

SYSTEMATIC REVIEW

OPEN



Machine learning in paediatric haematological malignancies: a systematic review of prognosis, toxicity and treatment response models

Gerard Gurumurthy¹✉, Juditha Gurumurthy² and Samantha Gurumurthy³

© The Author(s) 2024

BACKGROUND: Machine Learning (ML) has demonstrated potential in enhancing care in adult oncology. However, its application in paediatric haematological malignancies is still emerging, necessitating a comprehensive review of its capabilities and limitations in this area.

METHODS: A literature search was conducted through Ovid. Studies included focused on ML models in paediatric patients with haematological malignancies. Studies were categorised into thematic groups for analysis.

RESULTS: Twenty studies, primarily on leukaemia, were included in this review. Studies were organised into thematic categories such as prognoses, treatment responses and toxicity predictions. Prognostic studies showed AUC scores between 0.685 and 0.929, indicating moderate-high predictive accuracy. Treatment response studies demonstrated AUC scores between 0.840 and 0.875, reflecting moderate accuracy. Toxicity prediction studies reported high accuracy with AUC scores from 0.870 to 0.927. Only five studies (25%) performed external validation. Significant heterogeneity was noted in ML tasks, reporting formats, and effect measures across studies, highlighting a lack of standardised reporting and challenges in data comparability.

CONCLUSION: The clinical applicability of these ML models remains limited by the lack of external validation and methodological heterogeneity. Addressing these challenges through standardised reporting and rigorous external validation is needed to translate ML from a promising research tool into a reliable clinical practice component.

Paediatric Research; <https://doi.org/10.1038/s41390-024-03494-9>

IMPACT:

- Key message: Machine Learning (ML) significantly enhances predictive models in paediatric haematological cancers, offering new avenues for personalised treatment strategies. Future research should focus on developing ML models that can integrate with real-time clinical workflows.
- Addition to literature: Provides a comprehensive overview of current ML applications and trends. It identifies limitations to its applicability, including the limited diversity in datasets, which may affect the generalisability of ML models across different populations.
- Impact: Encourages standardisation and external validation in ML studies, aiming to improve patient outcomes through precision medicine in paediatric haematological oncology.

INTRODUCTION

Machine Learning (ML), a subset of artificial intelligence, is capable of identifying complex patterns within large datasets. By leveraging advanced algorithms, ML can facilitate significant advancements in diagnostics, prognostics, and therapeutic decision-making. Despite its potential, the application of ML in healthcare remains largely limited to adult oncology, radiology, and pathology, where it has shown promise in enhancing diagnostic accuracy and treatment planning.^{1–4} However, its utilisation in paediatric haematological malignancies is still in its infant stages,

primarily due to the unique challenges and complexities associated with paediatric cancers.

Paediatric haematological cancers present an area where ML can be beneficially utilised. Children with haematological malignancies exhibit diverse biological behaviours and responses to treatment, necessitating highly individualised therapeutic approaches.⁵ The heterogeneity of these diseases, coupled with the varying responses to existing therapies, underscores the need for a nuanced approach that balances effective treatment with the minimisation of long-term adverse effects.^{6–8} ML, with its ability to

¹School of Medicine, University of Manchester, Manchester, UK. ²School of Cancer and Pharmaceutical Sciences, King's College London, London, UK. ³Department of Infectious Diseases & Immunology, Imperial College London, London, UK. ✉email: gerardgurumurthy@doctors.org.uk

Received: 17 February 2024 Revised: 22 June 2024 Accepted: 5 August 2024

Published online: 31 August 2024

process and analyse vast amounts of data, offers the potential to develop more precise and personalised treatment strategies, thereby improving prognosis and reducing treatment-related toxicity in paediatric patients.

The European Union's Beating Cancer Plan underscores the importance of integrating advanced technologies, including ML, into cancer care.⁹ This initiative aims to exploit the predictive/classification power of ML to enhance cancer prevention, diagnosis, and treatment across Europe. In the context of paediatric haematological malignancies, the potential benefits of ML are particularly significant. The ability to predict disease progression, treatment response, and adverse effects with greater accuracy can transform clinical care, enabling more targeted and effective interventions. It is therefore necessary to address the current limitations of ML, including the need for diverse and representative datasets, standardised reporting, and rigorous external validation. This systematic review aims to provide a comprehensive overview of the current applications of ML in paediatric haematological malignancies, assessing its potential to enhance diagnostic accuracy, prognostic predictions, and treatment strategies.

METHODS

This review was conducted in accordance with the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines and was registered with PROSPERO (CRD42024507811). A comprehensive systematic review of the literature was carried out in February 2024 using the OVID platform. The databases searched included AMED, EMBASE, MEDLINE, and Emcare. The detailed search strategy is provided in Supplementary Table S1.

Inclusion and exclusion criteria

To be included in the review, studies had to focus on the application of ML in paediatric haematological cancers, detailing the type of ML model and its methodology. Only original research articles were considered; reviews, case reports, and other non-original research articles were excluded. Additionally, only studies exclusively involving paediatric populations were included; mixed studies with both paediatric and adult cohorts were excluded. Articles were limited to those published in English.

