Interdisciplinary Research in Medical Sciences Specialty
ISSN: 3062-4401
2026, Volume 6, Issue 1, Page No: 255-270

Available online at: www.galaxypub.co/page/journals

Galaxy Publication

Progression of Prognostic Prediction in Pediatric Cancer from Conventional Factors to Artificial Intelligence: A Systematic Review and Meta-Analysis

Sara Ben Youssef[1]*, Amal Trabelsi[1], Karim Boudiaf[2], Nabil Jebali[3]

[1]Department of Pediatric Oncology and Prognostic Modeling, Faculty of Medicine, University of Tunis, Tunis, Tunisia.
[2]Department of Systematic Review and Meta-Analysis, Faculty of Medicine, University of Sousse, Sousse, Tunisia.
[3]Department of AI in Pediatric Cancer Prediction, Faculty of Medicine, University of Sfax, Sfax, Tunisia.

***E-mail** ✉ sara.benyoussef@inat.ucar.tn

Received: 27 March 2026; **Revised:** 04 June 2026; **Accepted:** 06 June 2026

ABSTRACT

Modern paediatric cancer management needs improved methods for forecasting patient outcomes to enable tailored risk grouping. However, research evaluating the effectiveness, makeup, and shortcomings of current prognostic tools remains limited. This study sought to evaluate the predictive performance of conventional and more sophisticated prognostic approaches. This systematic review and meta-analysis (registered as CRTN42022370251) involved a comprehensive literature search across PubMed, Embase, Scopus, and the Cochrane Library, completed on 28 June 2024. Eligible investigations examined the precision of prognostic indicators or algorithms in paediatric blood cancers, central nervous system (CNS) tumours, or non-CNS solid tumours (NCNSST). Three main model types were examined: those based on (1) clinical variables, (2) genomic and transcriptomic information, and (3) artificial intelligence (AI) techniques. Key endpoints included the area under the receiver operating characteristic curve (AUROC) with 95% confidence intervals (CI) for different overall survival (OS) periods and event-free survival (EFS). Study selection and data gathering were carried out independently by two teams. Analyses drew on published results and open-access repositories [CRTN42022370251]. From 12,982 screened records, 358 studies contributed to the meta-analysis and 27 to the systematic review, though information on AI methods was scarce. The largest body of evidence concerned NCNSST patients at the 5-year OS mark. Here, Category-1 models achieved a pooled AUROC of 0.75 (CI: 0.72–0.79), which was significantly lower than Category-2 models at 0.85 (CI: 0.82–0.88) ($p < 0.001$). No significant difference emerged between Category-2 and Category-3 models ($p = 0.2834$; pooled AUROC 0.82, CI: 0.77–0.88). Investigations relying on internal validation demonstrated markedly higher performance than those with external validation, underscoring elevated risk of bias (RoB) associated with internal approaches. The most frequent RoB concerns appeared in the outcome measurement and statistical analysis domains, evaluated via PROBAST and QUIPS tools. Incorporating Category-2 and Category-3 models into routine practice is recommended, particularly for NCNSST cases to support better risk classification. Although AI-based forecasting in childhood cancer is still developing, its capabilities warrant continued investigation. Achieving this will demand systematic collection and responsible sharing of high-quality data from sufficient numbers of paediatric oncology patients.

Keywords: Paediatric oncology, Prognosis prediction, Prognostic model, Next-generation sequencing, Artificial intelligence, Public database

How to Cite This Article: Ben Youssef S, Trabelsi A, Boudiaf K, Jebali N. Progression of Prognostic Prediction in Pediatric Cancer from Conventional Factors to Artificial Intelligence: A Systematic Review and Meta-Analysis. Interdiscip Res Med Sci Spec. 2026;6(1):255-70. https://doi.org/10.51847/EQqVi0kUtt

Introduction

Paediatric cancers display varied biological features, causes, and disease progressions [1-3]. In wealthy nations, cure rates surpass 80% [4]; yet outcomes worsen considerably after adverse events (for instance, the 5-year

survival after relapse of acute lymphoblastic leukaemia reaches only 0.58 in Nordic data) [5]. This underscores the need to distinguish patients likely to follow different clinical paths [6]. The limited size of paediatric groups restricts robust scientific publications and reliable data for treatment choices. Globally, roughly 400,000 new paediatric cancer diagnoses occur annually [7].

These cancers fall into three primary groups according to tumour origin, each with distinct treatments, predictive instruments, and results. Haematological malignancies (HM) and CNS tumours are the most frequent. Non-CNS solid tumours (NCNSST) encompass many uncommon histological subtypes. Worldwide age-adjusted incidence for common leukaemias is 5.41 per 100,000, while brain tumours stand at 2.04 per 100,000—contrasting sharply with rarer types such as liver cancer at 0.19 per 100,000 [8]. Although uncommon, NCNSST remain important; for example, 2020 US figures showed 5-year OS below 70% for bone tumours, soft tissue sarcomas, and hepatoblastoma [9].

Even with recent therapeutic advances, individualised treatment planning and high-risk patient handling continue to pose difficulties, highlighting the value of clinical decision aids. Intensive multi-arm regimens are often necessary, yet inadequate initial risk assessment can result in overtreatment for some and disease advancement for others [10, 11].

In the past ten years, prognostic instruments in paediatric oncology have progressed from basic reliance on visible characteristics [12-14] toward incorporation of detailed molecular data through next-generation sequencing (NGS) [15-17], thereby refining risk grouping [18]. AI-supported methods hold potential for modelling intricate relationships among diverse data types and facilitating their integration [19, 20].

Nevertheless, absent previous meta-analyses, impartial data on the real-world effectiveness of these tools in paediatric oncology is missing. The purpose of this systematic review and meta-analysis was to deliver a thorough summary of prognostic modelling progress in childhood cancer, offering fresh understanding of how to build precise models and their practical consequences, ultimately supporting creation of stronger prognostic resources.

Materials and Methods

The investigation adhered to PRISMA 2020 standards [21] and Cochrane guidance [22]. The protocol was preregistered on PROSPERO (CRTN42022370251). Following the initial search, the analysis was modestly expanded beyond the original protocol to cover extra endpoints [23].

Eligibility criteria

The PICO framework guided the clinical question [24]. Investigations involving paediatric cancer patients were accepted. Different prognostic strategies served as the intervention and comparator. Primary outcomes were AUROC values (with 95% CI) for 1-, 2-, 3-, and 5-year OS and EFS, plus 10-year OS, and non-time-specific OS/EFS predictions. Secondary outcomes included C-index statistics for OS, EFS, and cancer-specific survival.

Included studies involved patients younger than 21 years (more than half, or mean/median age criterion) diagnosed with primary malignant tumours. Reports had to supply AUROC or C-index measures of prognostic model or marker performance, along with supporting statistics (SE, SD, 95% CI, sensitivity, specificity, predictive values, or case counts). When training and validation cohorts were analysed, data from the validation set were required. Further details appear in Supplementary Methods S3.

Information sources

Searches were executed in Embase, MEDLINE/PubMed, Cochrane CENTRAL, and Scopus on 22 November 2022. Backward citation screening of included articles was performed on 7 June 2023 using a reference tool [25]. An update search was completed on 28 June 2024. No language filters were used.

Search strategy

The query strategy merged four core domains: paediatric, cancer, prognosis, and accuracy.

Data extraction

Two independent teams (P.V. and T.K.+Sz.K.D.) extracted information, with discrepancies settled by the corresponding author (E.T.). Data were recorded in an Excel file (Microsoft Office 365). Prognostic elements and models were grouped into categories. Category-1 relied on standard clinical and basic genetic variables; Category-

2 employed NGS-derived genomic-transcriptomic data. Category-3 integrated clinical and NGS information but applied machine-learning and AI techniques to select optimal variables. Category-0 was introduced for weakly predictive single factors reported in studies for comparative purposes. Screening and selection processes are detailed in Supplementary Methods S1.

