Journal of Innovative Science

ISSN(Online): 3082- 4435

Volume- 02| Issue- 01| 2026

Research Article

Received: 05-01-2026 | Accepted: 30-01-2026 | Published: 13-02-2026

Evaluation of Ai-Based Predictive Models for Early Cancer Detection: Statistical Considerations and Validation Methods

Robert Amevor[1], Mantey Joshua Kojo Aduampong[2] and Theophilus Asiedu Nketiah[3]

[1]*University of South Carolina, USA*

[2]*Northeastern University — Boston, MA, USA*

[3]*Kwame Nkrumah University of Science and Technology, Ghana*

Abstract: The earliest possible diagnosis is currently the most crucial element affecting the survival rate in cancer patients, but the existing screening and diagnostic methods are unable to provide early clues about malignancies. However, the last decade has witnessed the advent of artificial intelligence, or AI, in predictive oncology, which combines radiologic, molecular, and clinical variables for the improved specificity of diagnosis. The existing literature has revealed the efficacy of predictive models based on deep learning with the help of AI in the early diagnosis of malignancies, even beyond the capabilities of human interpretation. This narrative review synthesizes findings from recent studies that developed or validated AI models for early cancer diagnostics. The data show the tremendous success achieved in imaging radiomics, cfDNA fragmentomics, proteomics, or EHR-based model approaches with highly discriminative abilities, with AUC often well beyond 0.85, mostly for lung, colon, or breast cancers. However, their quality in terms of statistical validation or clinical transferability is often questionable, with many publications predominantly focusing on validation testing, without much care about the aspects of calibration, multicentric validation, or transferability to other populations. The article showcases the prominent paradigm shifts in the methods being employed, such as the need for explainable AI, the integration of deep learning with traditional machine learning, or the novel application of real-time or edge computing in the context of poor resources. Overfitting, imbalance, and lack of translucency in reporting are some of the challenges yet to be fully remedied in the pursuit of clinical-grade translatability. The future will require validation, benchmarking, or sound application practices to ensure these aspects. Although the predictive capabilities in cancer, utilizing AI, are revolutionizing the future of early cancer diagnosis, their successful implementation awaits improved statistical grounding, validation, and fair global application.

Keywords: Artificial intelligence, cancer detection, predictive modeling.

INTRODUCTION

The best mode of controlling cancer is detection, and the screening method currently being used has its shortcomings in detecting early lesions. decade of artificial intelligence and machine learning integration into the field of cancer care has challenged the limits of early diagnosis, particularly for complex, high-dimensional data, which exceeds the capacity of humans, which, when detected early on, has a high probability of being treated. The past ten years of artificial intelligence and machine learning integration into the field of cancer care have challenged the limits of the possibility of early diagnosis of complex high-dimensional data, beyond the capacity of people to process (Crosby *et al*., 2022; Fitzgerald *et al*., 2022). The past decade of integrating artificial intelligence and machine learning into cancer care has pushed the boundaries of early diagnosis, especially for complex, high-dimensional data that surpasses human processing capabilities. These deep learning algorithms can assist in screening people at risk of the more dangerous CT scans in imaging of the thorax, which shows the predictive power of artificial intelligence on common workflow in radiology imaging (Lu *et al*., 2020) in imaging. In addition, combinations of proteomics-based early markers and interpretable machine learning strategies have been able to predict the development of pancreatic adenocarcinoma, according to the work initiated by Blyuss and colleagues (2020).

The use of imaging-based AI to localize, segment, and estimate risks of lesions has increased tremendously. The results provided by Jiang *et al*. (2025) have indicated that three-dimensional convolutional networks are better as compared to two-dimensional networks when estimating the risk of lung cancer using the NLST CT scans, which has drawn attention to the impact of space encodings and pre-training methods on the predictive capabilities of such models. Gong and Bang (2025) suggested the use of optimized deep learning models on edge platforms in colonoscopy, which reduced the latency of real-time identification of lesions, which is a significant step towards the implementation of AI models in point-of-care applications. On the same note, Mendes *et al*. (2025) found that integration of deep neural features together with hand-designed attributes enhanced the prediction of risks of breast cancer based on prior mammograms. Similar studies

***Corresponding Author:** Robert Amevor

DOI- https://doi.org/10.5281/zenodo.18634390

by Javed *et al*. (2024) and Roslidar *et al*. (2020) aimed at describing the value of modality-specific optimization schemes, whether it is CT, thermal, or mammography, on the diagnostic performance of AI-based models.