Data extraction and synthesis

Analysis of records were conducted by two authors independently. Data extracted from each study included the specific type of haematological cancer investigated, the tasks performed by the ML program, the number of patients involved, the ML method employed, input and output variables, the method of cross-validation used, and any external validation performed. Studies were then grouped based on their primary objectives or outcomes related to paediatric haematological cancers, such as prognosis, treatment response, and toxicity models. This thematic grouping facilitated a narrative synthesis to highlight trends, patterns, and gaps in the current research. A minimum of three studies was required to synthesise a theme, ensuring sufficient data to capture the scope and trends of current research efforts.

Quality assessment

The quality of the studies was assessed using appropriate tools. For studies investigating prognostic ML models, the Quality in Prognosis Studies (QUIPS)¹⁰ checklist was utilised. For all other thematic groups, the Newcastle-Ottawa Scale (NOS)¹¹ was used to assess the quality, given the observational nature of the included studies.

Analysis

Due to the varying nature of the ML tasks, lack of uniform reporting formats, and diverse effect measures, formal meta-analyses were deemed unfeasible. Instead, heterogeneity was addressed qualitatively by describing differences in study populations, methodologies, outcomes, and effect measures.

RESULTS

Searches conducted through the available databases in Ovid yielded a total of 711 results (Fig. 1). Of which, 20 studies that applied ML in paediatric haematological malignancies met the inclusion criteria for this review (Table 1).^{12–31}

The included studies primarily focused on leukaemia, with specific emphasis on acute lymphoblastic leukaemia (ALL) in 13 studies, acute myeloid leukaemia (AML) in four studies, and an unspecified subtype in 1 study. Additionally, two studies addressed lymphoma. The most commonly used ML methods and algorithms were Random Forest (RF, $n = 8$), Least Absolute Shrinkage and Selection Operator (LASSO, $n = 6$), Gradient Boosting Model (GBM, $n = 4$), and Support Vector Machines (SVM, $n = 4$). Note that multiple papers utilised more than one ML method in a single study.

Cross-validation techniques were employed in 16 (80%) of the studies, including methods such as 5- to 1000-fold cross-validation, leave-one-out cross-validation, training-versus-testing sets, and C-index calculations. External validation was performed in 5 (25%) of all included studies.

Prognosis and relapse/recurrence studies

Eight studies utilised ML to predict disease outcomes in paediatric haematological malignancies.^{12–19} All studies focused on leukaemia, with five addressing ALL, two on AML, and one on an unspecified subtype. All studies were assessed as “low risk” of bias using the Quality in Prognosis Studies (QUIPS) checklist, which evaluates study participation, attrition, prognostic factor measurement, outcome measurement, study confounding, and statistical analysis,¹⁰ indicating good overall study quality in this category.

Survival analyses employing genetic data from databases such as TARGET were the most common methodology in this group. For example, one study identified key long non-coding RNAs (lncRNAs) associated with AML prognosis using LASSO Cox analysis, reporting Area Under the Curve (AUC) values of 0.701, 0.704, and 0.696 for 1-, 3-, and 5-year survival, respectively.¹⁵ AUC values below 0.50 indicate poor predictability, values between 0.51 and 0.70 indicate relatively poor accuracy, values between 0.71 and 0.90 indicate moderate accuracy, and values above 0.90 indicate high accuracy and strong discrimination capability.³² Notably, these findings were externally validated using comparative data from The Cancer Genome Atlas (TCGA), although with a lower concordance to the developed model. Overall, four (50%) of the studies in this category were externally validated. For instance, one study using the RF algorithm with 10-fold Monte Carlo cross-validation to predict relapse in ALL achieved an AUC of 0.901. The results were externally validated against an independent test set of 84 patients, demonstrating the robustness and potential clinical applicability of the predictive model. The use of external validation suggests a strong reinforcement of the predictive models' robustness and applicability in clinical settings. This process, such as using data from TCGA, underscores the potential of these models to generalise across different datasets, enhancing their reliability for clinical prognosis and treatment decision-making in paediatric leukaemia cases.

The studies showed a range of AUC scores from 0.685 to 0.929, indicating a wide variation in model performance. This heterogeneity could be attributed to differences in study design, including varying numbers of patients (range 156–1693) and primary endpoints (e.g., 3-year overall survival vs. 5-year overall survival). Seven (88%) of the studies used AUC as a primary