Study risk of bias assessment

Two teams (P.V. and Á.T.+G.M.) independently judged risk of bias. QUIPS was applied to single-factor studies [26], while PROBAST was used for multi-variable prognostic models [27]. Disagreements were resolved through discussion with the corresponding author (E.T.).

Synthesis methods

Analyses were conducted in R version 4.1.2 [28]. Statistical significance was set at $p < 0.05$. Given the high volume of tests, some false-positive findings may exist; smaller p-values reduce this likelihood. AUROC values were analysed separately by time horizon. Standard deviations were derived from reported CIs or estimated via Hanley and McNeil's method [29] using survival curves and patient counts when necessary. Results were displayed in forest plots.

C-index values were meta-analysed following Debray *et al*. [30]. Publication bias was examined with funnel plots. Conventional heterogeneity metrics were computed in selected instances (details in Supplementary Methods S4). Studies were grouped mainly by tumour type (HM, CNST, NCNSST) and model category. Subgroup analyses addressed validation method (internal/external/composite) and inclusion of treatment response as a predictor. Public datasets for neuroblastoma [31-33] and osteosarcoma [31, 34-37] were compared internally and against other sources.

Role of the funding source

No external funding supported this work. M.O., T.K., Sz.K.D., and V.P. had full data access and responsibility for submission decisions.

Results and Discussion

The systematic searches identified 10,870 relevant records from the four databases, plus 7,986 from reference lists and 1,252 from the update. In total, 385 articles qualified: 358 (including 92 from the update) entered the meta-analysis, and 27 additional studies were included in the systematic review **(Figure 1)**.

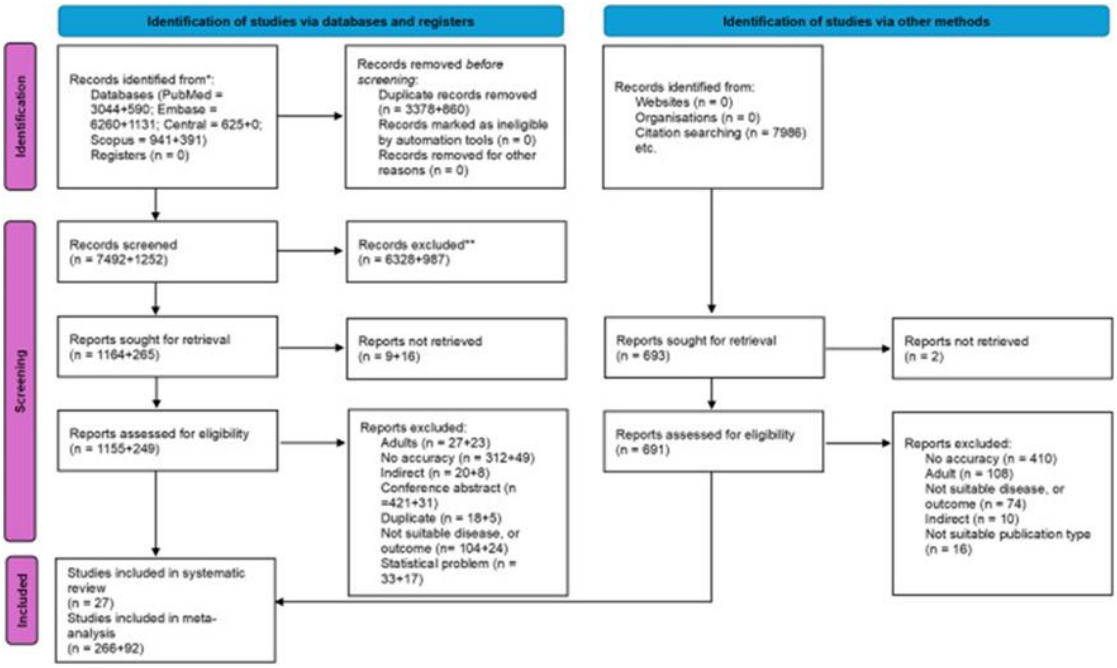


Figure 1. PRISMA diagram illustrating the study selection procedure.

The majority (379) of the selected investigations were retrospective cohort designs, supplemented by 5 prospective cohort studies and 1 cross-sectional study. These publications spanned the years 1991 to 2024. Contributions came from multiple geographic areas, with the greatest share (73.25%) originating from China. Among the included papers, 272 primarily addressed prognosis in NCNSST, while 81 focused on HM and 32 on CNST. Within the meta-analysis subset, Category-2 approaches were most common (168 studies), followed by Category-1 (145 studies), and Category-3 with notably fewer reports (only 45 studies). The distribution of studies added during the updated search closely mirrored the original set with regard to both model categories and tumour types.

Regarding NCNSST, pooled AUROC (pAUROC) estimates for 1-year OS remained comparable across model types: 0.8 (CI: 0.75–0.85) for Category-1, 0.85 (CI: 0.80–0.91) for Category-2, and 0.81 (CI: 0.74–0.88) for Category-3. No statistically meaningful differences were detected between categories (Category-1 vs. -2: $p = 0.169$; Category-1 vs. -3: $p = 0.831$; Category-2 vs. -3: $p = 0.245$). At the 2-year interval, Category-2 models performed strongly with a pAUROC of 0.80 (CI: 0.72–0.89). Category-3 models declined slightly to 0.76 (CI: 0.64–0.88), yet the comparison did not reach significance ($p = 0.659$).

Beyond 3 years, Category-1 performance fell to 0.77 (CI: 0.73–0.80), while Category-2 sustained a strong pAUROC of 0.84 (CI: 0.81–0.86). Category-3 yielded 0.77 (CI: 0.73–0.80). A significant advantage was observed only between Categories 1 and 2 ($p = 0.035$), favouring the latter (Category-1 vs. -3: $p = 0.552$; Category-2 vs. -3: $p = 0.061$).

For 5-year OS predictions — regarded as the benchmark outcome in paediatric cancer prognosis and therefore featured as the primary forest plot in the main document **(Figure 2)** — Category-2 demonstrated solid results with a pAUROC of 0.85 (CI: 0.82–0.88). Category-3 achieved 0.82 (CI: 0.77–0.88), while Category-1 dropped to 0.75 (CI: 0.72–0.79). As seen at 3 years, a clear statistical difference existed between Category-1 and Category-2 ($p < 0.001$), again favouring Category-2. Category-3 models showed nearly significant superiority over Category-1 ($p = 0.0511$). Category-2 models continued to exhibit strong performance even at the 10-year time point (pAUROC: 0.79, CI: 0.74–0.85).

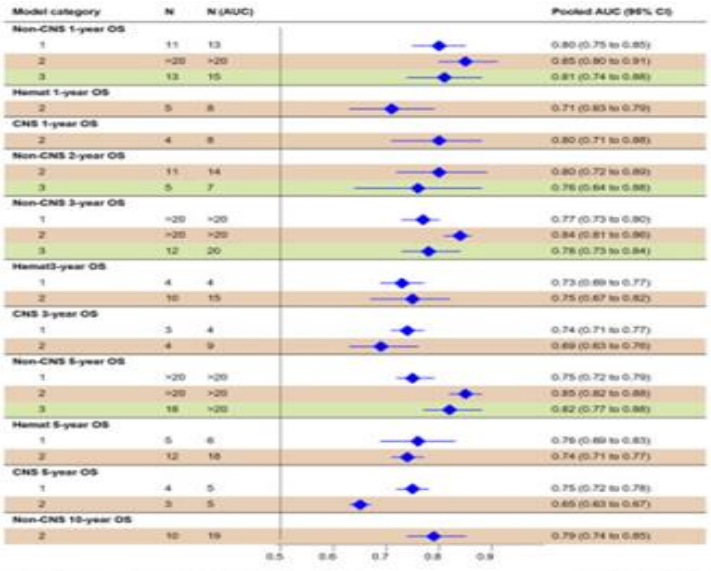
Figure 2. Forest plot displaying area under the receiver operating characteristic curve (AUROC) estimates along with their 95% confidence intervals (CI) for forecasting 5-year overall survival among paediatric cases involving non-CNS solid tumours, blood cancers, and CNS tumours, grouped according to the three prognostic model types.

