ISSN: 2476-2415

Research Article

International Journal of Clinical & Experimental Dermatology

# Artificial Intelligence Applied to Epigenomic Data in Leukemia: A Systematic Review of Diagnostic, Prognostic and Monitoring Models

**Jacopo Martelli***

*Convitto Nazionale Vittorio Emanuele II, Rome, Italy*

***Corresponding Author**
Jacopo Martelli, Convitto Nazionale Vittorio Emanuele II, Italy.

**Submitted:** 2026, Sep 28; **Accepted:** 2026, Oct 27; **Published:** 2026, Nov 06

**Citation:** Martelli, J. (2026). Artificial Intelligence Applied to Epigenomic Data in Leukemia: A Systematic Review of Diagnostic, Prognostic and Monitoring Models. *Int J Clin Expl Dermatol, 11*(3), 01-23.

## Abstract

***Background: Epigenetic Deregulation is a Defining Feature of Leukemia, and DNA Methylation*** *and chromatin accessibility profiles encode the cell of origin, lineage and clinical behaviour of the disease. Artificial intelligence (AI) and machine learning (ML) can extract predictive patterns from these high-dimensional data, but the published models have not been systematically appraised. We systematically reviewed AI models built on epigenomic data in leukemia for diagnosis, prognosis, relapse and treatment-response prediction, and measurable residual disease (MRD) detection.*

***Methods:*** *This review followed PRISMA 2020 and the SWiM guideline. PubMed was searched on 7 October 2026 with three structured queries, supplemented by a broader PubMed query, citation searching and web searching. Eligible studies were peer-reviewed primary studies that applied an ML or deep-learning algorithm to DNA methylation, chromatin-accessibility or histone-modification data from human leukemia samples and reported a quantitative diagnostic, prognostic, response or MRD outcome. Risk of bias and applicability were assessed with PROBAST, informed by PROBAST+AI. Owing to heterogeneity, results were synthesized narratively and in structured tables, without meta-analysis.*

***Results:*** *Of 68 unique database records and 32 reports from other sources, 30 studies (2010- 2026; 22 since 2020) were included: 16 in acute myeloid leukemia, 4 in acute lymphoblastic leukemia, 2 spanning acute leukemia lineages, 3 in juvenile myelomonocytic leukemia and 5 in chronic lymphocytic leukemia; 10 were pediatric. DNA methylation was the input in 28 studies and chromatin accessibility in 2; no eligible study modelled histone marks. Tree ensembles (n = 9) and support vector machines (n = 5) were the most frequent algorithms. Diagnostic classifiers reported accuracy, concordance or F1 scores of 0.87–1.00, and two classifiers were applied to nanopore sequencing data, one returning acute leukemia subtype calls within about 2 hours. Prognostic epigenomic scores were associated with survival (reported hazard ratios 2.36–10.82), but discrimination fell on transport to independent cohorts (C-index 0.677 internally vs 0.529 in one external cohort). Fifteen studies validated models on an external epigenomic cohort, one prospectively. Overall risk of bias was high in 21 studies, unclear in 8 and low in 1, driven by the analysis domain (small samples, absent calibration, possible data 38 leakage). Calibration was reported in 3 studies and code was available for 11.*

***Conclusions:*** *Epigenomic AI models classify leukemia subtypes with high reported accuracy and can deliver same-day diagnoses, and methylation-based scores add prognostic information. However, the evidence is dominated by retrospective, high-risk-of-bias development studies with limited external validation and almost no prospective or impact evaluation. Prospective, multi-center validation with calibrated, transparently reported models is required before clinical adoption.*

***Registration:*** *This review was not prospectively registered.*

**Keywords:** Leukemia, Epigenomics, DNA Methylation, Chromatin Accessibility, Artificial Intelligence, Machine Learning, Nanopore Sequencing, Systematic Review

## 1. Introduction

### 1.1. Clinical Context

Leukemias are clonal haematopoietic malignancies whose diagnosis, classification and treatment increasingly depend on molecular features [1]. Current classification systems define a large and growing number of genetically specified entities in acute myeloid leukemia (AML) and acute lymphoblastic leukemia (ALL) and risk-adapted therapy relies on integrating cytogenetics, targeted sequencing and measurable residual disease (MRD) [2-6]. Establishing the full diagnostic picture typically requires morphology, multiparameter flow cytometry, karyotyping, fluorescence in situ hybridization and sequencing, which take several days and are unevenly available worldwide [7]. Lineage-ambiguous cases, cryptic rearrangements and patients without a recognisable genetic driver remain difficult to classify and to risk-stratify.

### 1.2. Epigenomic Profiling in Leukemia

Epigenetic deregulation is a hallmark of leukemogenesis: recurrent mutations in *DNMT3A, TET2, IDH1/2* and chromatin regulators reshape the methylome, and leukemia-initiating fusions such as *KMT2A* rearrangements act through chromatin [8,9]. Genome-wide DNA methylation patterns reflect both the cell of origin and the transforming lesion, are chemically stable, can be measured from small amounts of DNA, and are captured natively by nanopore long-read sequencing [9,10]. Chromatin accessibility, measured by the assay for transposase-accessible chromatin (ATAC-seq), reports active regulatory elements and transcription-factor occupancy. These properties make epigenomic data attractive inputs for classifiers: in central nervous system tumours, a methylation-based random-forest classifier changed the diagnosis in up to 12% of prospectively analysed cases and is now widely used in diagnostic neuropathology [11].

### 1.3. Artificial intelligence Methods Applied to Epigenomic Data

Epigenomic datasets typically contain hundreds of thousands of features (CpG sites or accessible regions) and far fewer samples, a setting in which ML methods are used for feature selection, dimensionality reduction, classification and survival prediction [12,13]. The algorithm families encountered in this review, and the clinical questions they address, are summarised in Table 1 and Figure 1.

| Approach | Principle | Typical use in leukemia epigenomics |
|---|---|---|
| Nearest shrunken centroids | Each class is represented by a centroid of shrunken feature means; samples are assigned to the nearest centroid | Compact CpG panels for ALL subtyping [14,15] |
| Support vector machines | Find the maximum-margin boundary between classes, optionally in a kerneltransformed space | Assigning CLL and JMML epigenetic subgroups from few CpGs [16,17] |
| Random / survival forests | Ensembles of decision trees on bootstrapped samples and random feature subsets; survival forests model time-toevent outcomes | Prognostic CpG scores; targeted-assay classifiers [18-20] |
| Gradient boosting | Trees added sequentially to correct residual errors (LightGBM, XGBoost, CoxBoost) | Multiclass AML subtyping; JMML subgroups [21,22] |
| Neural networks and autoencoders | Layered non-linear transformations; autoencoders learn compressed latent representations | Sparse nanopore classification; multi-omics integration [7,23,24] |
| Unsupervised clustering | Groups samples by similarity without labels (k-medoids, consensus, Leiden) | Discovery of epitypes and chromatin subgroups later validated [25,26] |

**Table 1: Machine-Learning Approaches Encountered in the Included Studies**

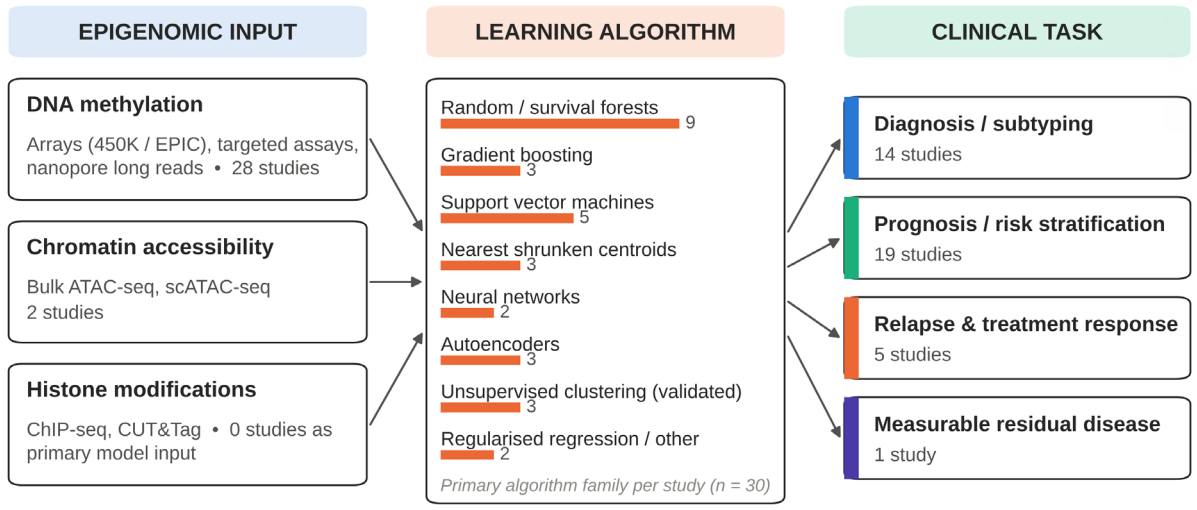



**Figure 1:** Conceptual Framework and Scope of the Review. Epigenomic inputs (left) are modelled with machine-learning algorithms (centre) to address clinical tasks (right). Counts are the number of included studies (n = 30) in each category; the centre panel shows the primary algorithm family of each study.

### 1.4. Rationale and Objectives

Studies applying AI to leukemia epigenomes span two decades, several leukemia types and a wide range of designs, from proof-of-concept classifiers developed on public data to multi-cohort models with real-time nanopore implementations. Their methodological quality, the extent of their validation and their readiness for clinical use have not been systematically appraised. This review aimed to identify all primary studies applying AI or ML to epigenomic data in leukemia and to answer three questions: (i) for which clinical tasks and leukemia types have such models been developed, and with which data and algorithms; (ii) what performance do they report, and in which validation settings; and (iii) what is their risk of bias and applicability, and what stands between them and clinical use.

## 2. Methods

### 2.1. Protocol, Registration and Reporting

The review was conducted and reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement (checklist in Supplementary Table S4) and the Synthesis Without Meta-analysis (SWiM) guideline [27,28]. TRIPOD+AI was used as a reference for the reporting items extracted from each study. The review protocol was not prospectively registered [29].

### 2.2. Eligibility criteria

Eligibility was defined with a PICOTS framework.

- Population: human leukemia samples of any type (AML, ALL, mixed-phenotype or lineageambiguous acute leukemia, chronic lymphocytic leukemia [CLL], chronic myeloid leukemia, juvenile myelomonocytic leukemia [JMML], chronic myelomonocytic leukemia and others) from patients of any age.
- Index model: an AI or ML algorithm explicitly used to develop or validate a model – for example random forest, gradient boosting, support vector machine, neural network, autoencoder, nearest-centroid classifier, or penalised regression framed by the authors as a machine-learning predictive model. Unsupervised clustering qualified only if the resulting subgroups were modelled or validated in an independent cohort.
- Data: DNA methylation (arrays, sequencing, nanopore or targeted assays), chromatin accessibility or histone-modification profiles used as model input or as the prediction target.
- Comparator: any (standard diagnostics, clinical risk scores or none).
- Outcomes: diagnostic classification, prognosis, relapse, treatment response or MRD, with at least one quantitative result.
- Study type: peer-reviewed primary research articles or full peer-reviewed conference papers in English.

Reviews, editorials, book chapters, preprints and conference abstracts without a corresponding full paper were excluded; an abstract describing the same study as an included paper was recorded as a companion report. Studies of lymphoma, myelodysplastic syndromes or solid tumours only, studies of RNA-modification (m6A) regulator expression, studies whose epigenomic data were not measured directly, and pan-cancer studies without leukemia-specific results were excluded.

