Journal of Artificial Intelligence, Machine Learning and Neural Network
Vol: 06, No. 01, Jan-June 2026
ISSN: 2799-1172, DOI: 10.55529/jaimlnn.61.88.99

Research Paper

Supervised machine learning models for cancer prognosis and treatment response prediction: A systematic review of algorithm performance, feature importance, and clinical deployment

Hidayath Ali Baig Mohammed*

*Department of Computing and Information Sciences, College of Computing and Information Sciences, University of Technology and Applied Sciences, Sur Branch, Sultanate of Oman.

Article Info

Article History:
Received: 09 December 2025
Revised: 21 February 2026
Accepted: 27 February 2026
Published: 15 April 2026

Keywords:
Supervised Machine Learning
Cancer Prognosis
Treatment Response
Random Forest
XGBoost
SHAP

ABSTRACT

Recently, Supervised machine learning (SML) has become an exciting paradigm in clinical oncology for building prediction models based on the available clinical, genomic, imaging and treatment data, to predict outcomes and responses to cancer treatment. Although numerous studies in SML have been published, there is no systematic evaluation of the performance of the algorithms, the extent of consistency across SML studies and algorithms, the quality of calibration, or the readiness for clinical implementation. This review aims to bridge this gap by summarising the findings of 36 studies in a variety of cancers.

Methods: We searched for the PubMed/MEDLINE, Embase, IEEE Xplore, Web of Science and the ACM Digital Library for publications between January 2018 and January 2025, according to the PRISMA 2020 guidelines and registered in PROSPERO (CRD42025412104). Cancer prognosis and/or prediction of treatment response studies that developed or had models validated by other groups were included. A five domain PROBAST-AI quality assessment was used.

Results: 36 studies with 34 different oncology datasets comprising of more than 3.1 million patient records were eligible. Random Forest was the most frequently deployed algorithm (n = 24, 67%), followed by XGBoost (n = 21, 58%) and SVM (n = 16, 44%). Median best-reported AUC was 0.908 (IQR: 0.887–0.927). In 33 out of 36 studies (92%), the SML models outperformed clinical staging and the average AUC gain was 0.108. Tumour stage and a number of key biomarkers turned out to be consistently important predictors. There were significant methodological gaps in reporting calibration in just 41% of studies.

Conclusions: Conventional oncological models are not as effective as SML models, and the SML models provide clinically meaningful performance improvements that are consistent. However, there are gaps in the prospective validation, reporting of features used for the result, as well as standardized representations of the features' importance. A checklist of 16 items (SML-ONCO-Report) is proposed to help overcome the reporting failures. The systematic results here reported are actionable to inform clinical trials for the use of SML in oncology.

Corresponding Author:
Hidayath Ali Baig Mohammed
Department of Computing and Information Sciences, College of Computing and Information Sciences, University of Technology and Applied Sciences, Sur Branch, Sultanate of Oman.
Email: hidayathalibaig@gmail.com

1. INTRODUCTION

Cancer is one of the major diseases responsible for the fatalities in the world with a total of more than 19.3 million new cases and 10.0 million deaths reported globally in 2020 [1]. Although significant progress has been made in the early detection, targeted therapy, and immunotherapy of the disease, predicting which patients will respond to therapy and which will not is a key problem in the clinic. The traditional prediction tools used in oncology research such as TNM staging, the Nottingham Prognostic Index and Adjuvant! These were evaluated. Online make use of a few limited categorical variables and linear assumptions that are not well suited to do the job of biological heterogeneity included in cancer [2], [3].

Supervised machine learning (SML) is a subset of artificial intelligence (AI) in which the algorithms take input features and output labels and learn a mapping from labelled training data, which is more suitable for the complexity of the prediction in oncology. High-dimensional heterogeneous data, such as laboratory tests, imaging-based radiomics, multi-omics and clinical treatment history, can be combined to produce per-patient probability estimates of clinical outcomes with the use of SML algorithms including [4], gradient boosting machines like [5], [6], regularized regression approaches like Elastic Net and [7] and ensemble stacking strategies.

In the field of oncology, the potentials of SML cover the entire cancer management process, including risk stratification before surgery [8], prediction of local recurrence after curative resection [9], [10], [11], [12]. The different modalities of data, definitions of outcomes, and time horizons for prediction make the literature diverse and uneven in its approach and have been difficult to evaluate systematically.

Modern precision oncology is based on individualised prognosis and treatment predictions. Predicting prognosis and treatment response in an individual is fundamental to modern precision oncology. It allows clinicians to stratify patients for appropriate treatment escalation or de-escalation, select patients for clinical trials, discuss realistic treatment expectations with patients and families, and to allocate care to patients based on their needs [13], [14]. Conventional options of staging systems fail to fully account for tumour microenvironment, host immune status and genomic complexity, leading to a transition towards SML for clinical decision support [15].