1, 2, 3 I-internal validation, E-external validation, C-composite validation.

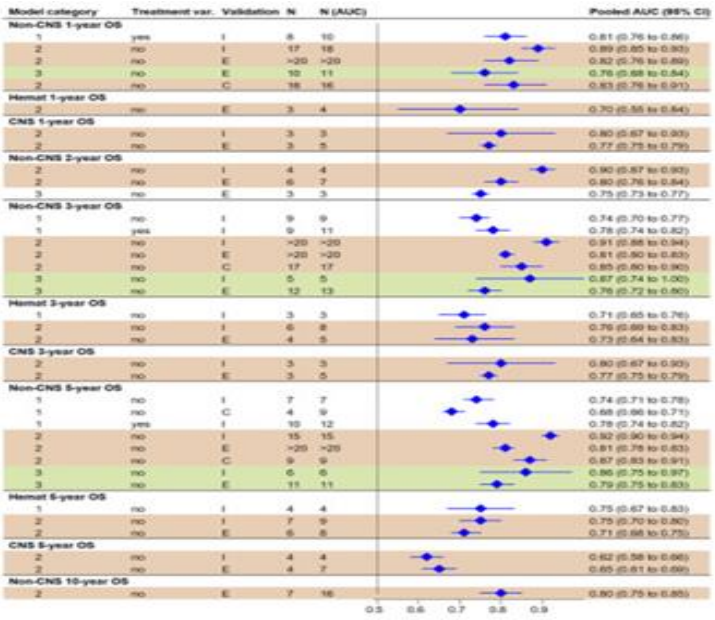
Analysis of how well the models performed for haematological malignancies (HM) was restricted due to scarce available information. Regarding 1-year OS, pooling was possible exclusively with Category-2 publications,

producing a pAUROC of 0.71 (CI: 0.63–0.79). For 3-year OS, outcomes from Category-1 and Category-2 were closely aligned, recording pAUROC figures of 0.73 (CI: 0.69–0.77) and 0.75 (CI: 0.67–0.82), respectively (Category-1 vs Category-2: p = 0.824). Turning to the 5-year timeframe shown in **Figure 2**, Category-1 recorded a pAUROC of 0.76 (CI: 0.69–0.83) compared with 0.74 (CI: 0.71–0.77) for Category-2 (Category-1 vs Category-2: p = 0.562).

For central nervous system tumours (CNST), adequate data for 1-year OS estimation was limited to Category-2 studies, which delivered a strong pAUROC of 0.8 (CI: 0.71–0.88). Assessments covering both Category-1 and Category-2 were feasible for the 3-year and 5-year OS endpoints. Category-1 achieved pAUROC values of 0.74 (CI: 0.71–0.77) at 3 years and 0.75 (CI: 0.72–0.78) at 5 years. Category-2, however, showed lower performance with 0.69 (CI: 0.63–0.76) at 3 years and 0.64 (CI: 0.63–0.65) at 5 years. The comparison between Category-1 and Category-2 for 3-year OS was not significant (p = 0.282), whereas the difference proved statistically meaningful for 5-year OS (p = 0.0141) **(Figure 3a)**.



a)



b)

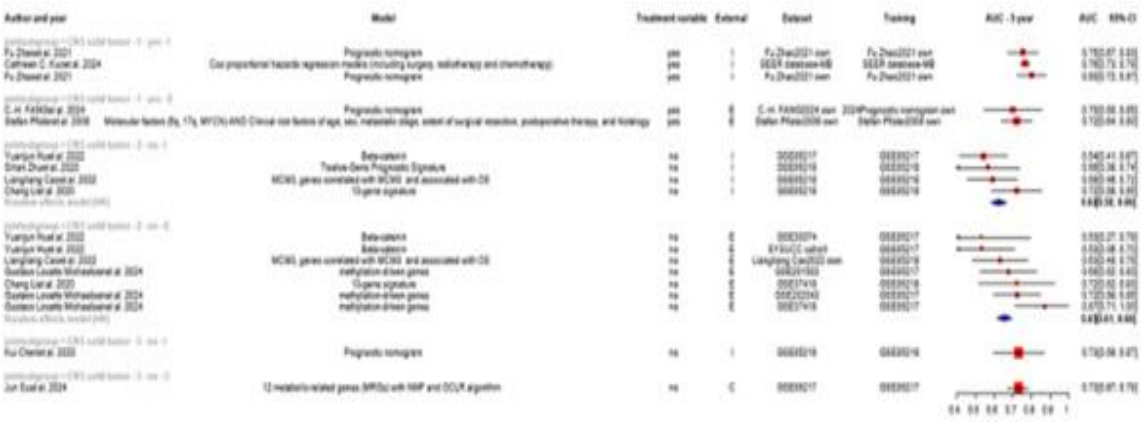
Figure 3. Consolidated forest plot illustrating the precision of the three prognostic model groups (1, 2, 3) in estimating 1-, 2-, 3-, 5-, and 10-year overall survival rates for paediatric oncology patients diagnosed with haematological malignancies, non-CNS solid tumours, or CNS tumours.

a) Core evaluation stratified according to tumour group and model classification.

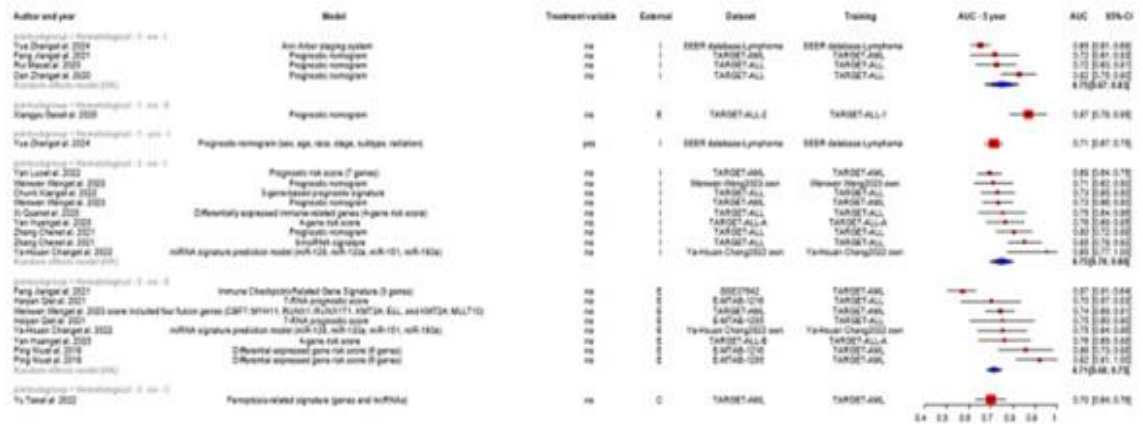
b) Additional subgroup evaluation that accounts for the type of validation performed and whether therapy response was included as a predictive variable.

AUC= area under the receiver operating characteristic curve. N= count of studies within each group. N (AUC)= count of AUC measurements within each group. CI= confidence interval. OS= overall survival. var.= variable. I=internal validation. E= external validation. C= composite validation. Non-CNS= non-central nervous system solid tumours. Hemat= haematological malignancies. CNS= central nervous system tumours. White= Category-1. Red= Category-2. Green= Category-3.