### 2.3. Information Sources and Search Strategy

PubMed/MEDLINE was searched on 7 October 2026, without date limits, using three queries combining leukemia, methylation and AI terms (Table 2). The search was supplemented by (i) a broader PubMed query ("leukemia AND methylation AND classifier", 385 records), of which the 78 most relevant records by PubMed Best Match ranking could be screened owing to access restrictions at the time of searching; (ii) citation searching of included studies and of an earlier draft of this review; and (iii) targeted web searching for studies of chromatin accessibility and histone modifications, which the database queries did not explicitly cover.

| Query | Search string | Records |
|---|---|---|
| Q1 | leukemia AND methylation AND "machine learning" | 47 |
| Q2 | leukemia AND methylation AND ("deep learning" OR "neural network") | 7 |
| Q3 | leukemia AND methylation AND ("random forest" OR "artificial intelligence" OR "gradient boosting") | 25 |
| | Total before deduplication | 79 |
| | Unique records after removing duplicates by PMID | 68 |
| Supplementary | leukemia AND methylation AND classifier (first 78 of 385 records screened) | 14 reports |
| PubMed applies Automatic Term Mapping, so free-text terms are also matched to MeSH headings | | |

**Table 2: Database Search Strategy (PubMed, 7 October 2026).**

### 2.4. Selection Process

Records were deduplicated by PubMed identifier. Titles and abstracts were screened against the eligibility criteria, and the full texts (or, where the full text was inaccessible, the complete abstract and supplementary information) of potentially eligible reports were assessed. Screening and eligibility assessment were performed by a single reviewer (J.M.) with computational assistance (see Declarations). Every decision and its reason was recorded in a structured log (Supplementary Tables S1 and S3), from which all PRISMA counts were computed.

### 2.5. Data Collection and Data Items

For each study the following items were extracted: first author, year, country, leukemia type and age group, number of patients or samples (training, test and validation), data source, epigenomic platform, AI algorithm, clinical task, validation design, comparator, performance measures exactly as reported (accuracy, concordance, F1 score, area under the receiver operating characteristic curve [AUC], sensitivity and specificity, concordance index [C-index], hazard ratio [HR]), calibration, handling of missing data, availability of code, and the limitations stated by the authors. Validation was classified by the highest level reached: cross-validation only; internal held-out split; external cohort assessed with a non-epigenomic proxy (for example, subtype labels transferred via a transcriptomic classifier); external epigenomic cohort; and prospective application.

### 2.6. Risk-of-bias and Applicability Assessment

Risk of bias and applicability were assessed with the Prediction model Risk of Bias Assessment Tool (PROBAST), applying the AI-specific considerations of PROBAST+AI in particular, data leakage between training and test sets, label derivation from the same data used as predictors, handling of class imbalance and high dimensionality, and the reporting of calibration [30-32]. Each of the four domains (participants, predictors, outcome, analysis) was judged as low, high or unclear risk. Overall risk of bias was rated high if any domain was high, low if all domains were low, and unclear otherwise. Concern regarding applicability to the review question was judged separately. Studies that externally validated a previously published model were assessed in the same way.

### 2.7. Synthesis Methods

Because of marked heterogeneity in leukemia type, task, platform, algorithm, outcome definitions and performance metrics, quantitative pooling was not appropriate. Studies were grouped by clinical question and disease into five themes - diagnostic classification of acute leukemia; prognostic stratification in AML; relapse, treatment response and MRD; JMML; and CLL - and synthesized narratively, with structured tabulation of characteristics and results. Reported performance metrics are displayed graphically by evaluation setting but were not combined. Vote counting based on statistical significance was not used. Reporting biases and the certainty of evidence were not formally assessed (Section 4.7).

## 3. Results

### 3.1. Study Selection

The three database queries returned 79 records; after removal of 11 duplicates, 68 were screened and 33 excluded on title and abstract. Of 35 reports sought, 2 could not be retrieved; 33 were assessed in full and 18 were included. Of 32 further reports identified through the supplementary query, citation searching and web searching, 12 were included. In total, 30 studies were included, together with one companion conference abstract (Figure 2) [33]. Reasons for exclusion of each record and report are given in Supplementary Tables S1 and S3. Several frequently cited studies were excluded at full text because the machine-learning component did not act on epigenomic data: for example, the classifiers of ibrutinib response

in CLL were trained on single-cell transcriptomes rather than on chromatin accessibility an epigenomewide study of pediatric ALL used association statistics without a predictive model and a surface-enhanced Raman spectroscopy MRD assay did not measure methylation directly [34-36].

### 3.2. Characteristics of Included Studies

The 30 included studies were published between 2010 and 2026, 22 of them from 2020 onwards (Figure 3A; Table 3). They originated from the USA (*n* = 7), Sweden (*n* = 5), China (*n* = 5), Germany (n = 4), Japan (*n* = 3), Spain (*n* = 2) and Austria, Italy, the Netherlands and Singapore (n = 1 each). AML was the most studied disease (n = 16), followed by CLL (*n* = 5), ALL (*n* = 4), JMML (*n* = 3) and panels spanning acute leukemia lineages (*n* = 2); ten studies were exclusively pediatric (Figure 4A). The median number of patients or samples analysed was 312 (interquartile range 147–1,050; range 49–3,314; Figure 4C), with frequent reuse of the same public cohorts, notably The Cancer Genome Atlas AML series [37].

DNA methylation was the model input in 28 studies: genome-wide arrays in 11, targeted methylation assays in 8, multi-omics designs including methylation in 6, nanopore sequencing combined with array-trained models in 2, and an older restriction-based array in 1 (Figure 4B). Two studies modelled chromatin accessibility none modelled histone modifications as a primary input [26,38]. Tree ensembles were the most common primary algorithm (*n* =9), followed by support vector machines (*n* =5), gradient boosting, nearest-centroid classifiers, autoencoders and validated unsupervised clustering (*n* = 3 each), and fully connected neural networks (*n* = 2) (Figure 3B). Diagnostic or subtype classification was addressed in 14 studies, prognosis in 19, relapse or treatment response in 5 and MRD in 1 (studies could address more than one task).

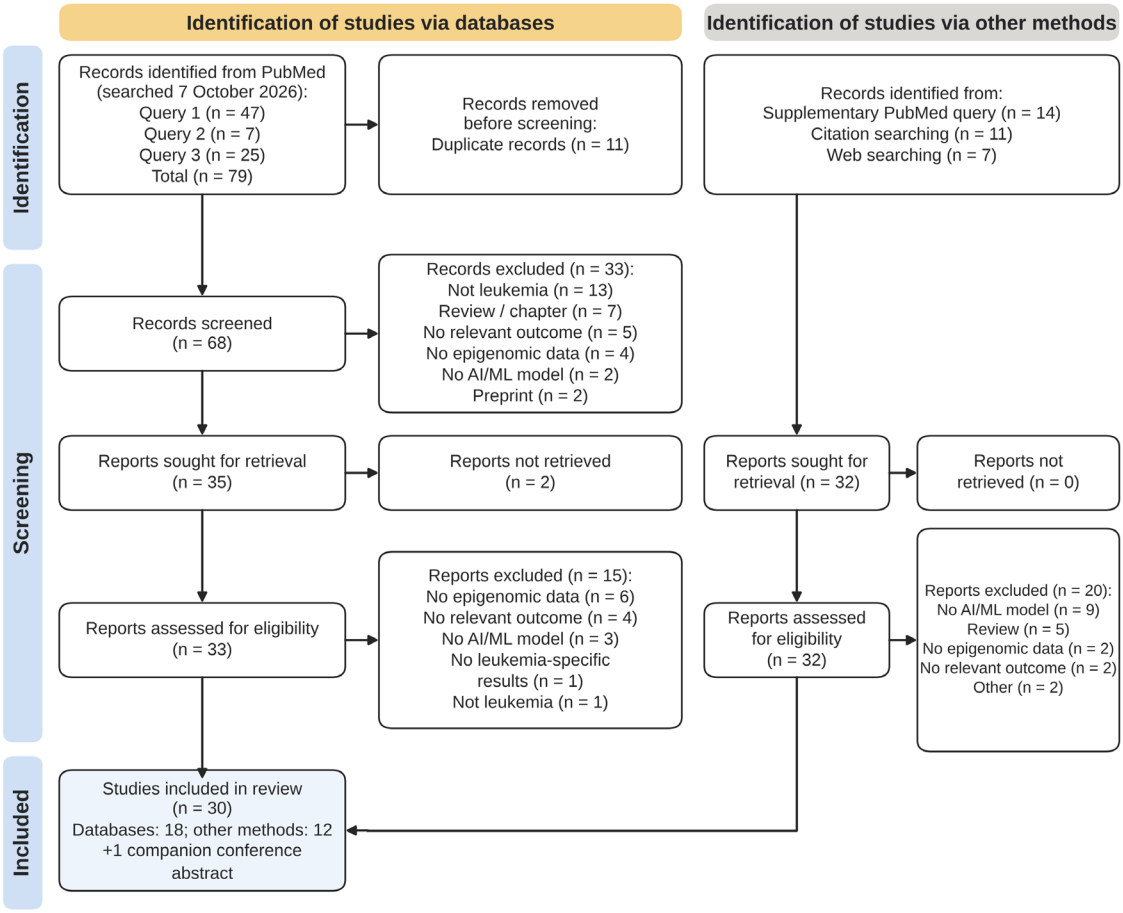



**Figure 2:** PRISMA 2020 flow diagram.Database records were identified with three PubMed queries (Table 2). "Other methods" comprises a broader supplementary PubMed query, citation searching and web searching. The companion report is a conference abstract describing the same study as Steinicke et al. [7,33].

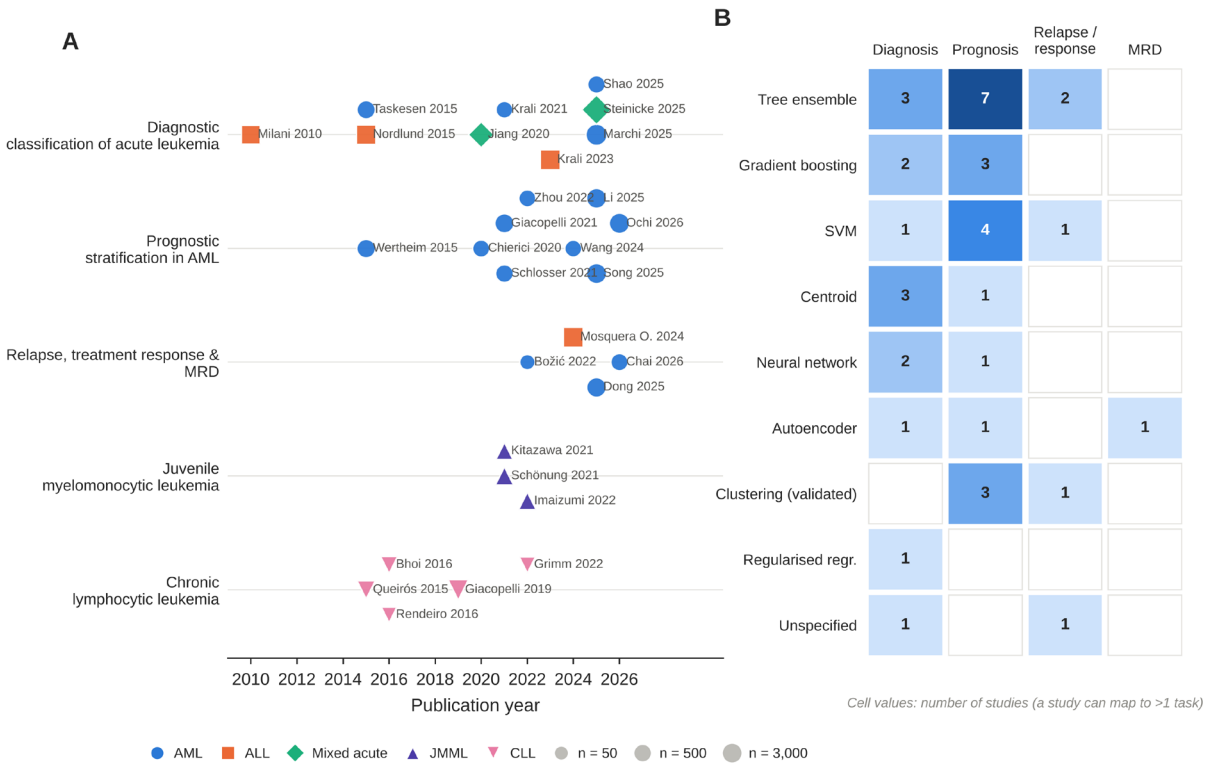



**Figure 3:** Evidence map. (A) Included studies by publication year and thematic group; marker shape and colour indicate leukemia type and marker area the number of patients or samples analysed (logscaled). (B) Number of studies by primary algorithm family and clinical task; a study addressing several tasks is counted in each