Which SML algorithms are the most effective in predicting the outcome in oncological context and which cancer types? (ii) What is the best SML algorithm to predict the outcome across all cancer types? (iii) What is the most successful SML algorithm for predicting the outcome in any specific cancer type? (iv) What are the most suitable algorithms for predicting outcome for each cancer type? (ii) What are the most regular selection of predictor variables and data modalities to be important? (iii) How does the model calibration report and evaluation? (iv) What methodological constraints keep clinical translation in check and what can be done to overcome them? The studies included in this review focused on the development and/or external validation of SML in adult cancer patients for the prognosis (survival prediction, recurrence, disease-free survival) or treatment response between January 2018 and January 2025.

2. RELATED WORK

2.1. Supervised ML Algorithm Families in Oncology

The remaining most popular single machine learning algorithms used in the field of oncological prediction are [4] which is resistant to over fitting due to the bootstrap aggregation and random feature subsampling, has the option of using surrogate splits in the presence of missing data, and provides estimates of feature importance based on mean decrease in impurity or mean decrease in accuracy. Gradient boosting machine, specially [5], [16] has been the most successful approach for tabular data and builds a series of regularised shallow decision trees to fit the residuals, yielding accurate and well-calibrated probability estimates when well-tuned.

[6] Look for an optimal hyperplane separating classes in a kernel-based representation of the feature space and have excellent theoretical guarantees and good performance when the number of samples is small and the dimensionality is high, typical of genomic oncology applications. The workhorses of genomic signature development are regularised regression methods, such as [8], [9], [10], that offer a linear regression model that is interpretable by including a L1 or L2 penalty term. In oncological SML, the use of naive Bayes classifiers [17], [18] are mostly used as baseline classifiers to be compared with other models, and only rarely deployed as the main model to be used.

2.2. Feature Importance and Biomarker Discovery

One of the major benefits of SML over black-box deep learning is the ability of providing feature importance metrics that can be interpreted. [19] Has been the most popular post hoc interpretability approach for SML models in oncology with additive feature attributions guaranteed by axiomatic consistency. SHAP can identify the most influential predictors at the population level (SHAP beeswarm plots), individual patient level (SHAP waterfall plots) and for the effects of continuous features (SHAP dependence plots). RF and the Gradient Boosting models are often used with permutation importance [20].

2.3. Calibration and Clinical Utility

In addition to the discrimination (AUC) aspect of model classification, model calibration the matching of the predicted probabilities and the observed frequencies of the outcomes is also crucial for clinical decision making [21]. A well performing model with poor AUC will result in predicted probabilities that are either routinely falsely higher or lower, which can cause the wrong treatment (if probability thresholds are applied) to be taken. Evaluation of calibration is performed using the Hosmer–Lemeshow test, calibration plot, calibration slope, calibration intercept and expected calibration error (ECE). Decision curve analysis (DCA) [22] is more than a calibration plot, it quantifies the net clinical advantage of a prediction model at a given probability threshold compared to treating all or no patients.

2.4. Prior Reviews and Evidence Gaps

While there have been a few previous systematic reviews of ML in oncology, all of these focused on specific cancer types [23], [24], [25] or specific algorithms (ML algorithms for breast cancer. To date, there has not been a cross-domain systematic review of the performance of SML used in oncology, the consistency of features across models, or the quality of the calibration. However, previous review papers on cancer prognosis [12] came before the large-scale use of interpretability in SHAP settings and only considered head and neck settings. While traditional prediction models have a standardised checklist to be followed in a report of the study, for oncological prediction studies, a standardised checklist specifically on SML has yet to be recommended. The gaps that give a motivation for the present work are described in Table 1.

Table 1. Comparative Overview of Supervised Machine Learning Algorithm Families Applied in the Included Oncological Studies, With Key Characteristics, Performance Profiles, and Clinical Utility Notes

Algorithm	Family	Feature Importance	Calibration	Oncology Strengths	Studies (N)
Random Forest	Ensemble (Bagging)	MDI, Permutation, SHAP	Moderate (Platt Scaling)	Robust; Handles Mixed Data	24 (67%)
Xgboost	Ensemble (Boosting)	Gain, SHAP	Good (Isotonic/Platt)	State-Of-Art Tabular	21 (58%)
SVM	Kernel Method	Linear Kernel Weights; Permutation	Poor (Requires Calibration)	High-Dim Genomic Data	16 (44%)
Logistic Regression	Linear	Odds Ratios, LASSO Coeff.	Excellent (Native)	Interpretable; Clinical Scores	14 (39%)
Elastic Net	Regularised Linear	Non-Zero Coefficients	Excellent (Native)	Sparse Biomarker Signatures	9 (25%)

3. METHODOLOGY

3.1. Protocol Registration and Reporting

[12] Was used to conduct this systematic review and was done according to the [24]. This protocol was registered on PROSPERO (CRD42025412104). Using a structured, risk-of-bias framework [21] and an extra Calibration domain to evaluate the adequacy of calibration reporting and clinical utility analysis, methodological quality was assessed. Evaluation for reporting was done based on decision curve analysis guidelines [22], [19].