A clear statistically meaningful distinction appeared in the performance of Category-2 models when predicting 1-year OS, with NCNSST outperforming HM (NCNSST vs HM: p = 0.046). Category-2 approaches also demonstrated substantially stronger predictive capability for 5-year OS in NCNSST patients compared with those having CNST (NCNSST vs CNST: p = 0.003).Further subgroup examinations divided the data beyond model categories to include validation strategy (internal, external, or composite) and the incorporation of treatment outcome as a prognostic element (yes or no). Within CNST patients, direct contrasts were feasible solely between the 2-no-I and 2-no-E subsets, yet these yielded no notable disparities at any examined interval (1-year OS: p = 0.691; 3-year OS: p = 0.694; 5-year OS: p = 0.556) **(Figure 4a)**. In HM cases, likewise, neither model category nor validation approach produced any significant distinctions among the subgroups (for 3-year OS: 1-no-I vs 2-no-I, p = 0.501 and 2-no-I vs 2-no-E, p = 0.520; for 5-year OS: 1-no-I vs 2-no-I, p = 0.995 and 2-no-I vs 2-no-E, p = 0.448) **(Figure 4b)**.



a)



b)

Figure 4. a) Area under the receiver operating characteristic curve (AUC) estimates with 95% confidence intervals (CI) for 5-year overall survival prediction in children diagnosed with central nervous system (CNS) tumours. I = internal validation, E = external validation, C = composite validation.
b) Area under the receiver operating characteristic curve (AUC) estimates with 95% confidence intervals (CI) for 5-year overall survival prediction in children with haematological malignancies. I = internal validation, E = external validation, C = composite validation.

For non-CNS solid tumours (NCNSST), available information was too limited to allow reliable comparisons at the 10-year OS mark. Nevertheless, clear statistical advantages emerged for 3-year and 5-year OS predictions when contrasting the 2-no-I subgroup against the 1-no-I subgroup ($p < 0.001$ for both time points), with superior results from the former. A comparable benefit was detected between the 2-no-I and 2-no-C subgroups specifically for 3-year OS forecasting ($p = 0.0074$). In addition, for 3-year and 5-year OS predictions displayed in **Figure 5**, analogous significant improvements were noted when comparing 2-no-I against 2-no-E subgroups ($p < 0.001$ in both cases).

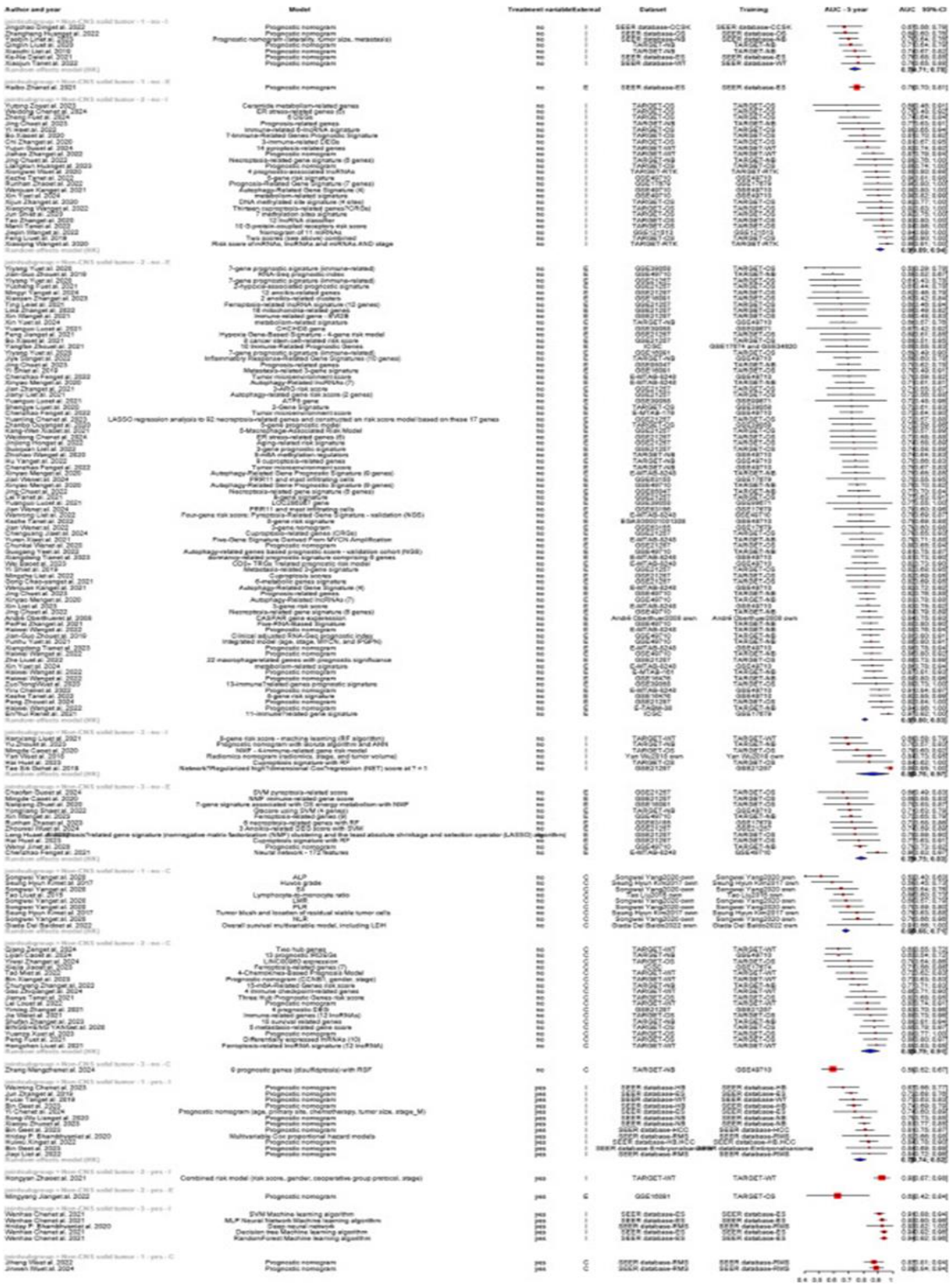
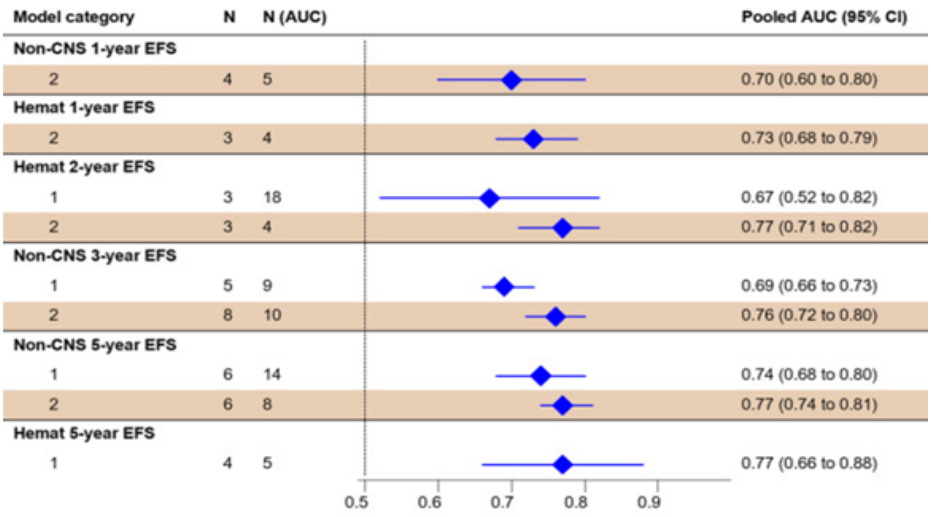


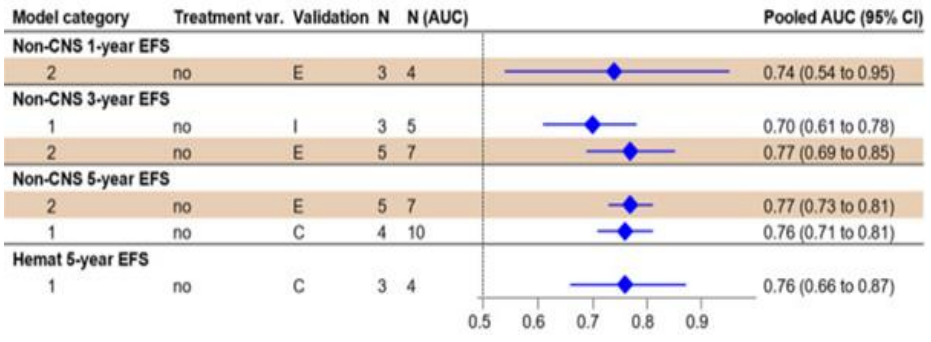
Figure 5. Plot displaying area under the receiver operating characteristic curve (AUC) figures with accompanying 95% confidence intervals (CI) for the prediction of 5-year overall survival in paediatric cases

of non-central nervous system (non-CNS) tumours. I-internal validation, E-external validation, C-composite validation.