### 3.3. Validation, Transparency and Risk of Bias

Fourteen studies evaluated their model in an external epigenomic cohort and one additionally applied it prospectively four validated subgroups in external cohorts only through a transcriptomic proxy, seven relied on an internal held-out split and four on cross-validation alone (Figure 4D) [7]. Calibration was reported in three studies [19,20,39]. Model code or software was publicly available for 11 studies, not available for 15 and could not be confirmed for 4. Overall risk of bias was high in 21 studies, unclear in 8 and low in 1 (Figure 5; Supplementary Table S2) [19]. The analysis domain drove most high ratings ($n$ =20): small development or test sets relative to the number of candidate features absent calibration cross validation without an independent test set, repeated samples from the same patient within cross-validation folds (a source of data leakage) and a reported training error of zero [40-42]. In the outcome domain, three studies were at high risk because class labels were derived by clustering the same methylation data that served as predictors, inflating apparent accuracy [17,22,24]. Participant selection was at high risk in one case–control design that compared

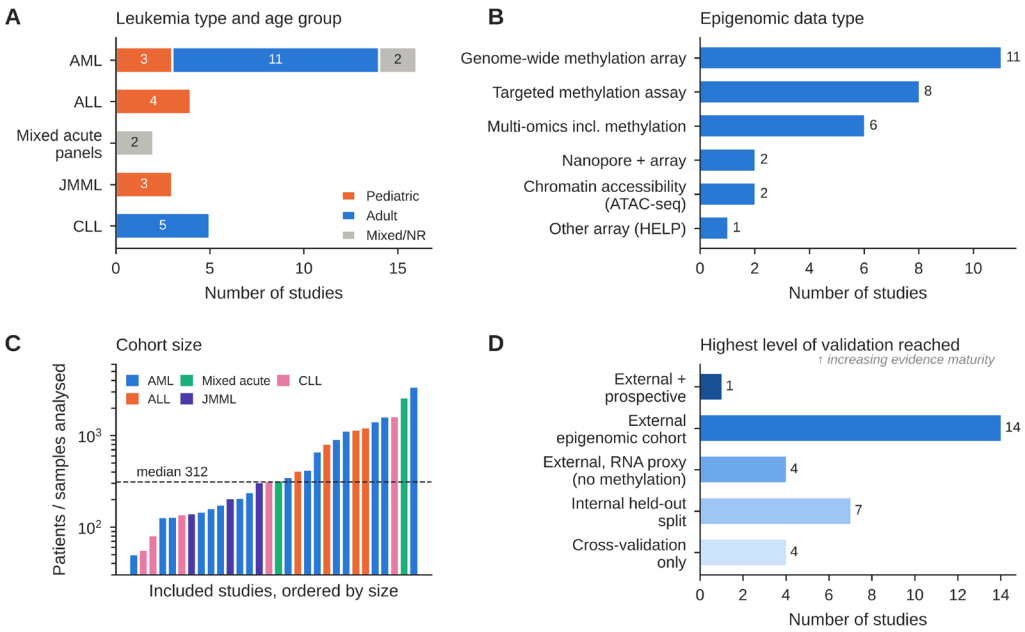



**Figure 4:** Study Characteristics. (A) Leukemia type by age group. (B) Epigenomic data type. (C) Number of patients or samples analysed in each study (log scale), coloured by leukemia type; the dashed line marks the median. (D) Highest level of validation reached by each study

Leukemia samples with healthy controls from a different source and unclear in 15 studies that pooled public datasets without describing inclusion criteria [43]. Applicability concerns were high for a model that used clinical rather than epigenomic predictors and for a model that discriminated diagnostic from relapse samples rather than predicting future relapse [44,45].

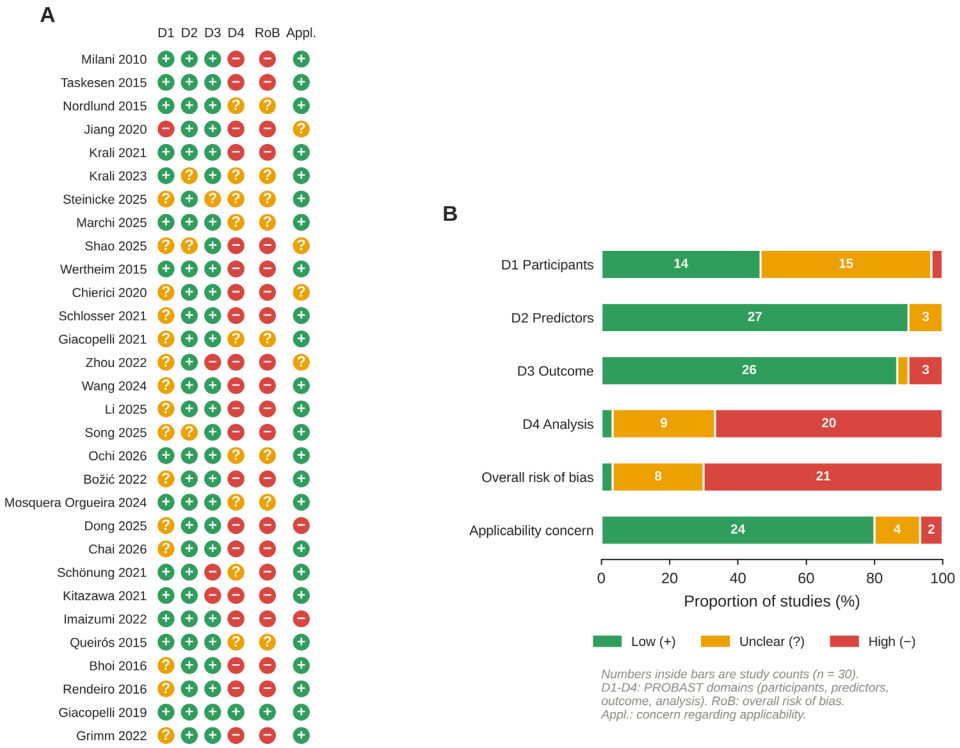



**Figure 5:** Risk of Bias and Applicability (PROBAST, Informed by PROBAST+AI). (A) Domain-level judgements for each study: D1, participants; D2, predictors; D3, outcome; D4, analysis; RoB, overall risk of bias; Appl., concern regarding applicability. (B) Distribution of judgements across studies. Rationales are given in Supplementary Table S2

| Study | Country | Leukemia; age group | Patients / samples | Epigenomic data | AI / ML method | Highest validation |
|---|---|---|---|---|---|---|
| Diagnostic classification of acute leukemia Milani 2010 [46] | Sweden | B-/T-ALL; pediatric | 401 patients | Targeted methylation array (1,320 CpGs, 416 genes) | Supervised multivariate classifiers (external cross-validation) | Cross-validation only |
| Taskesen 2015 [47] | Netherlands | AML; adult | 344 patients | HELP assay + expression arrays | LASSO logistic regression; nearest-mean late integration | Cross-validation only |
| Nordlund 2015 [14] | Sweden | B-/T-ALL; pediatric | 546 design + 39 blinded + 210 un-classified | 450K array | Nearest shrunken centroids (246 CpGs) | Internal held-out (blinded) |
| Jiang 2020 [43] | China | AML, ALL; age mixed or NR | 227 training + 89 validation cases; 754 controls | 450K array; padlock bisulfite sequencing | Nearest shrunken centroids; semi-supervised Cox clustering | Internal held-out (repeated split) |
| Krali 2021 [48] | Sweden | AML; pediatric | 142 samples, 125 patients | 450K array | Nine algorithms compared (neural network best) | Internal held-out |
| Krali 2023 [15] | Sweden | B-/T-ALL; pediatric | 1,131 patients + 3 external sets | 450K array + RNA-seq | One-vs-rest nearest shrunken centroids (ALLIUM) | External cohort |
| Steinicke 2025 [7] | USA | AML, B-/T-ALL, MPAL; mixed | 2,540 reference + 396 external + 39 nanopore | 450K/EPIC arrays; nanopore sequencing | Fully connected neural network with CpG dropout (MARLIN) | External + prospective |
| Marchi 2025 [21] | USA | AML; mixed | 3,314 samples, 11 cohorts; 20 nanopore specimens | 450K/EPIC arrays; nanopore sequencing | PaCMAP + LightGBM (diagnosis); penalised Cox, 38 CpGs (prognosis) (ALMA) | External cohort |
| Shao 2025 [23] | USA | AML (within 5 rare cancers); pediatric | 203 AML among 965 samples | 450K array | Variational autoencoder with transfer learning (RareNet) | Internal held-out |
| Prognostic stratification i Wertheim 2015 [18] | n AML USA | AML; adult | 344 training + 70 validation | xMELP targeted assay (17 loci) | Random forest, 17 loci (Mscore) | External cohort |
| Chierici 2020 [49] | Italy | AML (TCGA); adult | 157 patients | 450K + mRNA + miRNA | Random forest / linear SVM with network-fusion feature ranking | Internal held-out |
| Schlosser 2021 [50] | Germany | AML; adult | 180 training + 55 validation | 450K + expression | Netboost network filter + Cox-Boost | External cohort |
| Giacopelli 2021 [25] | USA | AML; adult | 649 discovery + 67 relapse pairs | 450K/EPIC arrays | k-medoids clustering into 13 epitypes | External cohort |
| Zhou 2022 [24] | USA | AML (TCGA); adult | 143 patients | Methylation + mRNA + miRNA | Incomplete fusion autoencoder | Internal held-out |
| Wang 2024 [51] | China | AML; adult | 126 clustered + external transcriptome set | 450K + RNA + miRNA + | Consensus of 10 multi-omics clustering algorithms; NTP/ PAM | External (nonepigenomic proxy) |
| Li 2025 [39] | China | AML; mixed | 1,103 patients, 5 cohorts (methylation n = 83) | 450K + multi-omics | Multi-omics clustering + random survival forest (82 combinations) | External (nonepigenomic proxy) |
| Song 2025 [52] | China | AML; adult | 1,391 patients across | 450K/850K + RNA + | Multi-omics clustering + random forest 16-gene panel | External (nonepigenomic proxy) |
| Ochi 2026 [26] | Japan | AML; adult | 1,563 ATAC-seq + | Bulk ATAC-seq (+ scATAC, ChIP-seq) | Leiden clustering; nearestcentroid (ClaNC) transfer; | External (nonepigenomic proxy) |
| Božic 2022 [53]′ | Germany | AML; adult | 117 samples, 49 patients; 63 controls | Targeted bisulfite amplicon sequencing | Random forest; autoencoder + k-means | Internal held-out |

| | | | | | | |
|---|---|---|---|---|---|---|
| Mosquera Orgueira | Spain | B-/T-ALL; pediatric | 573 / 190 / 42 / 384 (train / test / 2 external) | 450K array | Random survival forest (16 and 53 CpGs) | External cohort |
| Dong 2025 [45] | China | AML; pediatric | 890 samples (696 diagnosis, 194 relapse) | EPIC array | Boruta/LASSO/LightGBM/ MCFS selection; decision tree, random forest | Cross-validation only |
| Chai 2026 [54] Juvenile myelomonocytic leukemia | Singapore | AML; adult | Beat AML training + 25 local test (146 local) | EPIC array | SVM (RBF) vs random forest, XGBoost, elastic net | External cohort |
| Schönung 2021 [22] | Germany | JMML; pediatric | 255 pooled + 47 independent | 450K/EPIC arrays; MethylSeq | Consensus clustering; XGBoost (124 CpGs) | External cohort |
| Kitazawa 2021 [17] | Japan | JMML; pediatric | 99 discovery + 38 validation | DREAM methylation sequencing | SVM (84 CpGs) | External cohort |
| Imaizumi 2022 [44] Chronic lymphocytic leukemia | Japan | JMML; pediatric | 128 training + 72 validation | Clinical inputs; 450Kdefined labels | Decision tree, SVM, naive Bayes | External cohort |
| Queirós 2015 [16] | Spain | CLL; adult | 211 initial + 97 validation | 5-CpG pyrosequencing (450K-defined labels) | Support vector machine | External cohort |
| Bhoi 2016 [55] | Sweden | CLL; adult | 135 patients | 5-CpG pyrosequencing | Published SVM (external validation) | External cohort |
| Rendeiro 2016 [38] | Austria | CLL; adult | 88 samples, 55 patients | ATAC-seq | Random forest | Cross-validation only |
| Giacopelli 2019 [19] | USA | CLL; adult | 139 / 166 / 1,286 (train / validation / 4 test cohorts) | MassARRAY iPLEX (34 CpGs) | Random forest with calibrated probabilities | External cohort |
| Grimm 2022 [56] | Germany | CLL; adult | 79 patients | 5-CpG pyrosequencing | Published SVM (external validation) tion) | External cohort |

ALL, acute lymphoblastic leukemia; AML, acute myeloid leukemia; ATAC-seq, assay for transposase-accessible chromatin with sequencing; CLL, chronic lymphocytic leukemia; DREAM, digital restriction enzyme analysis of methylation; EPIC/450K/850K, Illumina Infinium methylation arrays; HELP, HpaII tiny fragment enrichment by ligation-mediated PCR; JMML, juvenile myelomonocytic leukemia; LASSO, least absolute shrinkage and selection operator; MPAL, mixed-phenotype acute leukemia; NR, not reported; NTP, nearest template prediction; PAM, partitioning around medoids; SVM, support vector machine; xMELP, multiplex methylation-specific ligation-mediated PCR.