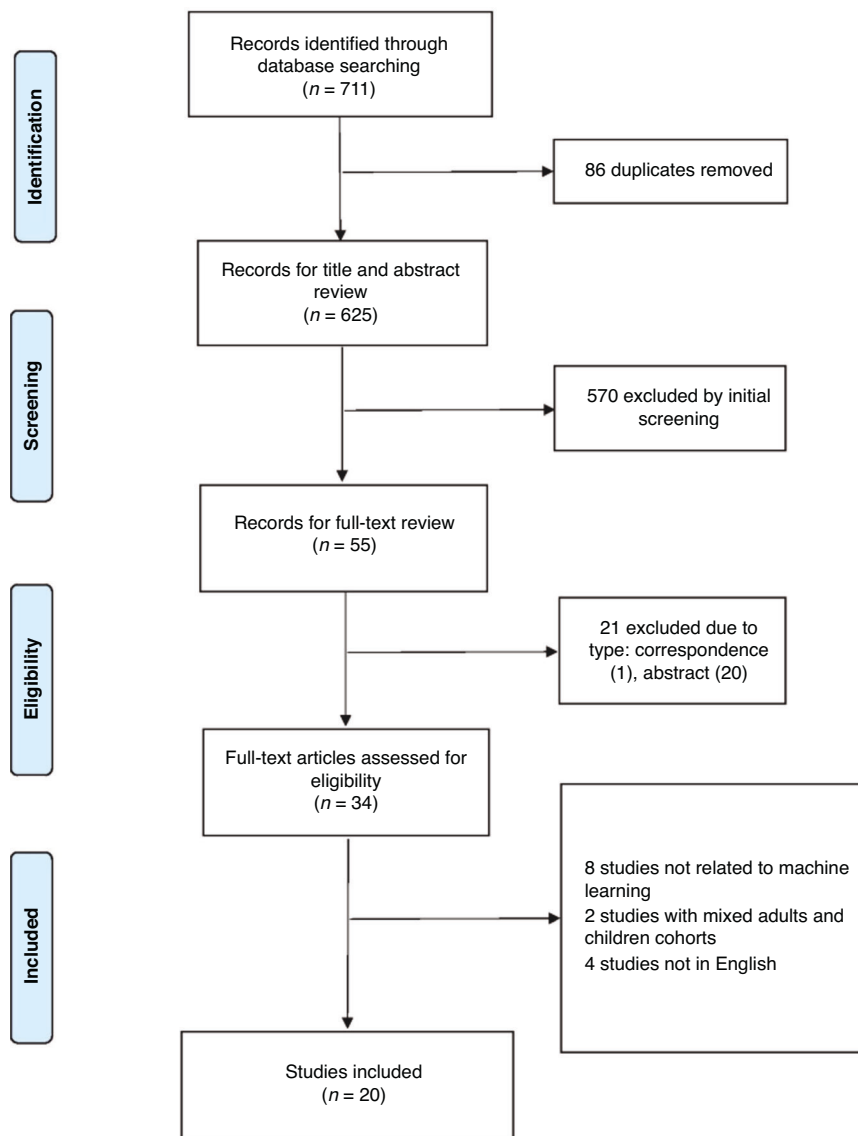


Fig. 1 Flow diagram of studies shortlisted for inclusion in this review from the initial search results ($n = 20$).

measure of predictive performance. All studies used either LASSO ($n = 4$) or RF ($n = 3$) methods. When grouped by ML method, LASSO models had AUC scores ranging from 0.685 to 0.898, indicating low to moderate accuracy, while RF models had AUC scores ranging from 0.803 to 0.929, indicating moderate to high accuracy. These results suggest that RF techniques may offer marginally superior predictive performance compared to LASSO.

Despite the promise shown by these models, limitations include the use of genetic data from publicly available databases and a lack of relevant paediatric cohort validation. One group of authors highlighted the need for future research to employ more prospective paediatric cohorts due to the limitations associated with using public databases.¹⁵

In summary, these studies highlight the significant potential of ML methods, particularly RF and LASSO, in predicting disease outcomes in paediatric leukaemia. The variation in AUC scores underscores the importance of strategic ML method selection, reflecting its role in study outcome heterogeneity. These findings highlight the need for a nuanced approach in selecting ML techniques, considering not only AUC scores but also factors like model interpretability and computational demands, to enhance predictive precision in leukaemia prognosis.

Treatment response studies

Five studies investigated the use of ML to predict treatment response in paediatric haematological malignancies, including three studies on ALL and two on AML.^{20–24} Four of these studies focused on classification tasks. All studies scored 6 or more on the Newcastle-Ottawa Scale (NOS), indicating a generally high standard of methodological quality and reliability in their findings.

In one study, a ten-gene DNA-damage response gene expression signature (CalDDR-GEx10 score) was used to predict responses to gemtuzumab ozogamicin (GO) in paediatric AML patients. The input variables included gene expression levels of 18 genes in DNA-damage response pathways. Patients with high CalDDR-GEx10 scores had lower complete remission (CR) rates and worse event-free survival when treated with GO. This score specifically predicted responses to calicheamicin-induced DNA damage, rather than general chemotherapy effects, with a sensitivity of 72.7%, specificity of 63.6%, and a Positive Predictive Value (PPV) of 61.1%.²⁰ Another study employed ML techniques, including k-nearest neighbours (K-NN), SVM, and RF, to RNA sequencing data to predict CR in paediatric AML patients post-induction therapy.²² The best result, achieved using a K-NN model with 50 genes, yielded an AUC of 0.812. Both studies were able to

Table 1. Characteristics of Included Studies Exploring Machine Learning in Paediatric Haematological Malignancies.