Comparisons between the 1-yes-I and 1-no-I subgroups — differentiated solely by the presence or absence of treatment response information as a predictor — revealed no notable variations for either 3-year OS (p = 0.479) or 5-year OS (p = 0.276). In contrast, a clear statistical distinction surfaced in 2-year OS forecasting that involved both the model category and the validation method applied (specifically, 2-no-E versus 3-no-E: p = 0.031; and 2-no-E versus 2-no-I: p < 0.001). Across the remaining subgroup analyses, Categories 2 and 3 showed no significant performance gaps (1-year OS: 2-no-I vs 3-no-I, p = 0.983; 3-year OS: 2-no-E vs 3-no-E, p = 0.126 and 2-no-I vs 3-no-I, p = 0.617; 5-year OS: 2-no-E vs 3-no-E, p = 0.2997 and 2-no-I vs 3-no-I, p = 0.353) **(Figure 3b)**.Turning to NCNSST patients, pooled estimates from Category-2 models for 1-year event-free survival (EFS) produced a pAUROC of 0.70 (CI: 0.60–0.80). Data for 3-year EFS permitted analysis of both Category-1 (pAUROC 0.69, CI: 0.66–0.73) and Category-2 (pAUROC 0.76, CI: 0.72–0.80), though the difference between them did not reach statistical significance (p = 0.065). Similarly, 5-year EFS results were close between Category-1 (pAUROC 0.74, CI: 0.68–0.80) and Category-2 (pAUROC 0.77, CI: 0.74–0.81), with no significant divergence (p = 0.32). For haematological malignancies (HM), Category-2 models achieved a pAUROC of 0.73 (CI: 0.68–0.79) when predicting 1-year EFS. At the 2-year EFS mark, Category-1 recorded 0.67 (CI: 0.52–0.82) while Category-2 reached 0.77 (CI: 0.71–0.82); the comparison was not significant (p = 0.302). Only Category-1 data were present for 5-year EFS in HM, yielding a pAUROC of 0.77 (CI: 0.66–0.88) **(Figure 6a)**.



a)

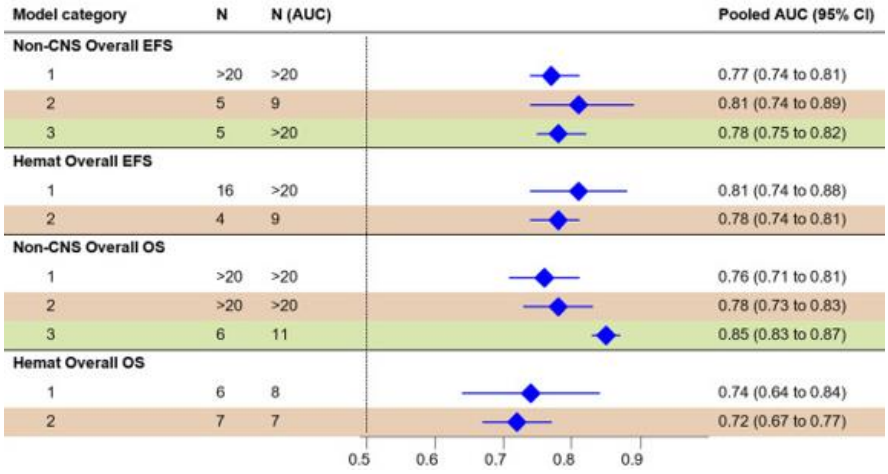


b)

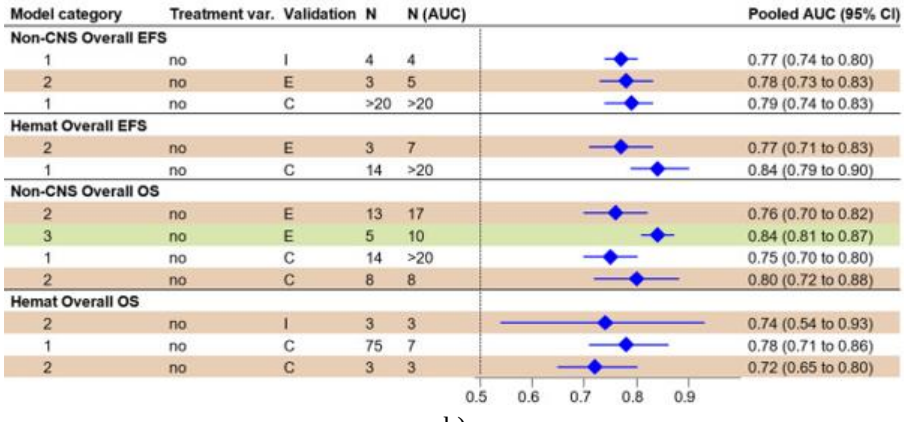
Figure 6. Consolidated forest plot presenting the predictive capability of the three prognostic model groups (1, 2, 3) for 1-, 2-, 3-, and 5-year event-free survival in paediatric cancer patients diagnosed with either haematological malignancies or non-central nervous system solid tumours.
a) Main evaluation divided by tumour category and model classification.
b) Additional subgroup evaluation that includes validation strategy and whether treatment results served as a prognostic element.

AUC= area under the receiver operating characteristic curve. N= number of articles in a certain group. N (AUC)= number of AUC values in a certain group. CI= confidence interval. EFS= event-free survival. var.= variable. I= internal validation. E= external validation. C= composite validation. Non-CNS= non-central nervous system solid tumours. Hemat= haematological malignancies. White= Category-1. Red= Category-2. Green= Category-3.

Substantially fewer records were accessible in the various subgroups for these particular endpoints, making it impossible to perform any informative statistical tests. A minor variation stood out in the 2-no-E subset for NCNSST cases, in which the pAUROC for 1-year EFS prediction stood at 0.74 (CI: 0.54–0.95) — lower than the figures obtained for 3-year EFS at 0.77 (CI: 0.69–0.85) and 5-year EFS at 0.77 (CI: 0.73–0.81) **(Figure 6b)**.
In the analysis of non-time-dependent overall survival for NCNSST patients, models from Category-3 led the way with a pAUROC of 0.85 (CI: 0.83–0.87). Category-1 and Category-2 delivered nearly equivalent results of 0.76 (CI: 0.71–0.81) and 0.78 (CI: 0.73–0.83), respectively (Category-1 vs -2: p = 0.578). However, both were surpassed in a statistically meaningful manner by Category-3 (Category-1 vs -3: p = 0.014; Category-2 vs -3: p = 0.035). For HM, the two categories performed on a similar level, recording pAUROC values of 0.74 (CI: 0.64–0.84) for Category-1 and 0.72 (CI: 0.67–0.77) for Category-2 (Category-1 vs -2: p = 0.805).
Shifting focus to event-free survival in the NCNSST population, information covered all model categories and produced closely matching outcomes: 0.77 (CI: 0.74–0.81) for Category-1, 0.81 (CI: 0.74–0.89) for Category-2, and 0.78 (CI: 0.75–0.82) for Category-3 (Category-1 vs -2: p = 0.377; Category-1 vs -3: p = 0.715). In HM patients, Category-1 achieved a pAUROC of 0.81 (CI: 0.74–0.88) compared with 0.78 (CI: 0.74–0.81) for Category-2 (p = 0.454) **(Figure 7a)**.



a)