All these models, combined with their emphasis on deep neural network architecture use, promise the future of mixed-method architectures that combine the benefits of statistical models with the representational power of deep networks. Molecular and liquid-biopsy modalities, similar to revitalizations in imaging, are being redesigned with AI integration. Guo *et al*. (2022) developed the cfDNA breakpoint motif logistic model that could reliably predict stage I lung adenocarcinoma in the outside data, whereas Xu and colleagues (2024) applied the cfDNA fragmentomics method in distinguishing between the malignant and benign pulmonary nodules with high AUCs on the external data. In the case of gastric cancer, Lee *et al*. (2021) managed to convert the tissue transcriptomics-based signature into the serum five-gene classifier, which implies the translational potential of the multi-omic AI workflows. Similarly, the same is achieved with the microRNA-based prediction models developed by Xu and colleagues (2024), whose ML models are able to reveal the latent connections between the circulating markers.

Although the progress rates are rather rapid, the absence of heterogeneity in the procedure of validation and the inconsistency of the process of validation are hampering the timeline of AI models from the development phase to the implementation phase of the healthcare system. Henriksen *et al*. (2025) emphasized the need for future multicenter validation of the population-based risk model LungFlag, stating that it was necessary. Other researchers, such as Wu and colleagues (2022), observed the absence of multicenter validation, in particular, the absence of validation in Asian populations, stating that the majority of studies are node-based, and they assess lung cancers using single-center datasets with tightly limited external validation, especially in Asian populations. The lack of reproducibility and fairness is caused by the prevalence of internal cross-validation, minimal calibration coverage reports, and inadequate representation of multiethnic data (Ben Ammar *et al*., 2024; Malik *et al*., 2024; Yao *et al*., 2025). Due to the growing regulatory attention to models, harmonized reporting formats, like TRIPOD-AI, sound external validation, and interpretable analytics (e.g., SHAP, LIME, or attention visualization) will become invaluable to establishing clinical trust and equitable translation (Li *et al*., 2025).

AI-BASED PREDICTIVE MODELS IN EARLY CANCER DETECTION

Evolution of Predictive Modeling

The application of predictive models in oncology began with mainstream statistical models, progressing to intricate data-driven models that combine radiologic, molecular, and clinical variables. Logistic regression, Cox proportional hazards regression, etc., enabled predictive scoring, but linearity in these models, coupled with the restrictive inclusion of variables, hampered predictive efficacy in high-dimensional biomedical datasets (Crosby *et al*., 2022; Fitzgerald *et al*., 2022). The availability of massive imaging, genomic, or EHR datasets triggered the shift towards the application of non-linear machine learning, followed by deep learning techniques that are efficient in uncovering complex relationships. For instance, Lu *et al*. (2020) illustrated the effectiveness of deep convolutional neural networks on chest radiographs in the early prediction of high-risk smokers, with the same efficacy as predicted outcomes from CT scans. Also, the study by Blyuss *et al*. (2020) recognized the effectiveness of gradient boosting in re-engineering proteomic predictions for pancreatic cancer, verifying the superiority of machine learning over other predictive models in uncovering complex biomedical relationships.

However, with the rising number of multimodal datasets, the integration of diverse streams has become common within predictive models. Lee *et al*. (2021) presented one such instance with the design of their serum transcriptomic classifier, which was developed from gene expression profiles of the gastric tissue, to demonstrate the applicability of omic data in developing minimally invasive diagnostic techniques. Li *et al*. (2025) presented another instance with their cross-attention transformer, which combined immune and histological information, to emphasize the shift toward biologically meaningful models. Simultaneous breakthroughs in the explainability of feature importance utilizing SHAP, Boruta, or attention maps emphasize the intersection of deep representations with classic ideas from stats to create models that are interpretable yet highly efficient (Ben Ammar *et al*., 2024).