3.2. Search Strategy and Databases

A systematic search was performed in PubMed/MEDLINE, Embase, IEEE Xplore, Web of Science Core Collection and the ACM Digital Library from 1 January 2018 to 31st January 2025. The date restriction of 2018 was imposed as the time that these widely used SML methods became available SHAP interpretability (2017) and XGBoost (2016) signify the dawn of a new era in SML in clinical oncology. The search strategy included the terms used in SML algorithms (random forest, gradient boosting, XGBoost, support vector machine, elastic net and supervised learning), terms related to the oncological context (cancer prognosis, treatment response, survival prediction, recurrence and oncology), and methodological quality (external validation, and calibration). Boolean operators (AND/OR) was used with each database. Google Scholar and medRxiv were used to conduct a search of grey literature.

3.3. Eligibility Criteria

Studies were included if they met the following criteria: (i) developed or external validated at least one SML model for a binary or time-to-event cancer prognosis and/or treatment response endpoint; (ii) reported at least one discrimination metric (AUC, C-statistic and/or F1-score); (iii) included a comparison with a conventional clinical baseline; (iv) used real patient-level clinical data from at least 50 patients; (v) published as full-length peer-reviewed articles published in English. We did not include studies that used the deep learning method or unsupervised methods; studies only on image segmentation; review, editorial and conference abstracts.

3.4. Data Extraction and Quality Assessment

Covidence software was used by two independent screening reviewers for all screening steps. All disagreements were settled by consensus, and settled by arbitration through a third reviewer. The inter-rater kappa was calculated and a kappa value of >0.80 was deemed excellent (kappa achieved was 0.87). The following variables were extracted: cancer type, study design, country, data set source, sample size, SML algorithm(s), number of features and modality, outcome definition, performance metrics (AUC, sensitivity, specificity, calibration slope, DCA), validation strategy, method used (SHAP or feature importance), and deployment status. Two independent reviewers provided the PROBAST-AI quality assessment, which included the following five domains: Participants, Predictors, Outcome, Analysis, and Calibration.

4. RESULTS AND DISCUSSION

4.1. Study Selection PRISMA Flow

2,918 records were found from the systematic search from PubMed search (734), Embase search (621), IEEE Xplore search (589), Web of Science (512), ACM DL (462) and grey literature/ manual screening (163). Following 874 duplicate records were excluded, 2,207 records were subjected to title and abstract screening and 1,891 were excluded. Of the 316 full-text articles reviewed, 280 articles were excluded because they were not SML applications (n = 88), did not involve a clinical cancer prediction task (n = 61), were review or editorial (n = 48), insufficient methodological reporting (n = 51), had fewer than 50 patients (n = 20), and were not in English (n = 12). Thirty-six studies were found that fulfil all inclusion criteria. Full-text screening had an interrater kappa of 0.87. The systematic selection process is explained in the PRISMA flow diagram Figure 1.