**Table 3: Characteristics of the 30 Included Studies**

### 3.4. Diagnostic Classification of Acute Leukemia

Nine studies developed classifiers for acute leukemia subtypes or lineage. The pediatric ALL work of the Nordic Society of Paediatric Haematology and Oncology (NOPHO) illustrates the evolution of the field. Milani et al. first showed in 401 children that methylation at 1,320 CpG sites discriminated cytogenetic subtypes and that methylation of 20 genes was associated with relapse [46]. Nordlund et al. then trained a nearest-shrunken-centroid classifier on 246 CpGs from 450K arrays in 546 patients, achieving a mean sensitivity of 0.90 and specificity of 0.99 across subtypes and correctly classifying 36 of 39 blinded samples; 106 of 210 patients with undefined karyotype or non-recurrent aberrations showed methylation profiles highly similar to known subtypes [14]. Krali et al. extended this approach to 17 molecular subtypes in 1,131 patients (ALLIUM), with 87.0% concordance between the methylation classifier and the reference subtype overall and 93.8% in the 242 patients who also had RNA-sequencing data, in whom the expression classifier reached 95.9% [15]. In pediatric AML, nine algorithms were compared on 450K data, and the best-performing neural network reached a test accuracy of 91% for cytogenetic subtype [48].

The two largest studies paired array-trained models with nanopore sequencing to enable rapid diagnosis. Steinicke et al. trained a fully connected neural network (MARLIN) on 2,540 acute leukemia and control samples, randomly dropping 99% of CpG features in each training epoch so that the model would tolerate the sparse, random coverage of shallow nanopore sequencing [7]. MARLIN reached a median F1 score of 0.91 for 38 methylation classes in cross-validation, and on nanopore data it agreed with the reference diagnosis in 25 of 26 cases (96.2%) with a high-confidence prediction. Results were typically available within 2

hours of sample receipt, whereas conventional integrated work-up takes days (3 and 4 days in two cases described by the authors), and the first prospective case required 56 minutes of processing and 40 minutes of sequencing (Figure 7). Marchi et al. assembled 3,314 samples from 11 cohorts to train a LightGBM classifier of AML subtypes (ALMA), with weighted F1 scores of 0.924 and 0.925 in two held-out pediatric trial cohorts, and applied it to 20 nanopore specimens with sequencing started 2 hours after specimen acquisition [21]. Both groups noted that rare entities were 239 under-represented in the training data; ALMA, for example, misclassified FUS::ERG and Down syndrome-associated AML more often than common subtypes [21].

Earlier, smaller studies reported near-perfect discrimination that is unlikely to transfer to clinical practice. A nearest-shrunken-centroid model distinguished AML and ALL from normal blood with AUCs of 0.9998 and 0.9995 using four and seven CpGs, but in a design that compared 244 patients with healthy controls from a different source a LASSO model on restriction-based methylation data predicted inv(16) AML with an AUC of 0.9995 under cross-validation only and a variational autoencoder trained across five rare cancers classified all 11 pediatric AML test samples correctly [23,43,47]. The reported diagnostic performance of all studies is displayed by 248 metric and evaluation setting in Figure 6.

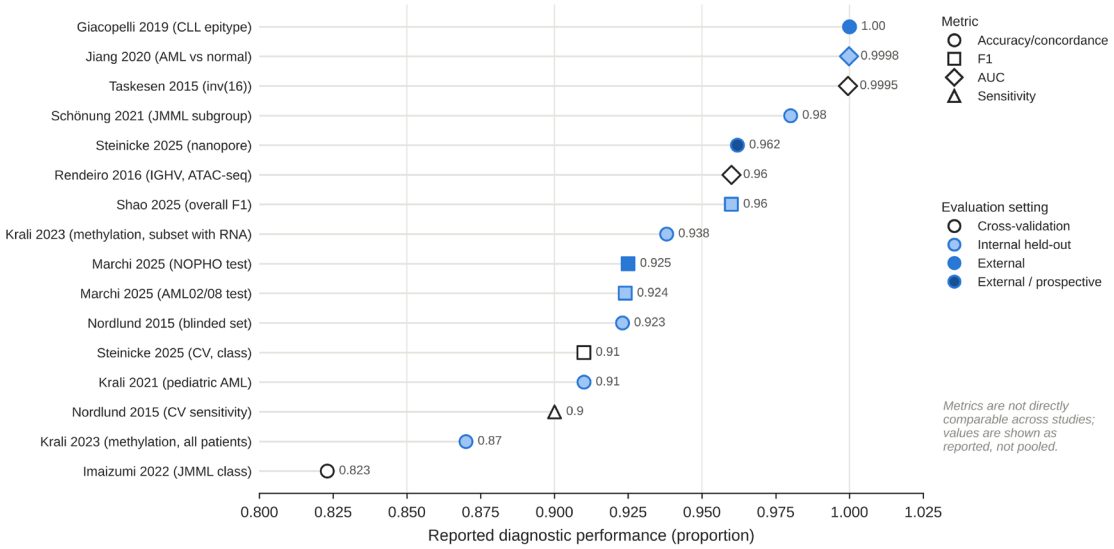



**Figure 6:** Reported Diagnostic Performance of Epigenomic AI Classifiers. Each marker shows one performance estimate as reported by the authors; marker shape indicates the metric and fill the evaluation setting (open, cross-validation; light, internal held-out; dark, external or external and prospective). Metrics differ in meaning and are not pooled. The highest values were obtained under cross-validation or case–control designs

### 3.5. Prognostic Stratification in AML

Nine studies modelled prognosis in AML. The earliest implemented a 17-locus methylation assay (xMELP) with a random-forest classifier trained on 344 patients; the resulting score separated mortality risk in an independent cohort of 70 patients (P = 0.009) [18]. In the largest prognostic analysis, an epigenomic risk score and a penalised Cox signature of 38 CpGs, both derived within the ALMA framework, were associated with overall survival in the held-out AML02/08 pediatric trial cohort (n = 200; HR 4.20, 95% CI 2.36–7.45; and HR 3.23, 95% CI 1.75–5.97) and were evaluated in adult cohorts (Figure 8A) [21].

Unsupervised approaches identified biologically coherent subgroups with prognostic value. k- medoids clustering of 649 adult AML methylomes defined 13 “epitypes”, of which three had inferior overall survival in two cohorts independently of the LSC17 stemness score epitypes were stable between diagnosis and relapse in 85–95% of paired samples [25,57]. Ochi et al. profiled chromatin accessibility in 1,563 patients and defined 16 subgroups whose addition to European LeukemiaNet (ELN) risk improved the C-index in bootstrap analyses; external validation in 1,079 patients was performed through a transcriptomic nearest-centroid proxy because ATAC-seq data were not available in those cohorts [26].

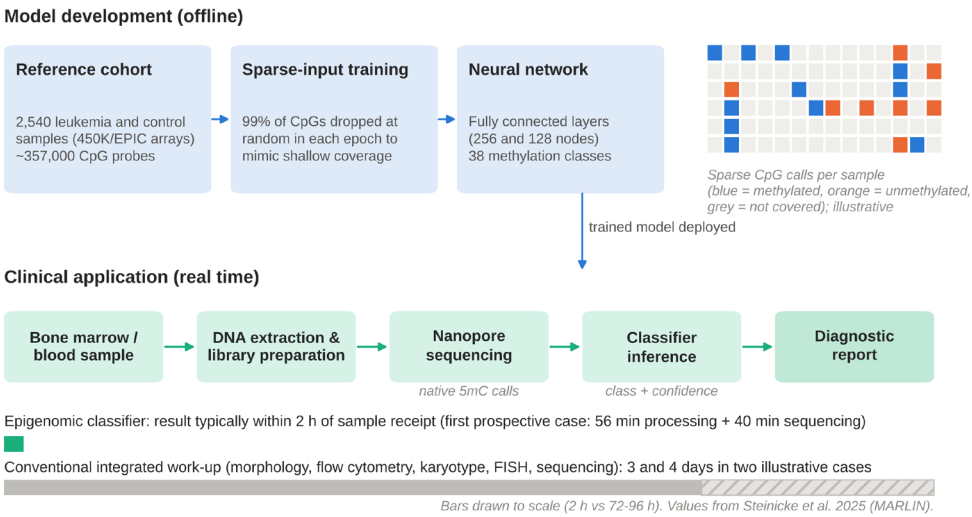



**Figure 7:** Real-time epigenomic classification of acute leukemia with nanopore sequencing. A neural network is trained offline on array methylation profiles with random feature dropout so that it tolerates the sparse coverage of shallow nanopore runs; in the clinic, native methylation calls from a short nanopore run are classified within about 2 hours, whereas conventional integrated work-up took 3 and 4 days in two cases described by the authors. The bars at the bottom are drawn to scale. Values from Steinicke et al. Marchi et al. report a comparable workflow [7,21]

The remaining six studies integrated methylation with other omics, mostly using public TCGA data [24,39,49-52]. Their reported performance was moderate - for example, a Matthews correlation coefficient of 0.18-0.26 on held-out data, an AUC of 0.75 and C-indices of 0.72–0.78 and in three of them methylation contributed only to subgroup discovery in fewer than 130 patients, whereas external validation used transcriptomic data alone [24,39,49,51,52]. A 16-gene panel derived from a multi-omics classification reached a C-statistic of 0.87, not significantly different from the 0.82 of ELN 2022 (P =0.162) [4,52].

### 3.6. Relapse, Treatment Response and MRD

Four studies addressed relapse, response or MRD. In pediatric ALL, random survival forests on 450K data predicted relapse (16 CpGs) and death (53 CpGs) in a development cohort of 763 patients with two external cohorts of 42 and 384 patients [20]. The relapse model achieved a C-index of 0.677 in the internal test set and 0.667 in a Canadian cohort, but only 0.529 - close to chance - in a Nordic cohort treated on different protocols; the mortality model fell from 0.754 to 0.621 (Figure 8B). In adult AML, a support vector machine trained on Beat

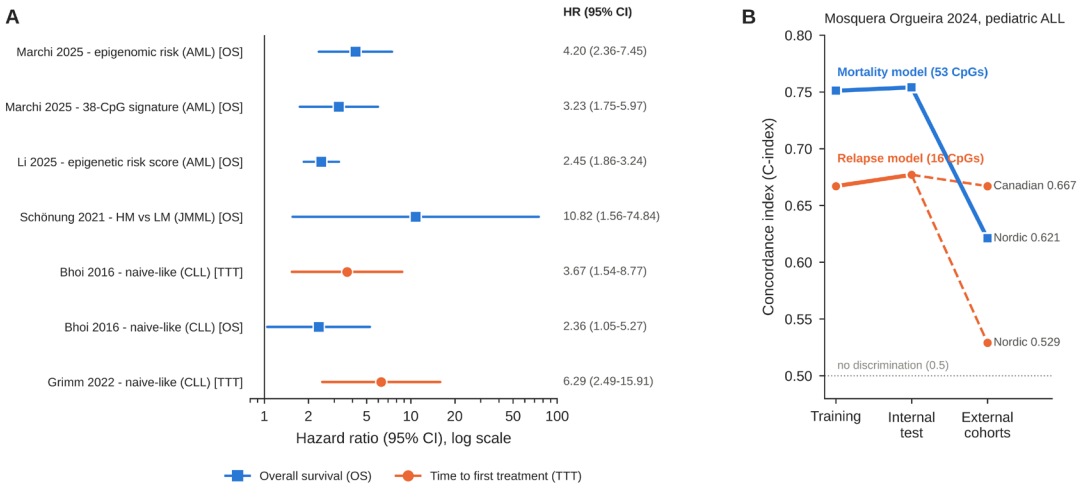



**Figure 8:** Prognostic performance. (A) Hazard ratios with 95% confidence intervals for epigenomic risk groups or scores, as reported (log scale; OS, overall survival; TTT, time to first treatment). Hazard ratios compare high-risk with low-risk groups and are not directly comparable across studies. (B) C-index of the relapse and mortality models of Mosquera Orgueira et al. in the training set, the internal test set and two external cohorts, illustrating loss of discrimination on transport [20]

AML methylomes predicted primary refractoriness to induction chemotherapy with an AUROC of 0.86 and, in 25 local patients, a sensitivity of 1.00 and specificity of 0.81 [54]. A pediatric AML study reported F1 scores of about 0.99 for distinguishing diagnostic from relapse samples because diagnosis and relapse samples may have come from the same patients and the model classifies sample state rather than predicting future relapse, this result does not establish relapse prediction [45]. The only MRD study combined targeted bisulfite amplicon sequencing with a random forest and an autoencoder: the approach flagged 76% of MRD-positive samples but also 71% of MRD-negative samples as abnormal, and methylation scores declined towards background at about 5% malignant DNA [53].