Publisher: SARC Publisher

Categories of AI Models

Artificial intelligence predictive models for early cancer diagnosis can be classified generally into four different areas, which are imaging, molecular/liquid biopsy, EHR, and integration approaches. Imaging-based models remain the most extensively validated. Jiang *et al*. (2025) compared the performance of several three-dimensional CNN models on the NLST dataset, clearly demonstrating the effectiveness of volumetric encoding in increasing the discriminative capabilities of the model for predicting lung cancer risk. an ultra-light edge-deployable CNN solution that enabled the real-time classification of endoscopic lesions with high AUC values but with a tiny latency. Mendes *et al*. (2025) clearly observed the effectiveness of the convolutional feature-random forest classifier combination in the risk prediction of breast cancer, while the study by T. Wu *et al*. (2025) applied histopathology-based CNNs for automated cervical cancer triage. These studies illustrate how imaging AI is expanding from high-resource radiology to low-resource clinical environments (Aftab *et al*., 2025; Ahmad & Alqurashi, 2024).

Molecular, liquid biopsy, and sequencing platforms have made rapid strides with the advancement of sequencing, biosensing, and other ML platforms. A cfDNA breakpoint motif logistic model greater than 90% sensitive to stage I lung adenocarcinoma was proposed by Guo *et al*. (2022), with Xu *et al*. (2024) differentiating malignant from benign lesions in lung cancer by utilizing cfDNA fragmentomics. In gastric cancers, the validation of the five-gene serum transcriptomic assay was undertaken by Metcalf (2024), who proposed the cm-iMI candidates' multi-cancer early diagnostic signature based on circulating microRNAs. Other techniques, consisting of proteomics, metabolomics, pancreatic, or hepatic malignancies, have presented orthogonal validation with high sensitivity to pancreatic, hepatic malignancies, and other cancers developed by Bi *et al*. (2025). These findings highlight that ML's adaptability allows integration of complex molecular features across platforms.

EHR-driven models are currently showing promise as scalable solutions that exploit available lab variables and demographics. The applicability of their LungFlag approach, which used longitudinal blood test variables to predict lung cancer with an average of two years' advance, was proven better than rule-based machine learning models by Henriksen *et al*. (2025), who demonstrated that longitudinal blood test trends could predict lung cancer up to two years before diagnosis through their LungFlag algorithm, outperforming rule-based systems. Similar Gradient boosting and SVM-driven approaches have also been explored for colorectal and prostatic cancer, with the feasibility of passive surveillance established (Ahmad & Alqurashi, 2024).

The current state-of-the-art prediction with AI is covered by multimodal models. Li *et al*. (2025) proposed the CAMFormer, which is the integration of transcriptomics, imaging, and cross-attention transformers for the early diagnosis of cancers across the board with AUC values above 0.9. Similar models, which integrate radiomics with the profile of cfDNA, have presented improved discriminative capabilities with superior interpretations from the biological side (Hajjar *et al*., 2024; Malik *et al*., 2024). Such frameworks establish a paradigm in which image-derived phenotypes and molecular fingerprints coalesce into unified, biologically grounded predictors.

General Workflow

The standard process for developing predictive models with AI consists of the following steps: data collection, processing, selection of variables, model development, and validation. Data collection, which may be imaging, molecular, or EHR data, involves intensive selection, processing, augmentation, noise reduction, and standardization to maintain quality in the data set (Ben Ammar *et al*., 2024; Roslidar *et al*., 2020). Univariate or multivariate feature selection strategies, including the LASSO, Boruta, or SHAP algorithms, promote explainability by identifying biologically rational variables that are crucial within the predictive models, hence reducing complexity by identifying the least important variables for the predictive model prediction (Blyuss *et al*., 2020; Chen *et al*., 2025). Training approaches vary by modality; random forest, SVM, and gradient boosting remain common for structured datasets, whereas CNNs and transformer networks dominate imaging and omic contexts.