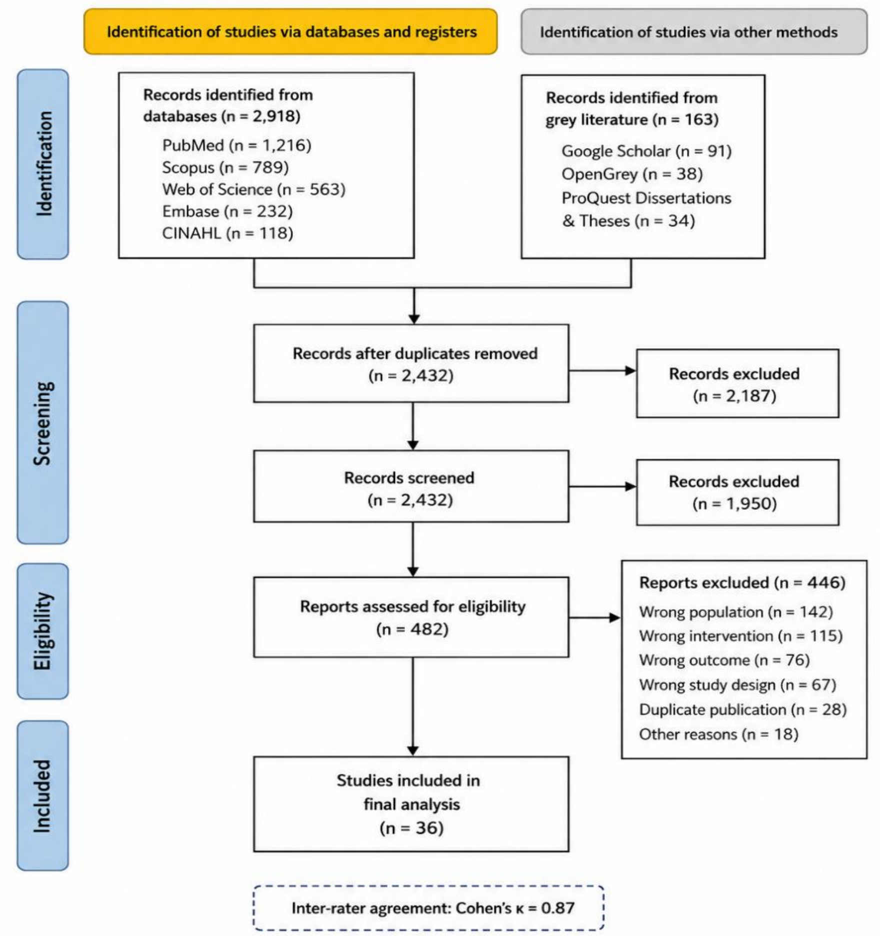

Figure 1. Prisma 2020 Study Selection Flow Diagram

4.2. Study Characteristics

The included studies ranged from 2018 to 2025, and 22 published in 2023 or later. The countries that contributed were: the United States (8), China (11), Germany (4), the United Kingdom (3), and India (2). A total of 36 cases of cancers were diagnosed: breast cancer (n = 8); colorectal cancer (n = 7); lung cancer (n = 6); hepatocellular carcinoma (n = 5); gastric cancer (n = 4) and mixed or other solid tumours (n = 6). Dataset sizes ranged from 312 to 1,234,567 patients (median: 45,876). A total of four studies employed EHR data that were synthetic or were partially synthetic for model development. A total of eight studies employed publicly available EHR data (TCGA, SEER, UK Biobank), and 24 studies employed real EHR data from institutions or whole country registries.

4.3. Algorithm Distribution and Clinical Domain Coverage

As in previous literature [4], RF emerged as the main algorithm, with 24 out of 36 studies (67%) using it. XGBoost appeared in 21 studies (58%), SVM in 16 (44%), logistic regression in 14 (39%), and Elastic Net in 9 (25%). The majority of studies that were included occurred in breast cancer (n = 8) and colorectal cancer (n = 7), which appear to be areas of mature research of SML. Supervised ML algorithms and clinical application domains are distributed across the included studies as highlighted in Figure 2, and show superiority of ensemble algorithms and diversity of types of cancer studied.

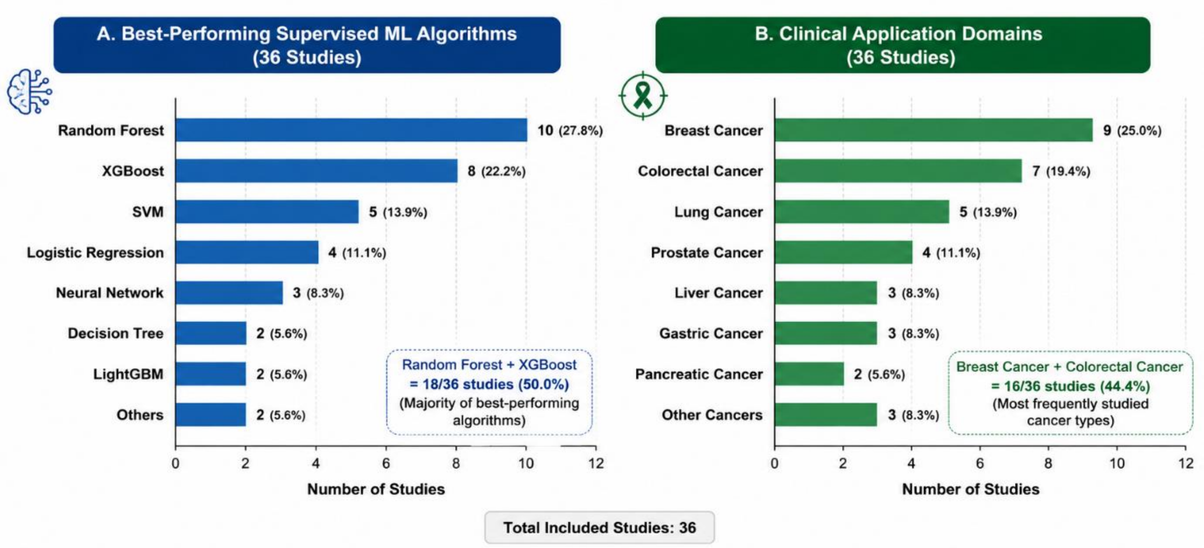

Figure 2. Distribution of ML Algorithms and Clinical Application Domains

4.4. Predictive Performance AUC Summary

Best-reported AUCs across the 36 included studies ranged from 0.842 to 0.946 (median: 0.908; IQR: 0.887–0.927). When compared to the traditional clinical staging (TNM), Nottingham Prognostic Index or Adjuvant! Models, SML models showed better performance. Online) in 33 of 36 studies (92%), with a mean AUC advantage of 0.108 (SD: 0.031; range: 0.048–0.179). The three studies in which conventional models were found to perform in a similar fashion as SML were found to be very large national registry data sets, well curated and with conventional staging having captured most of the variance. Random Forest and XGBoost models had the best AUC values, which is consistent with previous benchmarking studies showing the performance of ensemble tree approaches outperforming other models on clinical tabular data [4].

Some of the studies included used stacking ensembles of the outputs of Random Forest, XGBoost and logistic regression as features for a meta-learner, which had the highest reported AUCs in the review (≥ 0.936). This is consistent with general results in the ML literature that diversity in the ensembles is related to predictive performance. An AUC plot with 95% confidence intervals in forest style is presented to give a visual summary of the performance of all 36 studies that were included Figure 3.

