preserve information and avoid arbitrary cutoffs; (4) Adoption of transparent reporting with complete model equations and evaluation metrics for reproducibility; (5) Clearly defining the intended context and format for model application, including timing, setting, end-user, and presentation (e.g., online calculators or Clinical Decision Support Systems integration), to facilitate practical clinical use. Models incorporating these features are better positioned for translation into real-world leukemia management, as they address methodological rigor, clinical relevance, and usability.

Strength and limitations

The review conducted a systematic literature search and reviewed the methodological details of the leukemia prediction models. We firstly collected lists of candidate predictors, included predictors and their variable transformation methods, summarized the handling of missing data and utilized PROBAST to assess ROB. However, the study still had several limitations. Firstly, due to the high heterogeneity among the included leukemia models in terms of study populations, outcomes, and prediction timeframes, coupled with an insufficient number of models designed for similar purposes and inadequate reporting of performance parameters including 95% confidence intervals (Supplementary eTable 6), a meta-analysis was not feasible, thereby limiting the ability to quantitatively compare and synthesize the predictive value of existing models. Secondly, the publication bias remains a concern, due to the high heterogeneity among the included models, limited sample size, and incomplete reporting of model performance metrics, we were unable to assess publication bias using a funnel plot. Furthermore, the restriction to English-language publications may have introduced language bias, excluding relevant studies published in other languages and potentially limiting the global generalizability of our findings. Additionally, despite a comprehensive search across four major databases, it is possible that we missed eligible studies not indexed in these sources or published in grey literature, which may have led to an incomplete summary of existing models. Thirdly, considerable methodological heterogeneity in predictor definitions and measurement, coupled with substantial variations in the characteristics and quality of external validation datasets, hampered meaningful comparison between models and introduced uncertainty in assessing their reliability and generalizability.

Conclusion

In conclusion, numerous prediction models for leukemia were developed, but they had many methodological flaws and were severely limited for clinical practice, mainly for incomplete reporting, lack of external validation and having high overall ROB. Future research should pay attention to the multi-dimensional verification of model performance, enrich the model presentation forms, and construct

prediction models with low ROB and strong practicability, so as to make them truly a powerful tool in the leukemia diagnosis and treatment system and play a more critical role in improving the overall medical status of leukemia patients.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00432-025-06314-7.

Acknowledgements We express deep thanks to the authors of the original studies who developed or validated prediction models of leukemia, which are critical to safeguard the quality of the study.

Author contributions WSF and YYZ conceived and designed the study, and supervised the conduct of the review. AYZT and CXQ searched literature, and screened the abstract and full-text of searched literature. AYZT, CXQ, CXY, ZZY, ZY and LZN extracted data and conducted ROB assessment. YYZ, AYZT and CXQ checked the data and ROB assessment. AYZT and CXQ analyzed the data. AYZT and CXQ wrote the first draft, and WSF and YYZ gave critical comments and contributed to the writing of the manuscript. WSF and YYZ take full responsibility for the work as a whole, including the study design, access to data and decision to submit. All the authors agreed to submit and publish the manuscript.

Funding National Nature Science Foundation of China—"Design and Empirical Research on Rare Disease Multi-Security Mechanisms Driven by Data Intelligence" (ID: 72342015). National Nature Science Foundation of China—"Processing strategies and case studies of the time-related variables in predicting adverse drug reactions based on dynamic Bayesian networks and transfer learning" (ID: 82,173,616).

Data availability No datasets were generated or analysed during the current study.

Declarations

Competing interests The authors declare no competing interests.

References

Arber DA et al (2016) The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood 127(20):2391–2405

Boldú L et al (2021) A deep learning model (ALNet) for the diagnosis of acute leukaemia lineage using peripheral blood cell images. Comput Methods Programs Biomed 202:105999

Brown PA et al (2021) Acute lymphoblastic leukemia, Version 2.2021, NCCN clinical practice guidelines in oncology. J Natl Compr Canc Netw 19(9):1079–1109

Bzdok D et al (2018) Statistics versus machine learning. Nat Methods 15(4):233–234

Cai P et al (2020) An effective early death scoring system for predicting early death risk in de novo acute promyelocytic leukemia. Leuk Lymphoma 61(8):1989–1995

Collins GS et al (2015) Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ 350:g7594

Collins GS et al (2024) TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ 385:e078378

Damen JA et al (2016) Prediction models for cardiovascular disease risk in the general population: systematic review. BMJ 353:i2416

de Jong Y et al (2021) Appraising prediction research: a guide and meta-review on bias and applicability assessment using the prediction model risk of bias ASsessment tool (PROBAST). Nephrology 26(12):939–947

Della Porta MG et al (2024) A practical algorithm for acute myeloid leukaemia diagnosis following the updated 2022 classifications. Crit Rev Oncol Hematol 198:104358

DiNardo CD et al (2023) Acute myeloid leukaemia. Lancet 401(10393):2073–2086

Döhner H et al (2022) Diagnosis and management of AML in adults: 2022 recommendations from an international expert panel on behalf of the ELN. Blood 140(12):1345–1377

Efthimiou O et al (2024) Developing clinical prediction models: a step-by-step guide. BMJ 386:e078276

El Chaer F et al (2023) How I treat AML incorporating the updated classifications and guidelines. Blood 141(23):2813–2823

Elhadary M et al (2023) Revolutionizing chronic lymphocytic leukemia diagnosis: a deep dive into the diverse applications of machine learning. Blood Rev 62:101134