Study	Cancer Type	Task/ Problem	Input Variable	Output Variable	ML Method	No. of Patients	Cross validation	External Validation	Highest AUC	Other Comparative Statistical Scores Used
Prognosis and Relapse/ Recurrence Studies										
He et al. ¹²	AML	Prognostic prediction	Expression levels of pyroptosis-related genes	Risk score	LASSO	NA	Not specified	E-MTAB-1216 dataset	0.893	HR = 2.04
He et al. ¹³	ALL	Survival prediction	Clinical characteristics, immunophenotype, genetic data	EFS predictive model	LASSO	1693	10-fold CV	Not specified	0.822	C-index = 0.81
Cui et al. ¹⁴	Leukaemia (sub-type not specified)	Survival prediction	Clinical characteristics	Survival predictive model	Bayesian inference	17539	Not specified	Not specified		C-index = 0.93
Zheng et al. ¹⁵	AML	Prognostic prediction	Gene expression data of m6A-related lncRNAs	OS predictive model	LASSO	646	Not specified	TCGA database	0.685	C-index = 0.82
Bohannon et al. ¹⁶	ALL	Survival prediction	Genomic data	EFS predictive model	RF	156	Training (70%) and testing (30%) cohorts used	Not specified	0.929	HR = 5.41 C-index = 0.82
Gao et al. ¹⁷	B-cell ALL	Survival prediction	Clinical characteristics	OS predictive model	LASSO	1316	C-index	TARGET database	0.898	C-index = 0.87
Lin et al. ¹⁸	B-cell ALL	Relapse prediction	NAD+ metabolism-related genes	Relapse predictive model	RF	NA	Mentioned but not specified	Not specified	0.8031	
Pan et al. ¹⁹	ALL	Relapse prediction	Sociodemographic, clinical, immunological, and cytogenetic data	Relapse predictive model	RF, SVM, LR & DT	570	10-fold CV	Independent test set of 84 patients	0.904	Accuracy = 82.9%
Treatment Response Studies										
Gbadamosi et al. ²⁰	AML	GO-related response prediction	Genetic data	Treatment response outcome model	LASSO	301	1000-fold CV	Not specified		OS = 0.676 HR = 0.565
Pedreira et al. ²¹	ALL	Treatment intensity decision support model	Clinical data	Treatment decision model	NN	158	Leave-one-out	Not specified		RHR = 98% ROR = 21% RSR = 0%
Gal et al. ²²	AML	Complete remission prediction	Gene expression data	Complete remission predictive model	K-NN, SVM & RF	473	5-fold CV	Not specified	0.840	
Kashef et al. ²³	ALL	Treatment prediction	Clinical data	Complete remission predictive model	GBM, RF, GLM	241	5-fold CV	Not specified	0.8725	
Kashef et al. ²⁴	ALL	Treatment outcome classification	Clinical characteristics & treatment related toxicity	Classification of treatment outcomes model	DT, SVM, RF, LDA, MLR, GBM,	241	10-fold CV	Not specified	0.870	Accuracy = 94.9%

Table 1. continued

Study	Cancer Type	Task/ Problem	Input Variable	Output Variable	ML Method	No. of Patients	Cross validation	External Validation	Highest AUC	Other Comparative Statistical Scores Used
Treatment Toxicity Studies										
Al-Fahad et al. ²⁵	ALL	Treatment toxicity prediction	MRI-derived information	Classification of cognitive abilities	LASSO	200	Training (80%) and testing (20%) cohorts used	Not specified	0.870	
Ramalingam et al. ²⁶	ALL	Treatment toxicity prediction	Genotypes of SLC19A1, MTHFR, TYMS, and cytogenetic data	Methotrexate-related toxicities	MDR	115	10-fold CV	Not specified		OR = 5.71 -2 Log Likelihood of Reduced Model = 97.104
Zhan et al. ²⁷	B-cell ALL	Treatment toxicity prediction	SNPs in 16 genes, clinical characteristics and methotrexate delayed clearance	Predictive models for the risk of neutropenia and fever	RF with ADASYN SVM, DT	139	Training (70%) and testing (30%) cohorts used	Not specified	0.927	
Tram et al. ²⁸	Lymphoma	Treatment toxicity prediction	CT image	Risk of treatment-related late effects	NN	100	5-fold CV	Against human raters		Dice value = 0.988 HR = 3.1
Theruvath et al. ²⁹	Lymphoma	Dosing prediction	PET/MRI scan data	Enhanced PET/MRI images	NN	20	Not specified	Against neural network called SubtlePET	1	K statistic = 1
Disease Susceptibility/ Diagnosis Studies										
Mahmood et al. ³⁰	ALL	Classifying risk factors	Clinical, genomic and socio-environmental data	Risk score for ALL	CART, RF, GBM & DT	50	10-fold CV	Not specified		Accuracy = 99.83%
Kulis et al. ³¹	B-Cell Precursor ALL	Classifying risk factors	Antigens measured through flow cytometry	Identification of specific genetic aberrations	GBM	818	5-fold CV	Not specified		OR = 16.90

ADASYN Adaptive Synthetic, ALL Acute Lymphoblastic Leukaemia, AML Acute Myeloid Leukaemia, AUC Area Under the Curve, CART Classification and Regression Tree, CV Cross-validation, DT Decision Tree, GBM Gradient Boosting Model, GLM Generalised Linear Model, GO gemtuzumab ozogamicin, HR Hazard Ratio, LncRNA Long noncoding RNA, K-NN k-nearest neighbour, LASSO Least Absolute Shrinkage and Selection Operator, LDA Linear Discriminant Analysis, LR Linear Regression, MDR Multifactor Dimensionality Reduction, NN Neural Network, OR Odds Ratio, RF Random Forest, RHR Rate of High Risk, ROR Rate of Overestimated Risk, RSR Rate of Subestimation Risk, SNP Single Nucleotide Polymorphism, SVM Support Vector Machine

predict CR based on genetic data through the utilisation of ML and were cross-validated, highlighting the potential of ML and gene expression signatures in personalised medicine for cancer treatment.