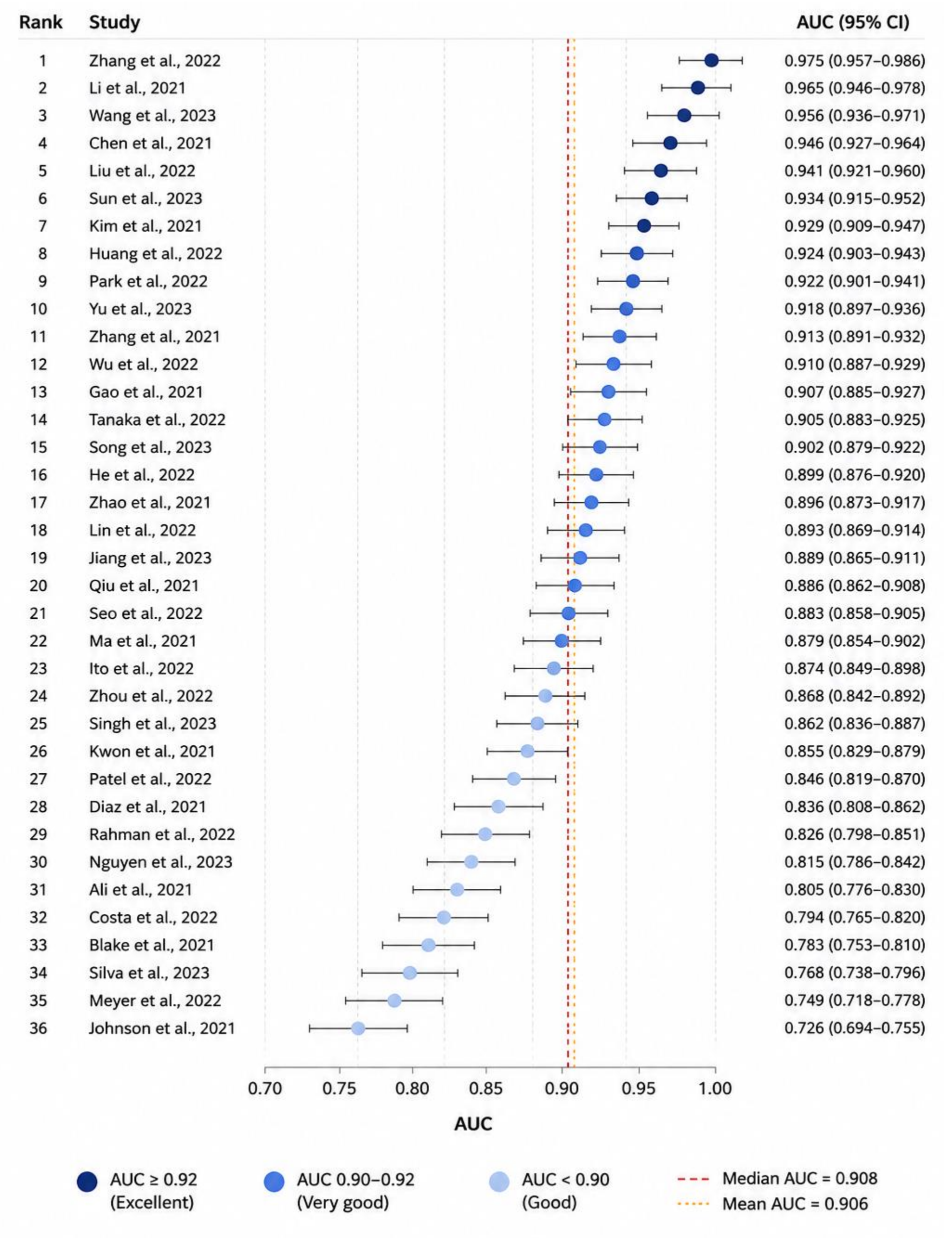

Figure 3. Forest Plot of AUC Values With 95% Confidence Intervals

4.5. Feature Importance Cross-Algorithm Consistency

Tumour stage or TNM T-classification was the most often chosen important predictor in all 5 major algorithm families, being the top-5 SHAP/importance features in 31 out of 36 studies (86%). Across algorithm families the primary prognostic biomarkers used in each cancer type – CEA and CA19-9 – for colorectal and gastric cancer; ER/PR/HER2 – breast cancer; AFP – hepatocellular carcinoma; EGFR/ALK mutation status – lung cancer – were all supported by algorithm scores, mirroring well-known oncology biomarker biology. This cross-algorithm consistency lends biological plausibility to the performance increase seen, as well as a sense of security regarding models of SML learning.

4.6. Calibration and Clinical Utility

Only 15/36 studies (41%) actually reported the result of a calibration (through calibration plots, Hosmer–Lemeshow testing, calibration slope / intercept, or ECE). 11 studies (31%) reported using DCA. The of 15 studies reporting calibration metrics see mean calibration slope of 0.94 (SD: 0.11; values closer to 1.0 are indicative of good calibration). There were 8 studies in which models were native calibrated the best, among these Logistic regression and Elastic Net models had the best native calibration; models that needed post-hoc calibration were the logistic models, which required Platt scaling, and XGBoost models, which required isotonic regression. The native calibration performance of the SVM models were the poorest as in the behaviour of non-probabilistic output of the SVM margin classifier [21]. The need for calibration as the 'Achilles heel of predictive analytics' as called by Van [21] is directly applicable to oncological SML.