Feng Y et al (2023) Characterization of risk prediction models for acute kidney injury: a systematic review and meta-analysis. JAMA Netw Open 6(5):e2313359

Gao X, Liu W (2020) The establishment and evaluation of a new model for the prediction of children B-ALL based on TARGET: a SQUIRE-compliant study. Medicine (Baltimore) 99(19):e20115

González-Gascón YMI et al (2021) From biomarkers to models in the changing landscape of chronic lymphocytic leukemia: evolve or become extinct. Cancers (Basel) 13(8):1782

https://www.healthdata.org/research-analysis/diseases-injuries-risks/factsheets/2021-leukemia-level-3-disease#main-content

Hu Y et al (2023) Prediction models for perineal lacerations during childbirth: a systematic review and critical appraisal. Int J Nurs Stud 145:104546

International CLL-IPI Working Group (2016) An international prognostic index for patients with chronic lymphocytic leukaemia (CLL-IPI): a meta-analysis of individual patient data. Lancet Oncol 17(6):779–790

Kim HT et al (2019) Prognostic score and cytogenetic risk classification for chronic lymphocytic leukemia patients: Center for International Blood and Marrow Transplant Research report. Clin Cancer Res 25(16):5143–5155

Kreuzberger N et al (2020) Prognostic models for newly-diagnosed chronic lymphocytic leukaemia in adults: a systematic review and meta-analysis. Cochrane Database Syst Rev 7(7):Cd012022

Krug U et al (2010) Complete remission and early death after intensive chemotherapy in patients aged 60 years or older with acute

myeloid leukaemia: a web-based application for prediction of outcomes. Lancet 376(9757):2000–2008
Ley C et al (2022) Machine learning and conventional statistics: making sense of the differences. Knee Surg Sports Traumatol Arthrosc 30(3):753–757
Li J et al (2022) Development and validation of a prognostic scoring model to risk stratify childhood acute myeloid leukaemia. Br J Haematol 198(6):1041–1050
Li J et al (2024) Comparison of the effects of imputation methods for missing data in predictive modelling of cohort study datasets. BMC Med Res Methodol 24(1):41
Molica S, Giannarelli D (2021) Prognostic models for chronic lymphocytic leukemia (CLL): a systematic review and meta-analysis. Leukemia 35(2):615–618
Molinaro AM et al (2016) Statistical considerations on prognostic models for glioma. Neuro Oncol 18(5):609–623
Moons KG et al (2012) Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio) marker. Heart 98(9):683–690
Moons KG et al (2014) Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. PLoS Med 11(10):e1001744
Moons KG et al (2015) Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 162(1):W1-73
Odrobina I (2024) Clinical predictive modeling of heart failure: domain description, models' characteristics and literature review. Diagnostics. https://doi.org/10.3390/diagnostics14040443
Pastore F et al (2014) Combined molecular and clinical prognostic index for relapse and survival in cytogenetically normal acute myeloid leukemia. J Clin Oncol 32(15):1586–1594
Patnaik MM, Tefferi A (2022) Chronic myelomonocytic leukemia: 2022 update on diagnosis, risk stratification, and management. Am J Hematol 97(3):352–372
Peduzzi P et al (1995) Importance of events per independent variable in proportional hazards regression analysis. Part II. Accuracy and precision of regression estimates. J Clin Epidemiol 48(12):1503–1510
Pfirrmann M et al (2016) Prognosis of long-term survival considering disease-specific death in patients with chronic myeloid leukemia. Leukemia 30(1):48–56
Phung MT et al (2019) Prognostic models for breast cancer: a systematic review. BMC Cancer 19(1):230
Pollyea DA et al (2023) Acute myeloid leukemia, version 3.2023, NCCN clinical practice guidelines in oncology. J Natl Compr Canc Netw 21(5):503–513
Romijn M et al (2023) Prediction models for bronchopulmonary dysplasia in preterm infants: a systematic review and meta-analysis. J Pediatr 258:113370
Silver SA et al (2015) Risk prediction models for contrast induced nephropathy: systematic review. BMJ 351:h4395
Smit HA et al (2015) Childhood asthma prediction models: a systematic review. Lancet Respir Med 3(12):973–984
Sperrin M et al (2020) Missing data should be handled differently for prediction than for description or causal explanation. J Clin Epidemiol 125:183–187
Tan J et al (2023) Prediction models for depression risk among older adults: systematic review and critical appraisal. Ageing Res Rev 83:101803
van den Boorn HG et al (2018) Prediction models for patients with esophageal or gastric cancer: a systematic review and meta-analysis. PLoS ONE 13(2):e0192310
Walter RB, Estey EH (2020) Selection of initial therapy for newly-diagnosed adult acute myeloid leukemia: limitations of predictive models. Blood Rev 44:100679
Wolff RF et al (2019) Probast: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med 170(1):51–58
Wynants L et al (2015) A simulation study of sample size demonstrated the importance of the number of events per variable to develop prediction models in clustered data. J Clin Epidemiol 68(12):1406–1414
Yun X et al (2020) Recent progress of prognostic biomarkers and risk scoring systems in chronic lymphocytic leukemia. Biomark Res 8:40
Zhang Y et al (2018) Development of a risk grading system to identify patients with acute promyelocytic leukemia at high risk of early death. Cancer Manag Res 10:3619–3627