Three (60%) of the studies in this category used AUC as a measure of their models' ability to predict treatment response, with scores ranging from 0.840 to 0.875, indicating moderate accuracy. Despite the use of different ML algorithms (GBM, K-NN, and Decision Tree), the studies showed similar patient sizes (range 241–473) and endpoints, contributing to low heterogeneity in the evaluation of treatment response prediction. This consistency suggests a reliable evaluation of treatment response prediction across these studies.

However, none of these studies achieved a high accuracy AUC model (>0.900), indicating that while the models were moderately effective, they did not reach the threshold of high accuracy. Additionally, none of the studies conducted external validation, which limits the clinical utility of these models. Prospective studies with external validation are needed to assess the impact of these ML models on treatment decision-making and patient outcomes. Despite these limitations, the findings support the potential of ML to enhance personalised medicine in this field.

Treatment toxicity studies

ML was used to predict adverse treatment effects in five studies.^{25–29} Three studies focused on ALL and two on lymphoma. The Newcastle-Ottawa Scale (NOS) was used to assess the quality of these studies, with all scoring six or more, indicating high methodological quality.

One study explored the relationship between genetic variations and treatment-related adverse effects (TRAEs) in paediatric patients with ALL undergoing methotrexate therapy. It found a significant association between the SLC19A1 (c.80 G > A) genotype and increased TRAEs, with an odds ratio (OR) of 5.71 ($p < 0.01$). Multinomial logistic regression and multifactor dimensionality reduction analysis supported this association, confirming the genotype's strong correlation with TRAEs.²⁶ Another study also focused on methotrexate therapy, using ML to predict neutropenia and fever associated with high-dose methotrexate treatment in paediatric B-ALL. The best model, using a combined RF with Adaptive Synthetic (ADASYN) resampling, achieved an AUC of 0.870 to 0.927, sensitivity of 0.916–0.935, and specificity of 0.920–0.924.²⁷

In another study, CT images were used to predict late TRAEs. A deep learning model demonstrated high concordance with manual human analysis, evidenced by Dice scores greater than 0.950 and a K-statistic of 1.00. Notably, once trained, the model segmented body composition from CT datasets in under a second, highlighting the potential of ML models to rapidly and accurately process extensive datasets. Validated against external manual analysis, this model shows promise for clinical application due to its capability to deliver rapid and reliable results.²⁸

The studies varied widely in their statistical analyses, making it difficult to comment on heterogeneity. Only two studies used AUC as a measure of effect. These AUC values were 0.870 (moderate accuracy) and 0.927 (high accuracy), suggesting strong predictive capabilities of ML models in this context.

The primary limitation of these studies is the lack of uniform reporting of effect measures, which hampers the ability to review heterogeneity and draw robust conclusions. Additionally, the sample sizes in these studies (range 20 to 200) were smaller compared to other categories, limiting the statistical power to detect significant associations. Moreover, translating these findings into clinical practice requires validation in larger, multi-centre studies to confirm their utility in predicting treatment-related toxicities. Only one of these studies included external validation, underscoring the need for further validation efforts.

Others: disease susceptibility & diagnosis studies

This review identified two studies focused on developing predictive models for disease susceptibility and diagnostics in paediatric haematological malignancies.^{30,31} These studies provide insights into the early application of ML in identifying risk factors and diagnostic markers.

One study employed several ML algorithms, including Classification and Regression Tree (CART), RF, GBM, and C5.0 decision tree, to identify key attributes influencing ALL susceptibility.³⁰ Platelet count was identified as a crucial predictor, and the CART algorithm demonstrated a high model accuracy of 99.8%. However, this study lacked external validation, which limits the generalisability of the findings and highlights the need for further investigation in more varied and larger cohorts.

Similarly, the second study within this group also utilised ML models for disease susceptibility but did not perform external validation. The lack of validation is a significant limitation as it prevents the confirmation of the models' applicability in different clinical settings. Despite this, the preliminary findings suggest that ML can identify important predictive factors for disease susceptibility.

Overall, while the limited number of studies prevents a comprehensive thematic analysis, these findings indicate the potential of ML in enhancing early disease detection and risk assessment in paediatric haematological cancers. The absence of external validation across both studies underscores the need for further research to ensure the reliability and practical utility of these ML models.