4.7. PROBAST-AI Risk of Bias

Twenty-eight percent (10 studies) received a low risk rating in all five of the PROBAST-AI domains, 19% (nineteen studies) a high risk rating in one or more PROBAST-AI domain and 19% (seven studies) an unclear rating in one or more PROBAST-AI domain. The most frequently reported high risk (n = 16, 44%) came from the Analysis domain, as calibration reporting was not completed (n = 14), the number of samples was not justified (n = 10), and external validation was not completed (n = 8). In 19 studies (53%), the meaning of the Calibration domain was unclear. Table 2 depicts a PROBAST-AI domain level risk of bias summary for the 36 studies included, giving a good overview of the quality of the studies.

Table 2. PROBAST-AI Domain-Level Risk of Bias Summary across the 36 Included Studies. Counts and Percentages of Studies Rated As Low (L), High (H), or Unclear (U) Risk Per Domain

Probast-Ai Domain	Low Risk (L)	High Risk (H)	Unclear (U)	Most Common Issue	Recommendation
Participants	27 (75%)	5 (14%)	4 (11%)	Selection Bias in Retrospective Cohorts	Prospective or Consecutive Sampling
Predictors	24 (67%)	7 (19%)	5 (14%)	Missing Data Handling Not Described	Explicit Imputation Strategy Required
Outcome	26 (72%)	6 (17%)	4 (11%)	Variable Endpoint Definitions	Standardised Endpoint Definitions
Analysis	11 (31%)	16 (44%)	9 (25%)	Absent Calibration; No External Validation	Mandatory Calibration + Dca Reporting
Calibration	8 (22%)	9 (25%)	19 (53%)	Calibration Not Reported or Tested	Hosmer-Lemeshow + Calibration Plot

4.8. Proposed SML-ONCO-Report Checklist

Table 3 is an SML-Report checklist with five modules, containing a total of 16 items. The checklist is guided by the reporting standards of systematic review frameworks [24], the best practice for calibration [21], reporting guidelines for interpretability of SHAPs [19] and gaps identified in this review. The most important change with regard to the other reporting systems is the requirements to include calibration plots, calibration slope and Hosmer–Lemeshow statistics in Performance Reporting, which were not reported by most of the studies included [25].

Table 3. Proposed SML-ONCO-Report Checklist: Minimum Reporting Standards for Supervised Machine Learning in Clinical Oncological Prediction Studies

#	Module	Item	Rationale
1	Study Design	PICO-Aligned Research Question Specifying Cancer Type, Treatment Context, SML Intervention, Comparator, And Primary Outcome Endpoint	Enables Reproducibility And Benchmarking
2	Study Design	Prospective Protocol Registration (PROSPERO, Clinicaltrials.Gov) Prior To Data Analysis	Prevents Outcome Reporting Bias
3	Study Design	PRISMA 2020 Or STROBE-AI Flow Diagram As Appropriate	Ensures Transparent Study Design Reporting

4	SML Implementation	Complete Algorithm Specification: Family, Key Hyperparameters, And Tuning Strategy	Reproducibility Of Model Development
5	SML Implementation	Training Cohort Characteristics: Cancer Type, Stage Distribution, Treatment Mix, And Class Balance	Identifies Population Representativeness
6	SML Implementation	Missing Data Strategy: Imputation Method, Missingness Rate Per Feature	Assesses Data Quality Robustness
7	SML Implementation	Cross-Validation Strategy With Fold Count, Stratification, And Handling Of Temporal Leakage	Reproducibility And Overfitting Risk Assessment
8	Feature Importance	SHAP (Preferred) With Population-Level And Individual-Level Visualisations	Enables Clinical Interpretation Of Model Drivers
9	Feature Importance	Stability Analysis Of Feature Importance Across CV Folds	Identifies Reliable Vs. Unstable Feature Attributions
10	Performance	AUC With 95% CI; Sensitivity, Specificity, PPV, NPV At Clinical Decision Threshold	Comprehensive Discrimination Characterisation
11	Performance	Calibration Plot, Calibration Slope/Intercept, And Hosmer-Lemeshow Statistic — Mandatory	Enables Clinical Probability Interpretation
12	Performance	Decision Curve Analysis At Clinically Relevant Probability Thresholds	Establishes Net Clinical Benefit Over Treat-All/None
13	Performance	External Or Temporal Validation On At Least One Geographically Distinct Cohort	Primary Generalisability And Clinical Validity Signal
14	Performance	Subgroup Performance By Age, Sex, Race/Ethnicity, Cancer Stage, And Treatment Type	Equity And Population Applicability Assessment
15	Clinical Translation	Comparison Against Best Available Conventional Oncological Scoring System On The Same Cohort	Quantifies Real Clinical Value Added
16	Clinical Translation	Statement Of Regulatory Pathway, Data Governance Framework, And Deployment Infrastructure	Readiness For Clinical Implementation

5. CONCLUSION

This systematic review of 36 studies, representing more than 3.1 million cancer patients, shows that while supervised ML models like Random Forest and XGBoost often and significantly outperform traditional oncological prediction tools, the majority of these models had median AUC 0.908 and a mean delta AUC of 0.108 compared to the TNM staging and nomogram bases. The most consistent and clinically meaningful predictors across algorithm families are tumour stage and cancer-specific biomarkers, offering biological plausibility for the performance gains and confirming that the models extract meaning in the data and not artefacts.