### 3.7. Juvenile Myelomonocytic Leukemia

In JMML, methylation subgroups are established prognostic markers, and ML has been used to make them clinically measurable. An international consortium defined high-, intermediate and low-methylation subgroups by consensus clustering of 255 pooled cases and trained an XGBoost classifier on 124 CpGs that reached 98% accuracy in a held-out set and 95% concordance in an independent cohort; the high-methylation group had worse overall survival in multivariable analysis (HR 10.82, 95% CI 1.56–74.84) [22]. A support vector machine applied to a sequencing-based methylation assay reproduced the array classification and separated 5-year overall survival in a validation cohort of 38 patients (26.3% vs 80.5%, P = 0.024) [17]. A third study predicted the methylation class from routine clinical parameters with 82–84% 298 concordance, offering a surrogate where methylation profiling is unavailable [44]. Because the subgroup labels were derived from the same methylation data, the reported accuracies measure reproducibility of the clustering rather than agreement with an independent reference standard.

### 3.8. Chronic Lymphocytic Leukemia

In CLL, a five-CpG support vector machine assigning patients to naive-like, intermediate and memory-like epigenetic subgroups was developed in 211 patients and validated in 97 and was subsequently validated by two independent groups, in which the naive-like subgroup was associated with shorter time to first treatment (HR 3.67 and 6.29) [16,55,56]. A related random forest classifier applied to a targeted Mass ARRAY assay was tested in four independent cohorts totalling 1,286 patients and predicted time to first treatment, progression and survival after chemoimmunotherapy or ibrutinib; it was the only study judged at low risk of bias [19]. Using ATAC-seq of 88 samples, a random forest predicted immunoglobulin heavy-chain variable region (IGHV) mutation status with an AUC of 0.96 under leave-one-out cross-validation [38].

| Study | Task | Principal reported result | Code |
|---|---|---|---|
| *Diagnostic classification of acute leukemia* Milani 2010 [46] | Diagnosis; relapse | Methylation of 40 genes discriminated cytogenetic subtypes; 20 genes associated with relapse | NR |
| Taskesen 2015 [47] | Diagnosis (15 subtypes) | Methylation-only AUC 0.9995 for inv(16); integration best in 14/15 subtypes | No |
| Nordlund 2015 [14] | Diagnosis (subtypes) | Mean sensitivity 0.90, specificity 0.99; blinded set 36/39 (92%) correct | Yes |
| Jiang 2020 [43] | Diagnosis; prognosis | AUC 0.9998 (AML vs normal, 4 CpGs); 0.9995 (ALL vs normal, 7 CpGs) | No |
| Krali 2021 [48] | Diagnosis; prognosis | Test accuracy 91% for cytogenetic subtype | Yes |
| Krali 2023 [15] | Diagnosis (17 subtypes) | Methylation 87.0% concordance overall; 93.8% in 242 patients with RNA data (expression 95.9%) | Yes |
| Steinicke 2025 [7] | Diagnosis (38 classes) | CV median F1 0.91 (class); nanopore 25/26 (96.2%) concordant; result <2 h | NR |
| Marchi 2025 [21] | Diagnosis; prognosis | Test weighted F1 0.924–0.925; epigenomic risk HR 4.20 (2.36–7.45) | Yes |
| Shao 2025 [23] | Diagnosis | 11/11 AML test samples correct; overall F1 0.96 | Yes |
| *Prognostic stratification in AML* Wertheim 2015 [18] | Prognosis (OS) | Risk separation in validation cohort, P = 0.009 | Yes |
| Chierici 2020 [49] | Prognosis (vital status) | Test MCC 0.18–0.26 | Yes |
| Schlosser 2021 [50] | Prognosis (OS) | Improved prediction error vs WGCNA (P = 0.041); chromatin module replicated (P = 0.04) | Yes |
| Giacopelli 2021 [25] | Subtyping; prognosis | Epitypes E11-E13 inferior OS (P < 0.001) independent of LSC17 | No |
| Zhou 2022 [24] | Prognosis (survival subtype) | AUC 0.75 | Yes |
| Wang 2024 [51] | Prognostic subtyping | 4 subtypes; OS log-rank P = 0.006 | No |
| Li 2025 [39] | Prognosis (OS) | C-index 0.72–0.78; risk score HR 2.45 (1.86–3.24) | No |

| | | | |
|---|---|---|---|
| Song 2025 [52] | Prognostic subtyping | C-statistic 0.87 vs 0.82 for ELN 2022 (P = 0.162 | Yes |
| Ochi 2026 [26] | Subtyping; prognosis; drug sensitivity | 16 chromatin subgroups; adding them to ELN raised C-index (bootstrap) | Yes |
| *Relapse, treatment resp* Božic 2022 [53]′ | *onse and MRD MRD* | Autoencoder flagged 76% of MRD+ but also 71% of MRD- samples as abnormal | No |
| Mosquera Orgueira 2024 [20] | Relapse; mortality | Relapse C-index 0.677 test, 0.667 and 0.529 external | No |
| Dong 2025 [45] | Relapse (sample state) | F1 0.99 (diagnosis vs relapse samples) | No |
| Chai 2026 [54] | Treatment response (refractory) | AUROC 0.86; local test sensitivity 1.00, specificity 0.81 | No |
| Juvenile myelomonocytic Schönung 2021 [22] | leukemia Subgroup classification; prognosis | Test accuracy 98%; HM vs LM HR 10.82 (1.56–74.84) | No |
| Kitazawa 2021 [17] | Prognosis | Validation 5-y OS 26.3% vs 80.5% (P = 0.024) | NR |
| Imaizumi 2022 [44] | Methylation class; prognosis | 82.3–83.6% concordance with array class | No |
| *Chronic lymphocytic leu* Queirós 2015 [16] | kemia Subgroup classification; prognosis | Subgroups differ in TTT and OS (P < 0.001) | No |
| Bhoi 2016 [55] | Prognosis | TTT HR 3.67 (1.54–8.77), naive-like | No |
| Rendeiro 2016 [38] | Diagnosis (IGHV status) | AUC 0.96; sensitivity 95.6%, specificity 88.2% | NR |
| Giacopelli 2019 [19] | Epitype classification; prognosis | 100% validation accuracy; TTT P < 0.0001 | No |
| Grimm 2022 [56] | Prognosis | TTT HR 6.29 (2.49–15.91), naive-like | No |

AUC, area under the receiver operating characteristic curve; CV, cross-validation; ELN, European LeukemiaNet; F1, harmonic mean of precision and recall; HM/LM, high/low methylation; HR, hazard ratio (95% CI); MCC, Matthews correlation coefficient; MRD, measurable residual disease; NR, not reported or not confirmed; OS, overall survival; TTT, time to first treatment.

**Table 4: Clinical task, Principal Reported Results and Code Availability. Results are Quoted as Reported by the Authors in the Setting Described in Table 3**

## 4. Discussion

### 4.1. Summary of Evidence

This review identified 30 studies applying AI to leukemia epigenomes. Four findings stand out.

First, DNA methylation dominates: chromatin accessibility was modelled in only two studies and histone modifications in none. Second, diagnostic classifiers consistently report high accuracy, and two recent studies demonstrate that array-trained models can classify sparse nanopore methylation data, with results available within about 2 hours in one of them. Third, prognostic epigenomic scores are associated with survival in AML, JMML and CLL, but discrimination declines on transport to independent populations. Fourth, the evidence base is methodologically immature: 21 of 30 studies were at high risk of bias, only one model was applied prospectively, calibration was rarely reported, and no study assessed the effect of an epigenomic AI model on clinical decisions or patient outcomes.

### 4.2. Why DNA Methylation Suits Machine Learning

Methylation profiles are well suited to supervised learning for several reasons that emerge across the included studies. They are stable, quantitative and largely binary at the single molecule level; they encode the developmental state and cell of origin of the leukemic clone, so that a single assay captures information that otherwise requires several genetic tests and they can be read directly from native DNA by nanopore sequencing. The dropout-based training of MARLIN shows how an architecture can be designed around these properties, and parallels the trajectory of methylation classification in central nervous system tumours [7,9,10,11]. Epigenomic subgroups were also biologically interpretable: CLL subgroups mirror the maturation stage of the normal B cell of origin, AML epitypes and chromatin subgroups align with myeloid differentiation stages, and JMML methylation classes track clinical aggressiveness] [16,19,22,25,26]. This interpretability partly offsets concerns about “black-box” models.

### 4.3. Methodological Issues

The risk-of-bias assessment highlights recurrent problems that are characteristic of high dimensional ML studies [29,32].

Incorporation bias. When classes are first defined by clustering methylation data and a classifier is then trained to recover them from the same data, high accuracy is expected by construction [17,22,24]. Such models are valuable for standardising a classification across platforms, but their accuracy does not measure agreement with an independent reference standard; their clinical value rests on the prognostic validation of the subgroups.

Spectrum bias and optimism. Near-perfect AUCs arose in case–control comparisons with healthy controls or under cross-validation alone [38,43,47]. Where models were evaluated on independent data, performance was lower and more variable (Figures 6 and 8).

Data leakage. Repeated samples from the same patient may have

been present within cross validation in at least two studies, and feature selection outside the cross-validation loop was not always excluded; both inflate apparent performance [38,42,45].

Transportability. The drop of a relapse model from a C-index of 0.677 to 0.529 in a cohort treated on different protocols shows that methylation-based risk depends on treatment context. Prognostic models therefore require validation in each therapeutic setting in which they are intended for use [20].

Proxy validation. Four studies validated epigenomically defined subgroups in external cohorts only through transcriptomic surrogates [26,39,51,52]. This supports the biological reproducibility of the subgroups but not the performance of the epigenomic assay itself.

Calibration, sample size and batch effects. Only three studies reported calibration, which is essential if predicted probabilities are to guide decisions few justified their sample size and models trained on pooled array data from different laboratories and array versions are vulnerable to batch effects which several studies did not address [40,41,58].

### 4.4. Comparison with Previous Reviews and Other Omics

Previous reviews of AI and epigenetics have been narrative and spanned many cancers or haematological malignancies [12,13]. To our knowledge, this is the first systematic review with structured risk-of-bias assessment focused on AI models built on leukemia epigenomes. Transcriptomic classifiers, including low-cost nanopore RNA sequencing address similar diagnostic questions and are often more mature in ALL in that study the expression classifier was slightly more concordant than the methylation classifier in the same patients (95.9% vs 93.8%) [15,59]. Methylation may be preferable where RNA quality is limited or where cell-of-origin information is needed, and integration rather than competition between the two layers is likely to be most informative. Related epigenomic approaches that fell outside our eligibility criteria a pan-B-cell methylation classifier and epigenetic mitotic cloc a pan-haematolymphoid methylation classifier and Bayesian inference of clonal evolution from fluctuating methylation - point to the same trajectory [60-62].