DISCUSSION

The review reveals a promising trend of ML models achieving moderate to high accuracy across the examined thematic categories. ML methods such as RF and LASSO have emerged as effective tools in paediatric haematological malignancies, as reflected in their prevalence across the studies reviewed. These studies demonstrate a strong emphasis on predictive tasks, highlighting a growing interest in using ML for prognosis and treatment outcome prediction. Most research thus far lies in prognosis models, with further research warranted in diagnosis and treatment toxicity prediction models. An adequate number of studies exist in treatment response studies. It is crucial, however, to assess the real-world applicability of these findings through external validation, considering the diverse methodologies and sample sizes across studies.^{33,34}

The lack of external validation in many studies is a significant limitation that prevents the replication and generalisation of ML models across different datasets. The heterogeneity of the data collected, including variations in patient populations, data sources, and ML methodologies, complicates the replication of these models. To address this issue, future studies should focus on standardising data collection methods and reporting metrics. The need for standardised reporting guidelines, such as the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD),³⁵ is highlighted in a similar review of the literature.³⁶ Moreover, the limited clinical deployment of ML algorithms, with many studies showing limited clinical applicability, is a common criticism.^{37–39}

This review also illustrates the infancy of ML application in this field, marked by the limited number of studies included in most thematic categories identified. The initial literature search yielded over 20 abstract reports that, despite being excluded due to not meeting inclusion criteria, indicate a growing application of ML in paediatric haematology with vast use across the field. For example, a study using Prediction Analysis of Microarrays (PAM) to identify paediatric patients with B-ALL with a Ph-like signature for better clinical intervention employed ML on gene expression profiles from 811 patients, leading to a 15-gene classifier that showed high sensitivity (93.0%) and specificity (89.7%) in tests.⁴⁰

The classifier was also able to identify genomic lesions linked to Ph-like ALL, associated with poor clinical outcomes. The findings suggest that integrating this classifier in clinical practice could help identify patients for targeted therapy, potentially improving treatment outcomes. The numerous abstracts noted in the literature search highlight the rapidly growing application of ML in paediatric haematological cancers. As these datasets grow, they offer new opportunities for applying novel ML approaches, potentially transforming the field.

The review highlights the potential of ML to enhance patient care by providing clinicians and health professionals with data-driven insights that can inform diagnostic and treatment decisions. While promising, the integration of ML in clinical practice should support, not replace, healthcare providers. For instance, ML algorithms have achieved 97.0% accuracy in identifying leukaemia from peripheral blood smears, thereby supporting clinical investigations.⁴¹ Furthermore, the integration of ML-enhanced technologies such as the Countess 3 Automated Cell Counter in bone marrow transplant labs exemplifies a shift toward more precise and efficient diagnostic processes, showing notable improvements over traditional manual methods.^{42,43} This trend is supported by a previous reporting of methodology that calls for strong evaluation frameworks to measure the actual impact of ML on patient outcomes, thus ensuring its role as a complement to, not a replacement for, clinical decision-making.⁴⁴

Comparing findings between paediatric and adult/mixed studies reveals key insights. In adult/mixed cohorts, ML applications show significant improvements with AUC ranges of 0.71 – 0.93 for prognosis/relapse prediction,^{45,46} 0.85 – 0.97 for treatment response,^{22,47,48} and 0.59 – 0.90 in toxicity predictions.^{49,50} These studies have a superior AUC for prognosis/relapse and treatment response predictions as compared to the paediatric cohort. The superior AUC in adult models highlights their robustness, likely due to larger sample sizes and more extensive datasets. Adult studies may also benefit from standardised methodologies and larger, diverse cohorts, contributing to increased generalisability. Conversely, paediatric studies face challenges such as smaller sample sizes and heterogeneous designs, leading to broader AUC ranges and reduced generalisability.

The scope of this review is narrowed by the predominance of studies focused on leukaemia, specifically ALL and AML, with only two studies extending to non-leukemic haematological malignancies. This lack of diversity within the spectrum of paediatric haematological cancers limits our capacity to generalise the findings of ML across the broader field. Consequently, while our review suggests substantial advancements in the ML-driven management of leukaemia, the translatability of these insights to other haematological conditions remains to be ascertained. With only a small number of studies employing external validations, we are unable to comment on the feasibility of implementing these ML algorithms in the current clinical setting. This underscores an imperative for future research to encompass a wider range of haematological disorders, thus enhancing the robustness and clinical relevance of ML prognostic, diagnostic, and treatment response models in paediatric haematology.

Addressing these challenges of methodological heterogeneity and limited clinical deployment is crucial for the implementation of ML in paediatric malignancies.⁵¹ The expanding datasets in this domain offer an opportunity for applying novel ML approaches. However, increased standardisation in study designs and reporting standards, like the TRIPOD guidelines mentioned above, is essential to achieve this. Future research should focus on prospective studies and fostering interdisciplinary collaboration to develop and implement clinically relevant ML tools. Moreover, integrating ML with clinical workflows and validating these models in diverse, real-world settings will be vital in ensuring their practical utility and improving outcomes for children with cancer.

CONCLUSION

This systematic review highlights the growing role of ML in paediatric haematological malignancies, demonstrating its potential to significantly enhance diagnostic accuracy, prognostic predictions, and treatment strategies. Despite moderate to high accuracy achieved by ML models, the clinical applicability remains constrained due to the lack of external validation and methodological heterogeneity. Addressing these challenges through larger, diverse datasets, standardised reporting, and robust external validation is crucial for translating ML from a promising research tool into a reliable component of clinical practice. This advancement could lead to more precise and personalised treatment approaches, ultimately improving outcomes for children with cancer.