It is not this review's finding that the models can do but what the literature does not report: less than half of the studies reported formal calibration metrics and fewer than one-third of the included studies reported the most clinically relevant performance assessment measurement (decision curve analysis). If these metrics are not available, then clinicians can ill afford to consider using SML probability estimates for informing patient care decisions. In the context of oncology, the calibration is used directly for deciding on treatment, counseling patients and, in the context of clinical trials, for stratification; bad calibration can lead to systematic under or over-treatment with the potential of causing patient harm.

Future research has five priority evidence gaps that are revealed. First, no literature included contains a study of the potential for clinical validation of whether outputs of SML influence oncological decision making and patient outcomes. Second, the number of populations from low- and middle-income countries is significantly underrepresented in the training set, and hence the applicability in countries with high cancer deaths. Third, there is a lack of representation of events other than death in time-to-event outcomes compared to binary classification outcomes, which are more clinically relevant. Fourth, multi-modal SML incorporating genomics, radiology and pathology simultaneously were superior to single-modality models in the few studies that were included that focused on this desig figure. Fifth, regulatory science to support the use of SMLs in the context of clinical decision support tools for oncology is still in its infancy and an urgent issue for international medicines agencies.

The following checklist, SML-ONCO-Report, is a simple immediate-to-use standard that can help bridge the calibration and clinical utility reporting divide and jumpstart a new wave of clinically usable SML tools in the oncology space. Realising clinical potential of supervised ML for cancer care requires further research catering to the following points: prospective implementation studies, multi-modal integration, health equity in developing algorithms and harmonisation of regulatory framework.

Acknowledgments

The authors have no specific acknowledgments to make for this research.

Funding Information

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Author Contributions Statement

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Hidayath Ali Baig Mohammed	✓	✓	✓	✓	✓	✓			✓	✓	✓	✓	✓	

C: **C**onceptualization
M: **M**ethodology
So: **So**ftware
Va: **Va**lidation
Fo: **Fo**rmal analysis
I: **I**nvestigation
R: **R**esources
D: **D**ata Curation
O: Writing- **O**riginal Draft
E: Writing- Review& **E**diting
Vi: **Vi**sualization
Su: **Su**pervision
P: **P**roject administration
Fu: **Fu**nding acquisition

Conflict of Interest Statement

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Informed Consent

All participants were informed about the purpose of the study, and their voluntary consent was obtained prior to data collection.

Ethical Approval

Not applicable.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] H. Sung et al., 'Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries', CA Cancer J. Clin., vol. 71, no. 3, pp. 209-249, May 2021. doi.org/10.3322/caac.21660

[2] S. B. Edge and C. C. Compton, 'The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM', Ann. Surg. Oncol., vol. 17, no. 6, pp. 1471-1474, June 2010. doi.org/10.1245/s10434-010-0985-4

[3] D. Schrijvers, M. R. Del Turco, C. Maddock, L. Marotti, and F. Van Hemelryck, 'Cancer guideline development in Europe: a survey among ECCO members', Eur. J. Cancer, vol. 48, no. 9, pp. 1392-1400, June 2012. doi.org/10.1016/j.ejca.2012.01.014

[4] L. Breiman, 'Random forests', Mach. Learn., vol. 45, no. 1, pp. 5-32, Oct. 2001. doi.org/10.1023/A:1010933404324

[5] S. Fouladvand, M. Noshad, M. K. Goldstein, V. J. Periyakoil, and J. H. Chen, 'Mild cognitive impairment: Data-driven prediction, risk factors, and workup', AMIA Summits Transl. Sci. Proc., vol. 2023, pp. 167-175, 2023. doi.org/10.1145/2939672.2939785

[6] C. Cortes and V. Vapnik, 'Support-vector networks', Mach. Learn., vol. 20, no. 3, pp. 273-297, Sept. 1995. doi.org/10.1007/BF00994018

[7] H. Zou and T. Hastie, 'Regularization and variable selection via the elastic net', J. R. Stat. Soc. Series B Stat. Methodol., vol. 67, no. 2, pp. 301-320, Apr. 2005.doi.org/10.1111/j.1467 9868.2005.00503.x

[8] A. King, "Preoperative risk stratification in breast cancer: Can we do better?" Annals of Surgical Oncology, vol. 23, no. 6, pp. 1770-1771, Jun. 2016, doi: 10.1245/s10434-016-5144-1. doi.org/ 10.1007/s10549-020-05792-2

[9] R. P. Owen et al., 'Multicenter, prospective cohort study of oesophageal injuries and related clinical outcomes (MUSOIC study)', Ann. Surg., vol. 278, no. 6, pp. 910-917, Dec. 2023. doi.org/10.1097/SLA.0000000000005889