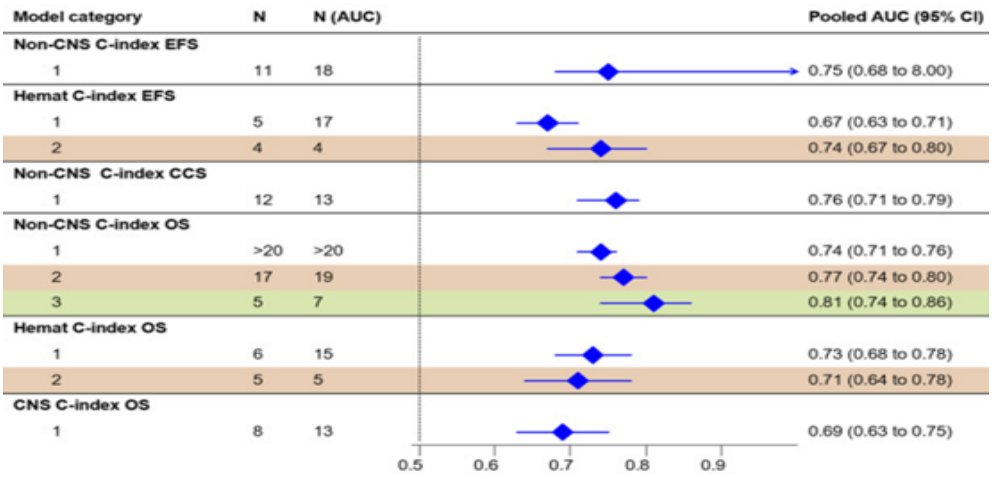
b)

Figure 7. Overview forest plot highlighting the effectiveness of the three prognostic model classes (1, 2, 3) when estimating overall survival and event-free survival in paediatric oncology patients affected by haematological malignancies or non-central nervous system solid tumours.

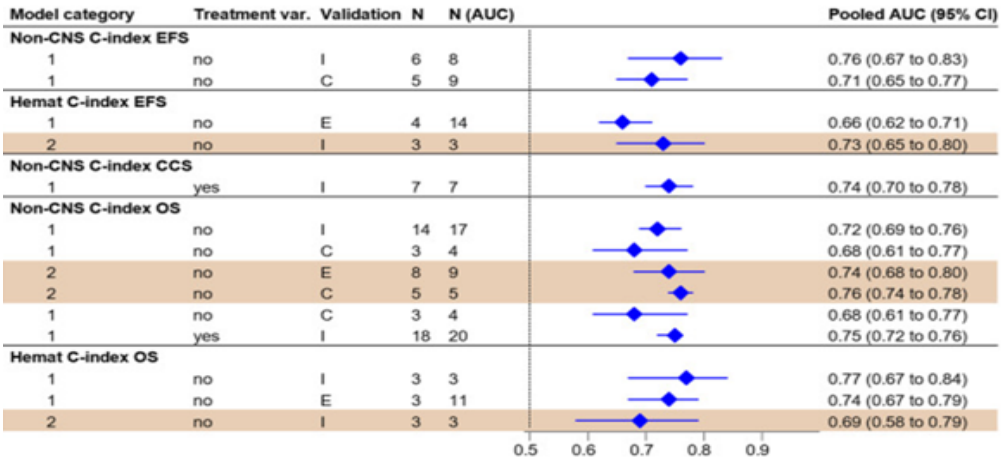
a) Core evaluation organized by tumour category and model type.
b) Extended subgroup evaluation that factors in the validation technique and the role of treatment outcome as a predictive element.

AUC= area under the receiver operating characteristic curve. N= number of articles in a certain group. N (AUC)= number of AUC values in a certain group. CI= confidence interval. OS= overall survival. EFS= event-free survival. var.= variable. I= internal validation. E= external validation. C= composite validation. Non-CNS= non-central nervous system solid tumours. Hemat= haematological malignancies. White= Category-1. Red= Category-2. Green= Category-3.

Limited data on non-time-dependent event-free survival restricted the subgroup analysis to a single feasible comparison, which did not demonstrate any meaningful distinction (1-no-I vs 1-no-C: p = 0.62). For overall survival forecasting, the 2-no-E and 3-no-E subgroups showed no significant separation (p = 0.063), mirroring the results from time-dependent assessments. A visible though non-significant pattern favoured Category-1 over Category-2 for OS prediction in both HM and NCNSST cohorts. In HM cases, Category-1 tended to yield stronger outcomes relative to Category-2 (1-no-C vs 2-no-C: p = 0.339), while a parallel non-significant trend was noted in NCNSST patients (1-no-C vs 2-no-C: p = 0.345) **(Figure 7b)**.Serving as an additional endpoint, the concordance index supplies extra detail on the reliability of predictions for overall survival (OS), event-free survival (EFS), and cancer-specific survival (CSS). Among NCNSST patients, all model categories delivered solid OS performance. Category-3 stood out with a pooled C-index (pC-index) of 0.81 (CI: 0.74–0.80), surpassing Category-1 at 0.74 (CI: 0.71–0.76) and Category-2 at 0.77 (CI: 0.74–0.80). OS estimates for HM and CNST were largely based on Category-1 models, producing pC-index figures of 0.73 (CI: 0.68–0.78) for HM and 0.69 (CI: 0.63–0.75) for CNST. Category-2 models were also applied to OS in HM, resulting in a pC-index of 0.71 (CI: 0.64–0.78). For EFS in NCNSST, Category-1 models generated a pC-index of 0.75 (CI: 0.68–0.80). In HM, Category-1 reached 0.67 (CI: 0.63–0.71) while Category-2 performed better at 0.74 (CI: 0.67–0.80). CSS prediction via Category-1 models in NCNSST patients achieved a pC-index of 0.76 (CI: 0.71–0.79) **(Figure 8a)**.



a)



b)

Figure 8. Consolidated forest plot illustrating the performance of the three model types (1, 2, 3) for predicting overall survival, event-free survival, and cancer-specific survival in paediatric cancer patients with haematological malignancies, non-central nervous system solid tumours, or central nervous system tumours.
a) Primary analysis stratified by tumour type and model category.
b) Subgroup analysis incorporating validation method and inclusion of treatment outcome as a prognostic variable.

C-index: concordance index. N: number of articles in a certain group. N (C-index): number of C-index values in a certain group. CI: confidence interval. OS: overall survival. EFS: event-free survival. CSS: cancer-specific survival. var.: variable. I: internal validation. E: external validation. C: composite validation. Non-CNS: non-central nervous system solid tumours. Hemat: haematological malignancies. CNS: central nervous system tumours. White: Category-1. Red: Category-2. Green: Category-3.

Due to insufficient data, no comparisons could be performed for cancer-specific survival (CSS). For event-free survival, only one comparison was possible and it showed no statistical significance (NCNSST patients: 1-no-C vs 1-no-I, p = 0.277).

In the subgroup analysis, overall survival was the most frequently reported outcome measured by C-index. No statistically significant difference was found between Category-1 (1-no-E) and Category-2 (2-no-E) models in NCNSST patients; in fact, Category-2 performed slightly worse (pC-index 0.81, CI: 0.75–0.94 vs 0.74, CI: 0.68–0.80; p = 0.1916). The previously observed strong effect of validation type was not evident here. External validation studies using Category-1 models tended to outperform those with internal validation (1-no-I vs 1-no-E: pC-index 0.72, CI: 0.69–0.76 vs 0.81, CI: 0.75–0.94; p = 0.11). When assessing the impact of including therapy response as a prognostic factor, its incorporation yielded modestly improved results, although the difference did not reach statistical significance (1-no-I vs 1-yes-I: pC-index 0.72, CI: 0.69–0.76 vs 0.75, CI: 0.72–0.76; p = 0.059). For HM, enough studies allowed comparison of validation types, but no significant difference emerged (1-no-E vs 1-no-I: p = 0.5553) **(Figure 8b)**.

The risk of bias (RoB) evaluation indicated low risk in the participant selection and outcome domains thanks to strict eligibility standards. In contrast, high RoB was common in the predictor and analysis domains, primarily stemming from problematic variable selection, reliance on univariable methods, and the frequent use of internal validation.

Results concerning Category-0, specific training/validation dataset pair comparisons, and studies included only in the systematic review are provided in the Supplementary Results.