REFERENCES

- Bertsimas, D. et al. Machine learning in oncology: methods, applications, and challenges. *JCO Clin. Cancer Inf.* **4**, 885–894 (2020).
- Nardini, C. Machine learning in oncology: a review. *Ecancermedicalscience* **14**, 1065 (2020). Published 2020 Jun 30.
- Nagy, M. et al. Machine learning in oncology: what should clinicians know? *JCO Clin. Cancer Inf.* **4**, 799–810 (2020).
- Hong, J. et al. System for high-intensity evaluation during radiation therapy (SHIELD-RT): A prospective randomized study of machine learning-directed clinical evaluations during radiation and chemoradiation. *J. Clin. Oncol.* **38**, 3652–3661 (2020).
- Widemann, B. Advances, challenges and progress in pediatric hematology and oncology. *Curr. Opin. Pediatr.* **35**, 39–40 (2023).
- Rosenquist Ret al. Novel precision medicine approaches and treatment strategies in hematological malignancies. *J. Intern. Med.* **294**, 413–436 (2023).
- Duncavage, E. et al. Genomic profiling for clinical decision making in myeloid neoplasms and acute leukemia. *Blood* **140**, 2228–2247 (2022).
- Khoury, J. et al. The 5th edition of the World Health Organization classification of haematolymphoid tumours: myeloid and histiocytic/dendritic neoplasms. *Leukemia* **36**, 1703–1719 (2022).
- European Commission. Europe's Beating Cancer Plan 2023. 2024. https://health.ec.europa.eu/system/files/2022-02/eu_cancer-plan_en_0.pdf.
- Cochrane Method. QUIPS tool. 2024. <https://methods.cochrane.org/sites/methods.cochrane.org/prognosis/files/uploads/QUIPS%20tool.pdf>.
- Ottawa Hospital Research Institute. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. 2024. https://www.ohri.ca/programs/clinical_epidemiology/oxford.asp.
- He, X. et al. A gene signature comprising seven pyroptosis-related genes predicts prognosis in pediatric patients with acute myeloid leukemia. *Acta Haematol.* **145**, 627–641 (2022).
- He, Y. et al. A nomogram for predicting event-free survival in childhood acute lymphoblastic leukemia: a multicenter retrospective study. *Front. Oncol.* **12**, 854798 (2022).
- Cui, Y. et al. Bayesian inference for survival prediction of childhood Leukemia. *Comput. Biol. Med.* **156**, 106713 (2023).
- Zheng, G. et al. Comprehensive analysis of N6-methyladenosine-related long noncoding RNA prognosis of acute myeloid leukemia and immune cell infiltration. *Front. Genet.* **13**, 888173 (2022).
- Bohannon, Z., Coffman, F. & Mitrofanova, A. Random survival forest model identifies novel biomarkers of event-free survival in high-risk pediatric acute lymphoblastic leukemia. *Comput. Struct. Biotechnol. J.* **20**, 583–597 (2022).
- Gao, X. & Liu, W. The establishment and evaluation of a new model for the prediction of Children B-ALL based on TARGET: A SQUIRE-compliant study. *Med. (Baltim.)* **99**, e20115 (2020).
- Lin, C. et al. Integrating RNA-seq and scRNA-seq to explore the biological significance of NAD⁺ metabolism-related genes in the initial diagnosis and relapse of childhood B-cell acute lymphoblastic leukemia. *Front. Immunol.* **13**, 1043111 (2022).
- Pan, L. et al. Machine learning applications for prediction of relapse in childhood acute lymphoblastic leukemia. *Sci. Rep.* **7**, 7402 (2017).
- Gbadamosi, M. et al. A ten-gene DNA-damage response pathway gene expression signature predicts gemtuzumab ozogamicin response in pediatric AML patients treated on COGAAML0531 and AAML03P1 trials. *Leukemia* **36**, 2022–2031 (2022).
- Pedreira, C. et al. New decision support tool for treatment intensity choice in childhood acute lymphoblastic leukemia. *IEEE Trans. Inf. Technol. Biomed.* **13**, 284–290 (2009).