Model evaluation encompasses both discrimination and calibration of metrics. The area under the receiver operating characteristic curve (AUC), sensitivity, specificity, and F1-score are standard indices, though calibration measures like the Hosmer-Lemeshow test and calibration plots

Publisher: SARC Publisher

are reported less frequently (Fitzgerald *et al*., 2022; Malik *et al*., 2024). Validation frameworks include k-fold cross-validation, bootstrapping, temporal validation, and multicenter replication, though external validation across ethnically diverse populations remains scarce (Henriksen *et al*., 2025; Z. Wu *et al*., 2022). This underscores the ongoing need for methodological harmonization and adherence to TRIPOD-AI guidelines.

Emerging Trends

There are current trends towards the development of explainable, real-time, or federated AI solutions with translational robustness in mind. The feasibility of edge AI was shown by the work of Gong & Bang (2025) in developing colonoscopy AI that can run on the same devices without the need for heavy computational processing. SHAP explanation was also used in the work of Chen *et al*. (2025) on pancreatic cancer prediction, with the inclusion of cross-attention mechanisms by Li *et al*. (2025) to improve explainability, consistency, or bioplausibility, indicating the movement towards more explainable AI without losing precision, supported by Yao *et al*. (2025).

Federated, privacy-preserving platforms for AI are also on the rise, facilitating collaborative work across institutions without the need for centralizing data, which is crucial for dealing with the validation and challenges of reproducibility deficits in existing models (Aftab *et al*., 2025; Hajjar *et al*., 2024). The challenges of overfitting, class imbalance, and insufficient reporting on calibration impede the pursuit of generality in deep learning, despite recent advancements in the field (Ben Ammar *et al*., 2024; Fitzgerald *et al*., 2022). Future progress will depend on integrating rigorous statistical governance through standardized validation pipelines and transparent reporting to ensure clinical-grade reliability across populations and healthcare systems.

CANCER-SPECIFIC INSIGHTS

Lung Cancer

Lung cancer is one of the most developed areas in terms of early AI-based diagnostics, with the integration of image, molecular, and clinical models. The application of deep convolutional neural networks (CNN) to conventional chest radiographs has indicated the possibility of detecting high-risk patients on a population level, with capabilities comparable to those of low-dose CT scans (Lu *et al*., 2020). S. Jiang *et al*. (2025) also illustrated the superiority of three-dimensional CNN models over conventional 2D models, utilizing the capabilities of three-dimensional encoding, with AUC ranging between 0.75 and 0.92. Other reports (Ahmad & Alqurashi, 2024; Nimmagadda *et al*., 2025) confirm the models' strong sensitivity to network complexity, pre-training, and augmentation strategies.

Molecular/EHR-based approaches have also developed with relevance largely on an even footing. Near-perfect sensitivity was reached with cfDNA breakpoint motifs separating stage I lung adenocarcinoma from normals by Guo *et al*. (2022), followed by the integration of fragment-length and end-motif features into an ML model differentiating between benign and malignant lesions with high repeatability by Xu *et al*. (2024). Non-imaging predictions have also made strides, indicated by the LungFlag model developed by Henriksen *et al*. (2025), separating the risk years before the diagnosis based on available clinical information. However, lack of proper calibration reporting and demographic imbalance, among other sources, have remained common over various datasets described by D. Liu *et al*. (2021) and Malik *et al*. (2024). These findings collectively underscore that, despite consistently high internal discrimination (AUC 0.75-0.98), reproducibility and fairness persist as the principal obstacles to clinical translation.

Breast Cancer

Breast cancer prediction has largely revolved around mammography-driven risk prediction and density estimation until now. Mendes *et al*. (2025) improved mammographic feature extraction with deep CNNs by adding random forest classification. The applicability of transfer learning on imaging type was validated for improved generalizability with fewer models by Y. Jiang *et al*. (2025), while the integration of breast density measurements with ML techniques was explored by L. Liu *et al*. (2025), showing improved AUC with the combination of human-engineered representations with deep representations.