This work represents the initial Level II evidence [38] regarding the accuracy of prognostic tools in paediatric oncology. It synthesised data from 385 investigations covering short-, intermediate-, and long-term survival as well as different adverse events. Superior performance of more advanced Category-2 and Category-3 models was evident for 5-year OS in NCNSST patients (Category-1 vs -2: p < 0.001; Category-1 vs -3: p = 0.0511). Category-3 models proved most effective for non-time-dependent OS (Category-1 vs -3: p = 0.014; Category-2 vs -3: p = 0.035). The substantial influence of validation strategy was confirmed through direct comparisons of internal versus external validation (3- and 5-year OS: 2-no-I vs 2-no-E, p < 0.001 for both).

Advances in biological knowledge through genetic and molecular profiling have enhanced tumour classification, outcome forecasting, and risk grouping [39]. This progress is reflected in the literature, where 161 included studies employed contemporary Category-2 or -3 approaches. Traditional Category-1 models with solid statistical foundations existed prior to 2000, with the oldest qualifying publications dating to the 1990s. Publication numbers surged after 2020, during which Category-2 studies became dominant (56.5% of 184 studies from the 2020s) alongside wider adoption of NGS. Category-1 remains widely used, accounting for nearly half (45.45%) of its articles published after 2020. Category-3 has shown limited growth over time. Although such studies first appeared in 2004, only 50 articles in total (45 in the meta-analysis and 5 in the systematic review) utilised AI methods.

Model development has evolved rapidly from Category-0 to Category-3. Next-generation techniques involving NGS (featured in both Category-2 and -3) evaluate hundreds of variables, far exceeding the limited factors considered in conventional clinical models (Category-1). Authors often began with standard prognostic elements, using them as benchmarks against their own developed systems (classified as Category-0). These basic factors generally underperformed compared with dedicated models (data not shown). Nevertheless, established clinical scoring systems retain their value as a reliable basis for paediatric cancer management [40]. Category-1 models incorporating pathological features (such as histology or stage) or radiomic data performed especially strongly. For non-time-dependent OS in HM, Category-1 even surpassed Category-2. Given the higher frequency of HM [8], Category-1 approaches in this area are mature and utilise more refined predictors. Most studies fell under Category-2, where models typically combined thousands of genes or RNA profiles with clinical variables or staging information, producing more precise and holistic tools. Category-3 models, which merge clinical, biomarker, and genomic-transcriptomic information through AI, were rarer but achieved performance comparable to Category-2.

The low incidence of paediatric cancers [7] and consequent data scarcity represent key barriers to building reliable prognostic tools. Prioritising data enrichment for rarer NCNSST subtypes [20, 41] focused on OS prediction appears strategic. In contrast to HM, where survival rates are generally higher [5], EFS serves as a more suitable target as it reflects both survival and quality of life. Across short-, mid-, and long-term horizons in NCNSST, all model categories performed well for 1-year OS, but divergence appeared at 3 and 5 years as Category-1 accuracy declined. This pattern was absent in HM, though results must be interpreted cautiously due to validation-related bias. Category-1 models may appear advantageous in HM for both OS and EFS, yet most were internally validated, inflating performance relative to the predominantly externally validated advanced models. External validation on independent cohorts remains the recommended standard, as supported by this meta-analysis.

Advanced models have shown particular strength in NCNSST prognosis, especially for OS — the established benchmark in oncology. Accurate risk stratification based on this endpoint is vital for tailoring intensive therapies [10, 11]. Significant superiority of Category-2 and Category-3 models was confirmed for 5-year OS, with Category-3 excelling in non-time-dependent predictions. However, the June 2024 update search identified 30 additional Category-1 studies published in the last 1.5 years, even for NCNSST OS forecasting. These findings support broader adoption of NGS-based (Category-2 or -3) approaches when feasible. Although AI models using extensive sequencing data show promise for superior accuracy, the current evidence base remains limited (168 Category-2 vs 45 Category-3 studies in the meta-analysis). Notably, HM research using Category-2 models has expanded rapidly (10 new articles added to the prior 17 between 2022 and 2024). These models, while not statistically superior, tended to predict EFS more accurately than OS. No Category-3 studies have yet addressed this tumour group. Given growing recognition of EFS as a valid surrogate in several HM [42-44], prioritising Category-3 evaluation in this setting is strongly advised.

Data handling and variable selection are equally critical as the choice of predictors themselves. Common practices such as univariate followed by multivariate Cox regression can introduce bias by overlooking inter-variable relationships. This underscores the value of alternative statistical techniques. Internally validated models consistently outperformed externally validated ones, highlighting generalisability issues. Including treatment outcome as a predictor showed a positive but non-significant trend due to its infrequent application.

High-quality, standardised biobanks are essential for advancing research and individualised care, yet they remain underused and poorly harmonised, particularly for paediatric populations [41, 45]. Public resources such as TARGET and SEER have proven invaluable for robust model development. Greater international collaboration on data sharing is needed to create large, high-quality repositories suitable for training AI systems. When paired with proper validation, modern models deliver enhanced clinical utility [46, 47].

This review incorporated a broad array of studies and endpoints for comprehensive coverage, including recent cutting-edge research. Multiple subgroups were examined to mitigate heterogeneity and bias. The methodology followed established standards, and most studies drew on transparent, quality-controlled public datasets.

Several analytical challenges were addressed, including estimation of missing standard errors for AUROC and C-index values from available information. Persistent moderate-to-high RoB across studies represents a limitation. Rigorous quality measures and transparent bias reporting were maintained. The lack of tumour-specific breakdowns introduces heterogeneity that may obscure subtype-specific patterns better revealed through finer subgroup analyses.

The findings highlight the enhanced predictive strength of genomic-transcriptomic and especially AI-supported models for NCNSST. Subgroup results confirmed important accuracy differences related to validation type, with internal validation producing optimistically biased estimates. Although AI methods are not yet standard in clinical workflows, they appear capable of matching or complementing NGS-based approaches in paediatric cancer forecasting. This emphasises the urgent need for systematic, ethical data collection and sharing to overcome current limitations, highlighting the value of well-curated paediatric cancer registries and biobanks.

Conclusion

In summary, while this systematic review and meta-analysis cannot offer conclusive guidance for every individual tumour type, it constitutes the most comprehensive and current synthesis in the domain. It delivers important observations and practical recommendations to guide subsequent research.

Acknowledgments: None

Conflict of Interest: None

Financial Support: None

Ethics Statement: None

References

1. Blattner-Johnson M, Jones DTW, Pfaff E. Precision medicine in pediatric solid cancers. Semin Cancer Biol. 2022;84:214-7.
2. Zou H, Poore B, Broniscer A, Pollack IF, Hu B. Molecular heterogeneity and cellular diversity: implications for precision treatment in medulloblastoma. Cancers. 2020;12(3):643.
3. Quessada J, Cuccuini W, Saultier P, Loosveld M, Harrison CJ, Lafage-Pochitaloff M. Cytogenetics of pediatric acute myeloid leukemia: a review of the current knowledge. Genes. 2021;12(6):924.
4. Erdmann F, Frederiksen LE, Bonaventure A, Botia B, Clavel J, Desandes E, et al. Childhood cancer: survival, treatment modalities, late effects and improvements over time. Cancer Epidemiol. 2021;71(Pt B):101733.
5. Oskarsson T, Soderhall S, Arvidson J, Forestier E, Montgomery S, Bottai M, et al. Relapsed childhood acute lymphoblastic leukemia in the Nordic countries: prognostic factors, treatment and outcome. Haematologica. 2016;101(1):68-76.
6. Krzyszczyk P, Acevedo A, Davidoff EJ, Timmins LM, Marrero-Berrios I, Patel M, et al. The growing role of precision and personalized medicine for cancer treatment. Technology. 2018;6(03n04):79-100.
7. Steliarova-Foucher E, Colombet M, Ries LAG, Moreno F, Dolya A, Bray F, et al. International incidence of childhood cancer, 2001–10: a population-based registry study. Lancet Oncol. 2017;18(6):719-31.
8. Wu Y, Deng Y, Wei B, Xiang D, Hu J, Zhao P, et al. Global, regional, and national childhood cancer burden, 1990–2019: an analysis based on the Global Burden of Disease Study 2019. J Adv Res. 2022;40:233-47.
9. Lupo PJ, Spector LG. Cancer progress and priorities: childhood cancer. Cancer Epidemiol Biomarkers Prev. 2020;29(6):1081-94.
10. Chen K, Huang B, Yan S, Xie Z, Mao J, Liu L, et al. Two machine learning methods identify a metastasis-related prognostic model that predicts overall survival in medulloblastoma patients. Aging. 2020;12(21):21481-503.