### 4.5. Translational Implications

The most mature application is rapid diagnosis of acute leukemia. A methylation classifier delivering a provisional subtype within hours could accelerate time-critical treatment decisions, help resolve lineage-ambiguous cases, and could support diagnosis in settings without access to the full panel of conventional tests. In CLL, the targeted assay of Giacopelli et al. is the closest to a deployable prognostic test, having been validated in multiple independent cohorts with calibrated outputs [19]. In JMML, methylation subgroups have reached international consensus [22]. By contrast, MRD detection and prediction of treatment response remain at an exploratory stage [53,54]. In every case, prospective studies measuring concordance with integrated diagnosis, turnaround time, cost and effect on management are still needed.

### 4.6. Limitations of the Evidence

The included studies were predominantly retrospective, reused a small number of public datasets, under-represented rare subtypes and non-European populations, and used heterogeneous platforms, algorithms and metrics, which precluded pooling. Only one model was applied prospectively, in five patients and none was evaluated for clinical impact [7].

### 4.7. Limitations of the Review Process

This review has several limitations. First, only one bibliographic database (PubMed) was searched systematically; Embase, Web of Science and Scopus were not searched, and the supplementary PubMed query could be screened only for its 78 most relevant records owing to access restrictions at the time of searching. Second, the database queries required methylation terms, so studies of chromatin accessibility or histone modifications were identified only through citation and web searching. Third, screening, data extraction and risk-of-bias assessment were performed by a single reviewer with computational assistance, without independent duplicate assessment. Fourth, two potentially eligible reports could not be retrieved, and for some studies data were extracted from the abstract and accessible supplementary material rather than the full text. Fifth, PROBAST was applied to clustering-based and multi-omics studies for which it is only partly designed. Sixth, preprints and non-English publications were excluded. Finally, certainty of evidence was not graded and reporting biases were not assessed. The protocol was not prospectively registered.

### 4.8. Implications for Research

Table 5 maps the shortcomings identified in this review to concrete recommendations for future studies.

| Shortcoming observed | Recommendation |
|---|---|
| External validation in 15 of 30 studies; prospective application in 1 | Prospective, multi-center validation in the intended clinical setting, with prespecified endpoints |
| Calibration reported in 3 studies | Report calibration (calibration plots, slope and intercept) together with discrimination [41] |
| Labels derived from the same data; case–control designs | Use reference standards independent of the predictors and consecutive clinical cohorts |
| Repeated samples per patient within cross-validation | Split data at the patient level; keep feature selection inside the cross-validation loop [42] |
| Code available for 11 studies | Release code, trained models and data dictionaries; follow TRI- POD+AI [29] |
| No study of histone marks; two of chromatin accessibility | Extend modelling to chromatin and histone layers and to multiomics integration |
| Rare subtypes misclassified; few non-European cohorts | Build shared, diverse reference cohorts, including rare and pediatric entities |
| No study of clinical impact | Assess turnaround time, cost, decision impact and patient outcomes |

**Table 5: Recommendations for future Studies of Epigenomic AI Models in Leukemia**

## 5. Conclusions

AI models built on leukemia epigenomes - above all on DNA methylation - can classify leukemia subtypes with high reported accuracy, deliver same-day diagnoses when combined with nanopore sequencing, and provide prognostic information in AML, JMML and CLL. The evidence nevertheless consists largely of retrospective development studies at high risk of bias, with limited external and almost no prospective validation, infrequent calibration and no assessment of clinical impact. Prospective, multi-center studies of calibrated and transparently reported models, extended to chromatin and histone layers and to under-represented populations, are needed before epigenomic AI can be incorporated into routine leukemia care.

## References


1. Tebbi, C. K. (2021). Etiology of acute leukemia: A review. *Cancers, 13*(9), 2256.
2. Khoury, J. D., Solary, E., Abla, O., Akkari, Y., & Alaggio, R., et al. (2022). The 5th edition of the World Health Organization classification of haematolymphoid tumours: myeloid and histiocytic/dendritic neoplasms. *leukemia, 36*(7), 1703-1719.
3. Arber, D. A., Orazi, A., Hasserjian, R. P., Borowitz, M. J., & Calvo, K. R., et al. (2022). International Consensus Classification of Myeloid Neoplasms and Acute Leukemias: integrating morphologic, clinical, and genomic data. *Blood, The Journal of the American Society of Hematology, 140*(11), 1200-1228.
4. Döhner, H., Wei, A. H., Appelbaum, F. R., Craddock, C., & DiNardo, C. D., et al. (2022). Diagnosis and management of AML in adults: 2022 recommendations from an international expert panel on behalf of the ELN. *blood, 140*(12), 1345-1377.
5. Campana, D. (2010). *Minimal residual disease in acute lymphoblastic leukemia*. Hematology 2010, the American Society of Hematology Education Program Book, 2010(1), 7-12.
6. Voso, M. T., Ottone, T., Lavorgna, S., Venditti, A., & Maurillo, L., et al. (2019). MRD in AML: the role of new techniques. *Frontiers in oncology, 9*, 448917.
7. Steinicke, T. L., Benfatto, S., Capilla-Guerra, M. R., Monteleone, A. B., & Young, J. H., et al. (2025). Rapid epigenomic classification of acute leukemia. *Nature genetics, 57*(10), 2456-2467.
8. Florean, C., Schnekenburger, M., Grandjenette, C., Dicato, M., & Diederich, M. (2011). Epigenomics of leukemia: from mechanisms to therapeutic applications. *Epigenomics, 3*(5), 581-609.
9. Dawson, M. A. (2017). The cancer epigenome: Concepts, challenges, and therapeutic opportunities. *Science, 355*(6330), 1147-1152.
10. Euskirchen, P., Bielle, F., Labreche, K., Kloosterman, W. P., & Rosenberg, S., et al. (2017). Same-day genomic and epigenomic diagnosis of brain tumors using real-time nanopore sequencing. *Acta neuropathologica, 134*(5), 691-703.
11. Capper, D., Jones, D. T., Sill, M., Hovestadt, V., & Schrimpf, D., et al. (2018). DNA methylation-based classification of central nervous system tumours. *Nature, 555*(7697), 469-474.
12. Hamamoto, R., Komatsu, M., Takasawa, K., Asada, K., & Kaneko, S. (2019). Epigenetics analysis and integrated analysis of multiomics data, including epigenetic data, using artificial intelligence in the era of precision medicine. *Biomolecules, 10*(1), 62.
13. Lauzon-Young, C., Silva, A., & Sadikovic, B. (2025). Epigenomic insights and computational advances in hematologic malignancies. *Molecular cytogenetics, 18*(1), 9.
14. Nordlund, J., Bäcklin, C. L., Zachariadis, V., Cavelier, L., & Dahlberg, J., et al. (2015). DNA methylation-based subtype prediction for pediatric acute lymphoblastic leukemia. *Clinical epigenetics, 7*(1), 11.
15. Krali, O., Marincevic-Zuniga, Y., Arvidsson, G., Enblad, A.

P., & Lundmark, A., et al. (2023). Multimodal classification of molecular subtypes in pediatric acute lymphoblastic leukemia. *NPJ precision oncology, 7*(1), 131.
16. Queirós, A. C., Villamor, N., Clot, G., Martinez-Trillos, A., & Kulis, M., et al. (2015). A B-cell epigenetic signature defines three biologic subgroups of chronic lymphocytic leukemia with clinical impact. *Leukemia, 29*(3), 598-605.
17. Kitazawa, H., Okuno, Y., Muramatsu, H., Aoki, K., & Murakami, N., et al. (2021). Simple and robust methylation test for risk stratification of patients with juvenile myelomonocytic leukemia. *Blood Advances, 5*(24), 5507-5518.
18. Wertheim, G. B., Smith, C., Luskin, M., Rager, A., & Figueroa, M. E., et al. (2015). Validation of DNA methylation to predict outcome in acute myeloid leukemia by use of xMELP. *Clinical chemistry, 61*(1), 249-258.
19. Giacopelli, B., Zhao, Q., Ruppert, A. S., Agyeman, A., & Weigel, C., et al. (2019). Developmental subtypes assessed by DNA methylation-iPLEX forecast the natural history of chronic lymphocytic leukemia. *Blood, The Journal of the American Society of Hematology, 134*(8), 688-698.
20. Mosquera Orgueira, A., Krali, O., Pérez Míguez, C., Peleteiro Raíndo, A., & Díaz Arias, J. Á., et al. (2024). Refining risk prediction in pediatric acute lymphoblastic leukemia through DNA methylation profiling. *Clinical Epigenetics, 16*(1), 49.
21. Marchi, F., Shastri, V. M., Marrero, R. J., Nguyen, N. H., & Oettl, A., et al. (2025). Epigenomic diagnosis and prognosis of Acute Myeloid Leukemia. *Nature communications, 16*(1), 6961.
22. Schönung, M., Meyer, J., Nöllke, P., Olshen, A. B., & Hartmann, M., et al. (2021). International consensus definition of DNA methylation subgroups in juvenile myelomonocytic leukemia. *Clinical cancer research, 27*(1), 158-168.
23. Shao, D., Addagudi, S., Cowles, J., Jain, A., & D'Souza, L., et al. (2025). Rarenet: a deep learning model for rare cancer diagnosis. *Scientific Reports, 15*(1), 22732.
24. Zhou, K., Kottoori, B. S., Munj, S. A., Zhang, Z., & Draghici, S., et al. (2022). Integration of multimodal data from disparate sources for identifying disease subtypes. *Biology, 11*(3), 360.
25. Giacopelli, B., Wang, M., Cleary, A., Wu, Y. Z., & Schultz, A. R., et al. (2021). DNA methylation epitypes highlight underlying developmental and disease pathways in acute myeloid leukemia. *Genome research, 31*(5), 747-761.
26. Ochi, Y., Liew-Littorin, M., Nannya, Y., Bengtzen, S., & Piauger, B., et al. (2026). Chromatin landscape and epigenetic heterogeneity of acute myeloid leukaemia. *Nature,* 1-11.
27. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., & Hoffmann, T. C., et al. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *bmj, 372.*
28. Campbell, M., McKenzie, J. E., Sowden, A., Katikireddi, S. V., & Brennan, S. E., et al. (2020). Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. *bmj, 368.*
29. Collins, G. S., Moons, K. G., Dhiman, P., Riley, R. D., & Beam, A. L., et al. (2024). TRIPOD+ AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *bmj, 385.*
30. Wolff, R. F., Moons, K. G., Riley, R. D., Whiting, P. F., & Westwood, M., et al. (2019). PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Annals of internal medicine, 170*(1), 51-58.
31. Moons, K. G., Wolff, R. F., Riley, R. D., Whiting, P. F., & Westwood, M., et al. (2019). PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. *Annals of internal medicine, 170*(1), W1-W33.
32. Moons, K. G., Damen, J. A., Kaul, T., Hooft, L., & Navarro, C. A., et al. (2025). PROBAST+ AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. *bmj, 388.*
33. Steinicke, T. L., Benfatto, S., Guerra, M. D. L. R. C., Monteleone, A. B., & Shankar, S., et al. (2024). Rapid epigenomic classification of acute leukemia. *Blood, 144*, 273.
34. Rendeiro, A. F., Krausgruber, T., Fortelny, N., Zhao, F., & Penz, T., et al. (2020). Chromatin mapping and single-cell immune profiling define the temporal dynamics of ibrutinib response in CLL. *Nature Communications, 11*(1), 577.
35. Ghantous, A., Nusslé, S. G., Nassar, F. J., Spitz, N., & Novoloaca, A., et al. (2024). Epigenome-wide analysis across the development span of pediatric acute lymphoblastic leukemia: backtracking to birth. *Molecular Cancer, 23*(1), 238.
36. Bancos, A., Iancu, S. D., Moisoiu, V., Ghiaur, A., & Tigu, A. B., et al. (2025). SERS-Based Assessment of DNA Methylation for the Evaluation of Measurable Residual Disease in Acute Promyelocytic Leukaemia. *Journal of Cellular and Molecular Medicine, 29*(5), e70244.
37. Cancer Genome Atlas Research Network. (2013). Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *New England Journal of Medicine, 368*(22), 2059-2074.
38. Rendeiro, A. F., Schmidl, C., Strefford, J. C., Walewska, R., & Davis, Z., et al. (2016). Chromatin accessibility maps of chronic lymphocytic leukaemia identify subtype-specific epigenome signatures and transcription regulatory networks. *Nature communications, 7*(1), 11938.
39. Li, J., & Wang, S. (2025). Integrative analysis of epigenetic subtypes in acute myeloid leukemia: a multi-center study combining machine learning for prognostic and therapeutic insights. *Plos one, 20*(5), e0324380.
40. Riley, R. D., Ensor, J., Snell, K. I., Harrell, F. E., & Martin, G. P., et al. (2020). Calculating the sample size required for developing a clinical prediction model. *Bmj, 368.*
41. Van Calster, B., McLernon, D. J., Van Smeden, M., & Wynants, L., et al. Steyerberg, E. W. (2019). Calibration: the Achilles heel of predictive analytics. *BMC Med, 17:*230.
42. Kapoor, S., & Narayanan, A. (2023). *Leakage and the reproducibility crisis in machine-learning-based science.* Patterns, 4(9).
43. Jiang, H., Ou, Z., He, Y., Yu, M., & Wu, S., et al. (2020). DNA methylation markers in the diagnosis and prognosis of common leukemias. *Signal transduction and targeted therapy,*