22. Gal, O., Auslander, N., Fan, Y. & Meerzaman, D. Predicting complete remission of acute myeloid leukemia: machine learning applied to gene expression. *Cancer Inf.* **18**, 1176935119835544 (2019).
23. Kashef, A., Khatibi, T. & Mehrvar, A. Prediction of cranial radiotherapy treatment in pediatric acute lymphoblastic leukemia patients using machine learning: a case study at MAHAK Hospital. *Asian Pac. J. Cancer Prev.* **21**, 3211–3219 (2020).
24. Kashef, A., Khatibi, T. & Mehrvar, A. Treatment outcome classification of pediatric acute lymphoblastic leukemia patients with clinical and medical data using machine learning: a case study at MAHAK Hospital. *Inform. Med. Unlocked* **20**, 100399 (2020).
25. Al-Fahad, R. et al. Early imaging based predictive modeling of cognitive performance following therapy for childhood ALL. *IEEE Access* **7**, 146662–146674 (2019).
26. Ramalingam, R. et al. Evaluation of cytogenetic and molecular markers with MTX-mediated toxicity in pediatric acute lymphoblastic leukemia patients. *Cancer Chemother. Pharmacol.* **89**, 393–400 (2022).
27. Zhan, M. et al. Machine learning to predict high-dose methotrexate-related neutropenia and fever in children with B-cell acute lymphoblastic leukemia. *Leuk. Lymphoma* **62**, 2502–2513 (2021).
28. Tram, N. et al. Deep learning of image-derived measures of body composition in pediatric, adolescent, and young adult lymphoma: association with late treatment effects. *Eur. Radiol.* **33**, 6599–6607 (2023).
29. Theruvath, A. et al. Validation of deep learning-based augmentation for reduced 18F-FDG Dose for PET/MRI in children and young adults with lymphoma. *Radiol. Artif. Intell.* **3**, e200232 (2021).
30. Mahmood, N. et al. Identification of significant risks in pediatric acute lymphoblastic leukemia (ALL) through machine learning (ML) approach. *Med. Biol. Eng. Comput.* **58**, 2631–2640 (2020).
31. Kulis, J. et al. Machine learning based analysis of relations between antigen expression and genetic aberrations in childhood B-cell precursor acute lymphoblastic leukaemia. *J. Clin. Med.* **11**, 2281 (2022).
32. Mandrekar, J. Receiver operating characteristic curve in diagnostic test assessment. *J. Thorac. Oncol.* **5**, 1315–1316 (2010).
33. Stuart, E., Bradshaw, C. & Leaf, P. Assessing the generalizability of randomized trial results to target populations. *Prev. Sci.* **16**, 475–485 (2015).
34. Khorsan, R. & Crawford, C. How to assess the external validity and model validity of therapeutic trials: a conceptual approach to systematic review methodology. *Evid. Based Complement Altern. Med.* **2014**, 694804 (2014).
35. Collins, G. & Moons, K. Reporting of artificial intelligence prediction models. *Lancet* **393**, 1577–1579 (2019).
36. Ramesh, S. et al. Applications of artificial intelligence in pediatric oncology: a systematic review. *JCO Clin. Cancer Inform.* **5**, 1208–1219 (2021).
37. Kelly, C., Karthikesalingam, A., Suleyman, M., Corrado, G. & King, D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med.* **17**, 195 (2019).
38. Ghassemi, M. et al. A review of challenges and opportunities in machine learning for health. *AMIA Jt. Summits Transl. Sci. Proc.* **2020**, 191–200 (2020).
39. Khan, B. et al. Drawbacks of Artificial Intelligence and Their Potential Solutions in the Healthcare Sector. *Biomed. Mater. Devices*. 1–8 <https://doi.org/10.1007/s44174-023-00063-2> (2023).
40. Harvey, R. et al. Development and validation of a highly sensitive and specific gene expression classifier to prospectively screen and identify B-precursor acute lymphoblastic leukemia (ALL) patients with a Philadelphia chromosome-like ('Ph-like') or 'Bcr-Abl1-like' signature for therapeutic targeting and clinical intervention. *Blood* **122**, 826 (2013).
41. Ghaderzadeh, M. et al. Machine learning in detection and classification of leukemia using smear blood images: a systematic review. *Sci. Prog.* **2021**, 9933481:1–14 (2021).
42. Lavitt, F., Rijlaarsdam, D., van der Linden, D., Weglarz-Tomczak, E. & Tomczak, J. Deep learning and transfer learning for automatic cell counting in microscope images of human cancer cell lines. *Appl. Sci.* **11**, 4912 (2021).
43. Lee, S.-J., Chen, P.-Y. & Lin, J.-W. Complete blood cell detection and counting based on deep neural networks. *Appl. Sci.* **12**, 8140 (2022).
44. Verma, A. et al. Grand rounds in methodology: key considerations for implementing machine learning solutions in quality improvement initiatives. *BMJ Qual. Saf.* **33**, 121–131 (2024).
45. Karami, K., Akbari, M., Moradi, M. T., Soleymani, B. & Fallahi, H. Survival prognostic factors in patients with acute myeloid leukemia using machine learning techniques. *PLoS One* **16**, e0254976 (2021).
46. Eckardt, J. et al. Prediction of complete remission and survival in acute myeloid leukemia using supervised machine learning. *Haematologica* **108**, 690–704 (2023). Published 2023 Mar 1.
47. Tong, Y. et al. Prediction of lymphoma response to CAR T cells by deep learning-based image analysis. *PLoS One* **18**, e0282573 (2023).
48. Elhadary, M. et al. Applications of machine learning in chronic myeloid leukemia. *Diagnostics* **13**, 1330 (2023).
49. Jian, C. et al. Predicting delayed methotrexate elimination in pediatric acute lymphoblastic leukemia patients: an innovative web-based machine learning tool developed through a multicenter, retrospective analysis. *BMC Med Inf. Decis. Mak.* **23**, 148 (2023).
50. Arai, Y. et al. Using a machine learning algorithm to predict acute graft-versus-host disease following allogeneic transplantation. *Blood Adv.* **3**, 3626–3634 (2019).
51. Tozzi, A. et al. Gaps and opportunities of artificial intelligence applications for pediatric oncology in european research: a systematic review of reviews and a bibliometric analysis. *Front. Oncol.* **12**, 905770 (2022).

AUTHOR CONTRIBUTIONS

GG, JG and SG were involved in the literature search and subsequent data extraction from the included studies. The initial draft of the manuscript was prepared by GG and JG, with SG providing further review and editing. All authors have given final approval for the version to be published.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41390-024-03494-9>.

Correspondence and requests for materials should be addressed to Gerard Gurumurthy.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024