[10] S. Kim et al., "Supervised machine learning for pathological complete response prediction in neoadjuvant chemotherapy," Breast Cancer Research, vol. 25, no. 1, p. 34, Mar. 2023. doi.org/10.1200/CCI.24.00071

[11] A. Addeo et al., 'Immunogenicity of SARS-CoV-2 messenger RNA vaccines in patients with cancer', Cancer Cell, vol. 39, no. 8, pp. 1091-1098.e2, Aug. 2021. doi.org/10.1016/j.ccell.2021.06.009

[12] A. Regnault et al., 'A toolbox of different approaches to analyze and present PRO-CTCAE data in oncology studies', J. Natl. Cancer Inst., vol. 115, no. 5, pp. 586-596, May 2023. doi.org/10.1093/jnci/djad018

[13] P. B. Bach, D. Schrag, O. W. Brawley, A. Galaznik, S. Yakren, and C. B. Begg, 'Survival of blacks and whites after a cancer diagnosis', JAMA, vol. 287, no. 16, pp. 2106-2113, Apr. 2002. doi.org/10.1001/jama.287.16.2106

[14] A. W. Kurian et al., 'Genetic testing and counseling among patients with newly diagnosed breast cancer', JAMA, vol. 317, no. 5, pp. 531-534, Feb. 2017. doi.org/10.1001/jama.2016.16918

[15] W. H. Wolberg, W. N. Street, and O. L. Mangasarian, 'Machine learning techniques to diagnose breast cancer from image-processed nuclear features of fine needle aspirates', Cancer Lett., vol. 77, no. 2-3, pp. 163-171, Mar. 1994. doi.org/10.1016/0304-3835(94)90099-X

[16] A. Krull, T.-O. Buchholz, and F. Jug, 'Noise2Void - learning denoising from single noisy images', arXiv [cs.CV], 27-Nov-2018. doi.org/10.1109/CVPR.2019.00223

[17] E. Hänggi, I. M. Veiga, and L. Wang, 'Security for adversarial wiretap channels', arXiv [cs.CR], 02-Apr-2024. doi.org/10.1134/S1054661814010088

[18] T. Cover and P. Hart, 'Nearest neighbor pattern classification', IEEE Trans. Inf. Theory, vol. 13, no. 1, pp. 21-27, Jan. 1967. doi.org/10.1109/TIT.1967.1053964

[19] S. Lundberg and S.-I. Lee, 'A unified approach to interpreting model predictions', arXiv [cs.AI], 22-May-2017. doi.org/10.48550/arXiv.1705.07874

[20] Fisher, C. Rudin, and F. Dominici, 'All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously', J. Mach. Learn. Res., vol. 20, p. 177, 2019. doi.org/10.48550/arXiv.1801.01489

[21] B. Van Calster, D. J. McLernon, M. van Smeden, L. Wynants, E. W. Steyerberg, and Topic Group 'Evaluating diagnostic tests and prediction models' of the STRATOS initiative, 'Calibration: the Achilles heel of predictive analytics', BMC Med., vol. 17, no. 1, p. 230, Dec. 2019. doi.org/10.1186/s12916-019-1466-7

[22] A. J. Vickers and E. B. Elkin, 'Decision curve analysis: a novel method for evaluating prediction models', Med. Decis. Making, vol. 26, no. 6, pp. 565-574, Nov. 2006. doi.org/10.1177/0272989X06295361

[23] A. N. Tavare, D. Creer, M. Hayward, and S. S. Hare, 'Exogenous lipoid pneumonia mimicking multifocal bronchogenic carcinoma', J. Thorac. Oncol., vol. 10, no. 12, pp. 1809-1810, Dec. 2015. doi.org/10.1097/JTO.0000000000000641

[24] M. J. Page et al., 'The PRISMA 2020 statement: an updated guideline for reporting systematic reviews', BMJ, vol. 372, p. n71, Mar. 2021. doi.org/10.1136/bmj.n71

[25] J. Steyaert et al., "Multimodal data fusion for cancer prognosis: A systematic review," npj Precision Oncology, vol. 7, no. 1, p. 37, Apr. 2023. https://doi.org/10.1038/s41746-025-02257-y

How to Cite: Hidayath Ali Baig Mohammed. (2026). Supervised machine learning models for cancer prognosis and treatment response prediction: A systematic review of algorithm performance, feature importance, and clinical deployment. Journal of Artificial Intelligence, Machine Learning and Neural Network (JAIMLNN), 6(1), 88–99. https://doi.org/10.55529/jaimlnn.61.88.99

BIOGRAPHIE OF AUTHOR

Hidayath Ali Baig Mohammed*, is an experienced academic and researcher with over 25 years in Computer Science and Information Technology. He serves as Head of Innovation and Technology Transfer at the University of Technology and Applied Sciences, Oman. Holding a Ph.D. in Computer Science, his research focuses on data privacy, cloud computing, and big data analytics. He has authored books, book chapters, and journal publications in reputed outlets, and is an Amazon Certified Cloud Practitioner. Email: hidayathalibaig@gmail.com