11. Liu Z, Liang M, Grant CN, Spiegelman VS, Wang HG. Interpretable models for high-risk neuroblastoma stratification with multi-cohort copy number profiles. Inform Med Unlocked. 2021;25:100701.
12. Scrideli CA, Queiroz RGPD, Bernardes JE, Valera ET, Tone LG. PCR detection of clonal IgH and TCR gene rearrangements at the end of induction as a non-remission criterion in children with ALL: comparison with standard morphologic analysis and risk group classification. Med Pediatr Oncol. 2003;41(1):10-6.
13. Bulzico D, De Faria PAS, De Paula MP, De Lima MAV, Cordeiro de Noronha Pessoa CH, Ferreira da Silva AL, et al. Recurrence and mortality prognostic factors in childhood adrenocortical tumors: analysis from the Brazilian National Institute of Cancer experience. Pediatr Hematol Oncol. 2016;33(4):248-58.
14. Morandi F, Corrias MV, Levreri I, Scaruffi P, Cangemi G, Carlini B, et al. Serum levels of cytoplasmic melanoma-associated antigen at diagnosis may predict clinical relapse in neuroblastoma patients. Cancer Immunol Immunother. 2011;60(10):1485-95.
15. Zhang WB, Han FM, Liu LM, Jin HB, Yuan XY, Shang HS. Characterizing the critical role of metabolism in osteosarcoma based on establishing novel molecular subtypes. Eur Rev Med Pharmacol Sci. 2022;26(8):2926-43.
16. Qi W, Yan Q, Lv M, Song D, Wang X, Tian K. Prognostic signature of osteosarcoma based on 14 autophagy-related genes. Pathol Oncol Res. 2021;27:1609782.
17. Qian H, Lei T, Hu Y, Lei P. Expression of lipid-metabolism genes is correlated with immune microenvironment and predicts prognosis in osteosarcoma. Front Cell Dev Biol. 2021;9:673827.
18. O'Donohue T, Sait SF, Bender JG. Progress in precision therapy in pediatric oncology. Curr Opin Pediatr. 2023;35(1):41-7.
19. Quintás G, Yáñez Y, Gargallo P, Piqueras M, Cañete A, Gil-Benso R, et al. Metabolomic profiling in neuroblastoma. Pediatr Blood Cancer. 2020;67(3):e28113.
20. Wang X, Wu X, Li T, Zhang H, Liu Y, Sun Y, et al. Identification of biomarkers to construct a competing endogenous RNA network and establishment of a genomic-clinicopathologic nomogram to predict survival for children with rhabdoid tumors of the kidney. Biomed Res Int. 2020;2020:1-27.
21. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71.
22. Chandler J, Hopewell S. Cochrane methods - twenty years experience in developing systematic review methods. Syst Rev. 2013;2(1):76.
23. Booth A, Clarke M, Dooley G, Ghersi D, Moher D, Petticrew M, et al. PROSPERO at one year: an evaluation of its utility. Syst Rev. 2013;2(1):4.
24. Munn Z, Stern C, Aromataris E, Lockwood C, Jordan Z. What kind of systematic review should I conduct? A proposed typology and guidance for systematic reviewers in the medical and health sciences. BMC Med Res Methodol. 2018;18(1):5.
25. Haddaway NR, Grainger MJ, Gray CT. Citationchaser: a tool for transparent and efficient forward and backward citation chasing in systematic searching. Res Synth Methods. 2022;13(4):533-45.
26. Hayden JA, Van Der Windt DA, Cartwright JL, Côté P, Bombardier C. Assessing bias in studies of prognostic factors. Ann Intern Med. 2013;158(4):280.
27. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med. 2019;170(1):51.
28. R Core Team. R: A language and environment for statistical computing [Internet]. Vienna: R Foundation for Statistical Computing; 2021.
29. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143(1):29-36.
30. Debray TP, Damen JA, Riley RD, Snell K, Reitsma JB, Hooft L, et al. A framework for meta-analysis of prediction model studies with binary and time-to-event outcomes. Stat Methods Med Res. 2019;28(9):2768-86.
31. Genomic Data Commons, Therapeutically Applicable Research to Generate Effective Treatments (TARGET) [Internet]. 2007-2024 [cited 2024]. Available from: https://www.cancer.gov/ccg/research/genome-sequencing/target
32. Zhang W, Shi L, Hertwig F, Kool M, Pfister SM, Jones DTW, et al. SuperSeries GSE47792; SubSeries GSE49710, GSE49711, GSE62564 [Internet]. 2014.
33. Hammerschmidt W, Buschle A. E-MTAB-8428 [Internet]. 2021.

34. Davis S. SuperSeries GSE16102; SubSeries GSE16091 [Internet]. 2009.
35. Buddingh EPK, Marieke L, Duim RAJ, Kuijjer ML, Koster J, Hogendoorn PCW, et al. Anne-marie series GSE21257 [Internet]. 2011.
36. Kelly A. SuperSeries GSE39058 [Internet]. 2013.
37. Ho D, Kõks S, Phung P, Degner A, Le J, Schmitz A, et al. Series GSE99671 [Internet]. 2017.
38. Burns PB, Rohrich RJ, Chung KC. The levels of evidence and their role in evidence-based medicine. Plast Reconstr Surg. 2011;128(1):305-10.
39. American Cancer Society. Cancer facts & figures 2023 [Internet]. Atlanta: American Cancer Society; 2023.
40. Lehrnbecher T, Robinson PD, Ammann RA, Fisher B, Groll AH, Haeusler GM, et al. Guideline for the management of fever and neutropenia in pediatric patients with cancer and hematopoietic cell transplantation recipients: 2023 update. J Clin Oncol. 2023;41(9):1774-85.
41. Joseph N, Roberts CK, Graham K, Bastarache L, Haines JL, Roden DM, et al. Biobanking in the twenty-first century: driving population metrics into biobanking quality. In: Karimi-Busheri F, editor. Biobanking in the 21st century. Cham: Springer; 2015. p. 95-114. (Advances in experimental medicine and biology; vol. 864).
42. Maurer MJ, Ellin F, Srour L, Ristow K, Habermann TM, Witzig TE, et al. International assessment of event-free survival at 24 Months and subsequent survival in peripheral T-cell lymphoma. J Clin Oncol. 2017;35(36):4019-26.
43. Zhu J, Yang Y, Tao J, Liu F, Chen H, Zhong Z, et al. Association of progression-free or event-free survival with overall survival in diffuse large B-cell lymphoma after immunochemotherapy: a systematic review. Leukemia. 2020;34(10):2576-91.
44. Norsworthy KJ, Gao X, Ko CW, Przepiorka D, Gormley NJ, Farrell AT, et al. Response rate, event-free survival, and overall survival in newly diagnosed acute myeloid leukemia: US food and drug administration trial-level and patient-level analyses. J Clin Oncol. 2022;40(8):847-54.
45. Rush A, Byrne JA, Watson PH. Applying findable, accessible, interoperable, and reusable principles to biospecimens and biobanks. Biopreserv Biobank. 2024.
46. Hegyi P, Eross B, Izbeki F, Parniczky A, Szentesi A. Accelerating the translational medicine cycle: the Academia Europaea pilot. Nat Med. 2021;27(8):1317-9.
47. Hegyi P, Petersen OH, Holgate S, Erjefalt J, McEntee K, Gallagher C, et al. Academia europaea position paper on translational medicine: the cycle model for translating scientific results into community benefits. J Clin Med. 2020;9(5):1532.