5(1), 3.
44. Imaizumi, T., Meyer, J., Wakamatsu, M., Kitazawa, H., & Murakami, N., et al. (2022). Clinical parameter-based prediction of DNA methylation classification generates a prediction model of prognosis in patients with juvenile myelomonocytic leukemia. *Scientific Reports, 12*(1), 14753.
45. Dong, Y., Liao, H., Huang, F., Bao, Y., & Guo, W., et al. (2025). Machine learning approaches reveal methylation signatures associated with pediatric acute myeloid leukemia recurrence. *Scientific Reports, 15*(1), 15815.
46. Milani, L., Lundmark, A., Kiialainen, A., Nordlund, J., & Flaegstad, T., et al. (2010). DNA methylation for subtype classification and prediction of treatment outcome in patients with childhood acute lymphoblastic leukemia. *Blood, The Journal of the American Society of Hematology, 115*(6), 1214-1225.
47. Zhan, Q., Ye, Y., Lam, T. W., Yiu, S. M., & Wang, Y., et al. (2015). Improving multiple sequence alignment by using better guide trees. *BMC bioinformatics, 16*(Suppl 5), S4.
48. Krali, O., Palle, J., Bäcklin, C. L., Abrahamsson, J., & Norén-Nyström, U., et al. (2021). DNA methylation signatures predict cytogenetic subtype and outcome in pediatric acute myeloid leukemia (AML). *Genes, 12*(6), 895.
49. Chierici, M., Bussola, N., Marcolini, A., Francescatto, M., & Zandonà, A., et al. (2020). Integrative network fusion: a multi-omics approach in molecular profiling. *Frontiers in oncology, 10,* 547844.
50. Schlosser, P., Knaus, J., Schmutz, M., Döhner, K., & Plass, C., et al. (2020). Netboost: boosting-supported network analysis improves high-dimensional omics prediction in acute myeloid leukemia and Huntington's disease. *IEEE/ACM transactions on computational biology and bioinformatics, 18*(6), 2635-2648.
51. Wang, T., Cui, S., Lyu, C., Wang, Z., & Li, Z., et al. (2024). Molecular precision medicine: Multi-omics-based stratification model for acute myeloid leukemia. *Heliyon, 10*(17).
52. Song, Y., Wang, Z., Zhang, G., Hou, J., & Liu, K., et al. (2025). Classification of acute myeloid leukemia based on multi-omics and prognosis prediction value. *Molecular Oncology, 19*(6), 1836-1854.
53. Božić, T., Kuo, C. C., Hapala, J., Franzen, J., & Eipel, M., Platzbecker, U., ... & Wagner, W. (2021). Investigation of measurable residual disease in acute myeloid leukemia by DNA methylation patterns. *Leukemia, 36*(1), 80.
54. Chai, Y., Bono, K. A. M., Tang, N. X. N., Ooi, M. G., & Jen, W. Y., et al. (2026). Methylome profiling reveals context-dependent chemo-resistance mechanisms and enhances risk stratification in AML. *Cancer Cell International, 26*(1), 229.
55. Bhoi, S., Ljungström, V., Baliakas, P., Mattsson, M., & Smedby, K. E., et al. (2016). Prognostic impact of epigenetic classification in chronic lymphocytic leukemia: the case of subset# 2. *Epigenetics, 11*(6), 449-455.
56. Grimm, C., Herling, C. D., Komnidi, A., Hussong, M., & Kreuzer, K. A., et al. (2022). Evaluation of a prognostic epigenetic classification system in chronic lymphocytic leukemia patients. *Biomarker Insights, 17,* 11772719211067972.
57. Ng, S. W., Mitchell, A., Kennedy, J. A., Chen, W. C., & McLeod, J., et al. (2016). A 17-gene stemness score for rapid determination of risk in acute leukaemia. *Nature, 540*(7633), 433-437.
58. Leek, J. T., Scharpf, R. B., Bravo, H. C., Simcha, D., & Langmead, B., et al. (2010). Tackling the widespread and critical impact of batch effects in high-throughput data. *Nature Reviews Genetics, 11*(10), 733-739.
59. Wang, J., Bhakta, N., Ayer Miller, V., Revsine, M., & Litzow, M. R., et al. (2022). Acute leukemia classification using transcriptional profiles from low-cost nanopore mRNA sequencing. *JCO Precision Oncology, 6*, e2100326.
60. Duran-Ferrer, M., Clot, G., Nadeu, F., Beekman, R., & Baumann, T., et al. (2020). The proliferative history shapes the DNA methylome of B-cell tumors and predicts clinical outcome. *Nature Cancer, 1*(11), 1066-1081.
61. Saksena, A., Siewert, C., Turakulov, R., Singh, O., & Abdullaev, Z., et al. (2026). DNA methylation-based classification of hematolymphoid neoplasms. *Blood advances, 10*(13), 4842-4859.
62. Gabbutt, C., Duran-Ferrer, M., Grant, H. E., Mallo, D., & Nadeu, F., et al. (2025). Fluctuating DNA methylation tracks cancer evolution at clinical scale. *Nature, 645*(8081), 764-773.

| PMID | First author | Year | Title (abbreviated) | Query | Decision |
|---|---|---|---|---|---|
| 42640858 | Abdelbaky SB | 2026 | Epigenetic markers expand genetic risk in AML | Q1 | Not retrieved |
| 30723113 | Arribas AJ | 2019 | Promoter methylation of hairy cell leukemia | Q3 | Excluded (full text): no AI/ML model |
| 42539087 | Ashford AJ | 2026 | Mixture-of-experts VAE (preprint) | Q1 | Excluded (title/abstract): preprint |
| 40042091 | Bancos A | 2025 | SERS DNA methylation MRD in APL | Q1 | Excluded (full text): no epigenomic data |
| 41894249 | Bendig S | 2026 | IGH::FENDRR B-ALL subtype | Q1 | Excluded (full text): no epigenomic data |
| 34131280 | Bozic T | 2022 | MRD in AML by DNA methylation patterns | Q2 | Included |
| 41952166 | Chai Y | 2026 | Methylome profiling and chemoresistance in AML | Q1 | Included |
| 32714870 | Chierici M | 2020 | Integrative Network Fusion | Q1, Q3 | Included |
| 38531950 | de Camargo Magalhaes ES | 2024 | Proteomics for AML therapy | Q1 | Excluded (title/abstract): no epigenomic data |
| 34105306 | de la Calle-Fabregat C | 2021 | Methylation in undifferentiated arthritis | Q1 | Excluded (title/abstract): not leukemia |
| 39450427 | Della Porta MG | 2025 | Future directions in MDS/AML classification (review) | Q1 | Excluded (title/abstract): review |
| 40328883 | Dong Y | 2025 | ML reveals methylation signatures of pediatric AML recurrence | Q1, Q3 | Included |
| 42064384 | Drewes C | 2026 | CLL with IGH::BCL3 | Q3 | Excluded (full text): no AI/ML model |
| 40065077 | Glaser S | 2025 | Germinal-centre B-cell lymphoma methylation modules | Q1, Q3 | Excluded (title/abstract): not leukemia |
| 40192513 | Glaser S | 2025 | Subtyping Burkitt lymphoma by DNA methylation | Q3 | Excluded (title/abstract): not leukemia |
| 39908143 | Gokbuget D | 2025 | KMT2C/KMT2D H3K4me1 and replication timing | Q1 | Excluded (title/abstract): no relevant outcome |
| 35095271 | Grimm C | 2022 | Prognostic epigenetic classification in CLL | Q1 | Included |
| 36042365 | Imaizumi T | 2022 | Clinical-parameter prediction of JMML methylation class | Q1 | Included |
| 27435615 | Jalali A | 2016 | Per-case weighted ensemble | Q1 | Excluded (full text): no epigenomic data |
| 32296024 | Jiang H | 2020 | DNA methylation markers in common leukemias | Q1 | Included |
| 41821663 | Josephine A | 2026 | IDH inhibition in cancer (review) | Q1, Q3 | Excluded (title/abstract): review |
| 35069208 | Khayat MT | 2022 | Phenylacetamide variants | Q3 | Excluded (title/abstract): not leukemia |
| 38731939 | Kim DY | 2024 | Azacitidine resistance biomarkers in MDS | Q1 | Excluded (title/abstract): not leukemia |
| 18229696 | Kim S | 2008 | Predicting methylation susceptibility | Q3 | Excluded (title/abstract): no relevant outcome |
| 34580726 | Kitazawa H | 2021 | Methylation test for JMML risk stratification | Q1 | **Included** |

| | | | | | |
|---|---|---|---|---|---|
| 33479225 | Koelsche C | 2021 | Sarcoma classification by DNA methylation profiling | Q1 | Excluded (title/abstract): not leukemia |
| 36686949 | Koowattanasuchat S | 2023 | AuNPs and ML smartphone for cancer DNA | Q1 | Excluded (full text): no epigenomic data |
| 38875909 | Kosvyra A | 2024 | ML and integrative multi-omics network analysis for survival prediction in AML | Q1, Q3 | Not retrieved |
| 38066241 | Krali O | 2023 | Multimodal classification of pediatric ALL subtypes | Q1 | **Included** |
| 39329449 | Latiri M | 2024 | DNA methylation: T-LBL vs thymoma | Q1 | Excluded (title/abstract): not leukemia |
| 38159337 | Le Rhun E | 2024 | Zotiraciclib in glioblastoma | Q3 | Excluded (title/abstract): not leukemia |
| 39794314 | Lee D | 2025 | Methylation disorder in DNMT3A AML | Q3 | Excluded (full text): no AI/ML model |
| 40435135 | Li J | 2025 | Epigenetic subtypes in AML with ML | Q1 | **Included** |
| 39223659 | Li Y | 2024 | Multi-omics survival benchmark | Q1 | Excluded (full text): not leukemia |
| 41118575 | Lin YY | 2025 | Pipeline Olympics benchmarking | Q1, Q3 | Excluded (title/abstract): no relevant outcome |
| 41074259 | Loscocco GG | 2025 | MPN mutation profiling (review) | Q3 | Excluded (title/abstract): review |
| 42801378 | Mallick S | 2026 | Runx1-Gata2 axis (review) | Q3 | Excluded (title/abstract): review |
| 31559392 | Mangaonkar AA | 2019 | Advances in CMML (review) | Q1 | Excluded (title/abstract): review |
| 40730747 | Marchi F | 2025 | Epigenomic diagnosis and prognosis of AML | Q1 | Included |
| 38549146 | Mosquera Orgueira A | 2024 | Risk prediction in pediatric ALL by DNA methylation | Q1, Q3 | Included |
| 40838101 | Mulet-Lazaro R | 2025 | Transcriptional profiling of ambiguouslineage leukemia | Q1 | Excluded (title/abstract): no epigenomic data |
| 30806759 | Naushad SM | 2019 | CART for 6-MP leucopenia in ALL | Q1 | Excluded (full text): no epigenomic data |
| 32760119 | Oriol-Tordera B | 2020 | Methylation and HIV control | Q3 | Excluded (title/abstract): not leukemia |
| 32365516 | Pasca S | 2020 | Methylating mutations in AML | Q1 | Excluded (full text): no relevant outcome |
| 42024465 | Saksena A | 2026 | DNA methylation classification of hematolymphoid neoplasms | Q1 | Excluded (full text): no leukemia-specific results |
| 32870144 | Saleh C | 2021 | Rare brain tumours unsupervised ML | Q1 | Excluded (title/abstract): not leukemia |
| 31717838 | Sanchez R | 2019 | Methylation signal vs background variation | Q1 | Excluded (full text): no relevant outcome |
| 32034170 | Sanchez R | 2020 | Integrative network analysis of DMGs in leukemia | Q1 | Excluded (full text): no relevant outcome |
| 32365034 | Schlosser P | 2021 | Netboost | Q1, Q3 | Included |
| 42369521 | Shah AJ | 2026 | MOSAIC classifier (preprint) | Q2 | Excluded (title/abstract): preprint |
| 35743098 | Shait Mohammed MR | 2022 | IDH1 targeting metabolome | Q3 | Excluded (title/abstract): no AI/ML model |

| | | | | | |
|---|---|---|---|---|---|
| 40594942 | Shao D | 2025 | RareNet deep learning for rare cancers | Q1, Q2, Q3 | Included |
| 36033519 | Song J | 2022 | Methylation signatures associated with CAR T cells | Q1, Q3 | Excluded (full text): no relevant outcome |
| 40983754 | Steinicke TL | 2025 | Rapid epigenomic classification of acute leukemia | Q2 | **Included** |
| 41481703 | Telonis AG | 2026 | IDH2/SRSF2 intragenic epigenetic deregulation | Q1 | Excluded (title/abstract): no relevant outcome |
| 39198401 | Truong P | 2024 | TOPORS mediates resistance to hypomethylating agents | Q1 | Excluded (title/abstract): no AI/ML model |
| 39263156 | Wang T | 2020 | Multi-omics stratification model for AML | Q1 | **Included** |
| 26797014 | Wang Y | 2016 | Predicting CpG methylation state with deep networks | Q2 | Excluded (title/abstract): no relevant outcome |
| 25348669 | Wertheim GB | 2015 | xMELP methylation predicts AML outcome | Q3 | **Included** |
| 28735484 | Wertheim GBW | 2017 | Microsphere methylation assay for AML (methods chapter) | Q3 | Excluded (title/abstract): review |
| 39564730 | Witzig TE | 2025 | Plasma methylated DNA in lymphoma | Q3 | Excluded (title/abstract): not leukemia |
| 34562613 | Xia D | 2021 | Methylation classification of small B-cell lymphomas | Q1 | Excluded (title/abstract): not leukemia |
| 30529546 | Xing J | 2019 | BRD4 inhibitor design | Q1 | Excluded (title/abstract): not leukemia |
| 37391640 | Yang J | 2023 | m6A modulator signature in AML | Q3 | Excluded (title/abstract): no epigenomic data |
| 40361241 | Zehtabcheh S | 2025 | KMT2A rearrangements in AML (review) | Q1 | Excluded (title/abstract): review |
| 38650628 | Zhang B | 2024 | Hub genes and drug sensitivity in AML based on ML | Q1 | Excluded (full text): no epigenomic data |
| 30601803 | Zhang SY | 2019 | m6A functions by deep learning | Q2 | Excluded (title/abstract): no epigenomic data |
| 35336734 | Zhou K | 2022 | Integration of multimodal data for disease subtypes | Q2 | **Included** |

**Supplementary Table S1: Screening log of Database Records (n = 68 unique records)**

| Study | D1 Particip. | D2 Predict. | D3 Outcome | D4 Analysis | Overall | Applic. concern | Main reasons |
|---|---|---|---|---|---|---|---|
| Milani 2010 [46] | Low | Low | Low | High | **High** | Low | Trial-based cohort; metrics not reported at classifier level; no external test. |
| Taskesen 2015 [47] | Low | Low | Low | High | **High** | Low | Trial cohort; double-loop cross-validation only; no external test or calibration. |
| Nordlund 2015 [14] | Low | Low | Low | Unclear | **Unclear** | Low | Population-based trial cohort; blinded heldout set from same source; no calibration. |
| Jiang 2020 [43] | High | Low | Low | High | **High** | Unclear | Cases and healthy controls from different sources (spectrum bias); near-perfect AUCs suggest optimism. |
| Krali 2021 [48] | Low | Low | Low | High | **High** | Low | Small subgroups (n = 18–30) for outcome signatures; 33-sample test set. |
| Krali 2023 [15] | Low | Unclear | Low | Unclear | **Unclear** | Low | No batch correction across arrays; large cohort with external sets. |
| Steinicke 2025 [7] | Unclear | Low | Unclear | Unclear | **Unclear** | Low | Pooled public reference set; some classes defined from the same data; small prospective cohort. |
| Marchi 2025 [21] | Low | Low | Low | Unclear | **Unclear** | Low | Trial cohorts with held-out and external tests; mean imputation; no calibration. |
| Shao 2025 [23] | Unclear | Unclear | Low | High | **High** | Unclear | Public data, no batch correction; only 11 AML test samples. |
| Wertheim 2015 [18] | Low | Low | Low | High | **High** | Low | Trial training cohort; 70-patient validation; no hazard ratio or calibration. |
| Chierici 2020 [49] | Unclear | Low | Low | High | **High** | Unclear | Public TCGA data; low events per variable; test MCC below CV MCC. |
| Schlosser 2021 [50] | Unclear | Low | Low | High | **High** | Low | 55-patient external cohort; no calibration. |
| Giacopelli 2021 [25] | Unclear | Low | Low | Unclear | **Unclear** | Low | Unsupervised subtypes validated in two cohorts; no individual-level classifier. |
| Zhou 2022 [24] | Unclear | Low | High | High | **High** | Unclear | Labels derived from the same omics data with manual adjustment. |
| Wang 2024 [51] | Unclear | Low | Low | High | **High** | Low | Methylation used only in a 126-patient discovery set; external validation without methylation. |
| Li 2025 [39] | Unclear | Low | Low | High | **High** | Low | Methylation subtypes derived from 83 patients; external validation transcriptomeonly. |
| Song 2025 [52] | Unclear | Unclear | Low | High | **High** | Low | Methylation available for 30 local patients; validation largely transcriptomic. |
| Ochi 2026 [26] | Low | Low | Low | Unclear | **Unclear** | Low | Large cohort-based ATAC-seq; external validation via RNA proxy; no calibration. |
| Božic 2022 [53]′ | Unclear | Low | Low | High | **High** | Low | Small MRD cohort; high false-positive rate. |
| Study | D1 Particip. | D2 Predict. | D3 Outcome | D4 Analysis | **Overall** | Applic. concern | Main reasons Mosquera Orgueira |
| 2024 [20] | Low | Low | Low | Unclear | **Unclear** | Low | External cohorts; marked performance drop in one external cohort; events per variable not reported. |
| Dong 2025 [45] | Unclear | Low | Low | High | **High** | High | Cross-validation only; diagnosis and relapse samples may share patients (leakage); predicts sample state, not future relapse. |
| Chai 2026 [54] | Unclear | Low | Low | High | **High** | Low | 25-patient test set; 34/146 patients without response data excluded. |
| Schönung 2021 [22] | Low | Low | High | Unclear | **High** | Low | Classification labels from clustering of the same data; wide CIs. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Kitazawa 2021 [17] | Low | Low | High | High | **High** | Low | Labels from same methylation data; 38patient validation; reported training error of zero. |
| Imaizumi 2022 [44] | Low | Low | Low | High | **High** | High | Clinical predictors only; validation cohort lacks methylation data. |
| Queirós 2015 [16] | Low | Low | Low | Unclear | **Unclear** | Low | Two cohorts; hazard ratios not available in accessible text. |
| Bhoi 2016 [55] | Unclear | Low | Low | High | **High** | Low | Small external cohort (n = 135; 12 in subset #2). |
| Rendeiro 2016 [38] | Unclear | Low | Low | High | **High** | Low | Leave-one-out CV with repeated samples per patient; 112,298 features for 88 samples. |
| Giacopelli 2019 [19] | Low | Low | Low | Low | **Low** | Low | Large multi-cohort external testing; calibrated probabilities with ambiguity threshold. |
| Grimm 2022 [56] | Unclear | Low | Low | High | **High** | Low | Single-centre external validation with few events. |

Overall risk of bias is high if any domain is high, low if all domains are low, and unclear otherwise. AI-specific signalling considerations from PROBAST+AI were applied within the four domains.

**Supplementary Table S2: PROBAST Judgements and Rationale**

| **Report** | **Source** | **Decision** |
|---|---|---|
| Milani 2010 | supplementary query | **Included** |
| Nordlund 2015 | supplementary query | **Included** |
| Taskesen 2015 | supplementary query | **Included** |
| Krali 2021 | supplementary query | **Included** |
| Giacopelli 2021 | supplementary query | **Included** |
| Song Y 2025 | supplementary query | **Included** |
| Schönung 2021 | supplementary query | **Included** |
| Ochi 2026 | supplementary query | **Included** |
| Kimura 2020 | supplementary query | Excluded: no AI/ML model |
| Haider 2020 | supplementary query | Excluded: no AI/ML model |
| Schönung 2025 | supplementary query | Excluded: no relevant outcome |
| Ghanjati 2025 | supplementary query | Excluded: no AI/ML model |
| Gabbutt 2025 | supplementary query | Excluded: no AI/ML model |
| Charalampopoulou 2024 | supplementary query | Excluded: no AI/ML model |
| Rendeiro 2016 | citation searching | **Included** |
| Queirós 2015 | web search | **Included** |
| Giacopelli 2019 | citation searching | **Included** |
| Bhoi 2016 | web search | **Included** |
| Rendeiro 2020 | citation searching | Excluded: no AI/ML model |
| Ghantous 2024 | citation searching | Excluded: no AI/ML model |
| Duran-Ferrer 2020 | web search | Excluded: no leukemia-specific results |
| Oakes 2016 | citation searching | Excluded: no AI/ML model |
| Wierzbinska 2020 | citation searching | Excluded: no relevant outcome |
| Charalampopoulou 2026 | web search | Excluded: no AI/ML model |
| Li XP 2025 | web search | Excluded: no epigenomic data |
| Zhang 2022 Front Oncol | web search | Excluded: no epigenomic data |
| Lyu 2025 | web search | Excluded: not leukemia |

| | | |
|---|---|---|
| Lauzon-Young 2025 | citation searching | Excluded: review |
| Dalyan 2025 | citation searching | Excluded: review |
| Shiwlani 2025 | citation searching | Excluded: review |
| Hamamoto 2020 | citation searching | Excluded: review |
| Drozak 2022˙ | citation searching | Excluded: review |

**Supplementary Table S3: Reports Identified through other Methods (n = 32)**

| Section / topic | Item | Location in manuscript |
|---|---|---|
| Title | 1 | Title page |
| Abstract | 2 | Structured abstract |
| Rationale | 3 | Sections 1.1–1.4 |
| Objectives | 4 | Section 1.4 |
| Eligibility criteria | 5 | Section 2.2 |
| Information sources | 6 | Section 2.3; Table 2 |
| Search strategy | 7 | Section 2.3; Table 2 |
| Selection process | 8 | Section 2.4; Supplementary Tables S1, S3 |
| Data collection process | 9 | Sections 2.4–2.5 |
| Data items | 10a–b | Section 2.5 |
| Study risk-of-bias assessment | 11 | Section 2.6 |
| Effect measures | 12 | Section 2.5 (metrics as reported) |
| Synthesis methods | 13a–f | Section 2.7 |
| Reporting bias assessment | 14 | Not performed (Section 4.7) |
| Certainty assessment | 15 | Not performed (Section 4.7) |
| Study selection | 16a–b | Section 3.1; Figure 2; Supplementary Tables S1, S3 |
| Study characteristics | 17 | Section 3.2; Table 3; Figures 3–4 |
| Risk of bias in studies | 18 | Section 3.3; Figure 5; Supplementary Table S2 |
| Results of individual studies | 19 | Table 4; Sections 3.4–3.8 |
| Results of syntheses | 20a–d | Sections 3.4–3.8; Figures 6–8 |
| Reporting biases | 21 | Not assessed (Section 4.7) |
| Certainty of evidence | 22 | Not assessed (Section 4.7) |
| Discussion | 23a–d | Sections 4.1–4.8 |
| Registration and protocol | 24a–c | Section 2.1 (not registered) |
| Support | 25 | Declarations — Funding |
| Competing interests | 26 | Declarations — Competing interests |
| Availability of data, code and other materials | 27 | Declarations — Data and code availability |

**Supplementary Table S4: PRISMA 2020 Checklist**

***Copyright:*** *©2026 Jacopo Martelli. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.*