Reported discrimination typically ranges between 0.68 and 0.76, representing incremental yet meaningful gains over logistic regression benchmarks (Ben Ammar *et al*., 2024). Cross-center validation studies underscore that imaging heterogeneity stemming from variable acquisition protocols poses major barriers to reproducibility (Ahmad & Alqurashi, 2024; Roslidar *et al*., 2020). Nonetheless, breast cancer remains a proving

Publisher: SARC Publisher

ground for multimodal systems that merge imaging, hormonal, and demographic features to achieve explainability and lower false-positive rates in population screening.

Colorectal and Gastric Cancers

The application of AI in colorectal and gastric cancers illustrates the scope and viability of predictive models. A real-time edge computing AI system with the potential to classify lesions of the colon with the same accuracy as humans in colonoscopy procedures was developed by Gong & Bang (2025). Similarly, Kinar *et al*. (2016) developed gradient-boosted models trained on complete blood count features capable of identifying individuals at elevated colorectal cancer risk years before diagnosis.

In gastric cancer, Lee *et al*. (2021) constructed a serum-based five-gene transcriptomic signature that achieved AUCs above 0.90 across independent validation cohorts. Su *et al*. (2023) further validated gastroscopic CNNs that detect precancerous gastric lesions with comparable performance. In total, internal and external validations show AUC values between 0.82 and 0.95. However, reproducibility and cross-platform comparability are still limited (Fitzgerald *et al*., 2022; Wang & Su, 2025). These results underscore that, despite advancements in predictive accuracy, standardized calibration and external validation are crucial for clinical legitimacy.

Prostate and Pancreatic cancers.

The gastroscopic CNN models for the prediction of precancerous gastric lesions have also been validated by Chen *et al*. (2025), utilizing SHAP feature ranking to identify the prominent proteomic drivers underlying the prediction. A set of meta-analyses validated the consistent internally observed proteomic performance on separate proteomic datasets (Bi *et al*., 2025).

The combination of ML with urine metabolite profiles, followed by logistic regression analysis, yielded greater specificity than PSA alone in prostatic cancer, thereby validating the predictive utility of the biosignature approach (Malik *et al*., 2024; Vosoughi *et al*., 2025). However, there is mostly only internal validation, with little attention paid to external validation or the reporting of calibration, as observed in most involved studies (Fitzgerald *et al*., 2022). These limitations illustrate the broader need to couple AI's predictive strength with rigorous biostatistical governance to achieve clinically interpreted and generalizable outcomes.

Cervical and Laryngeal Cancers

AI-driven diagnostic solutions are rapidly being developed in the context of gynecological and head & neck cancers. Regarding cervical cancer, the study conducted by T. Wu *et al*. (2025) involved the application of CNN models in the automatic detection of transformation zones, precancerous areas, or both from digital cervical imaging, with high sensitivity validated across centers. For laryngeal cancer, the study conducted by Lam *et al*. (2025) involved the creation of an AI-assisted endoscopic classifier that outperformed experienced otolaryngologists in early lesion detection, with the ability to be explained by attention maps. Each study illustrates the potential for mitigating diagnostic inequalities but also the lack of standard validation protocols or consistent reporting of AI-driven diagnostic performance across studies (Hajjar *et al*., 2024; Yao *et al*., 2025).

Multicancer/Pan-Cancer Approaches

The recent development of pan-cancer models also signifies a shift toward generalized early screening. Li *et al*. (2025), in CAMFormer, an immune transcriptomic and histologic cross-attention transformer, reported an AUC of about 0.92 on various cancers with excellent calibrations. Other experiments with multi-modal architecture have reaffirmed the possibility of utilizing common molecular signatures across the board for universal screening of all cancers, documented in other studies (Aftab *et al*., 2025; Metcalf, 2024).

These integrative models represent a transition from organ-specific detection to systemic oncogenic profiling. However, challenges persist in dataset harmonization, feature attribution, and biological interpretability (Malik *et al*., 2024; Wang & Su, 2025). Despite these obstacles, multicancer AI approaches represent a critical step toward clinically unified and statistically transparent early detection systems.

STATISTICAL AND VALIDATION CONSIDERATIONS

Data Quality and Bias

The problem of statistical robustness of predictive models of early cancer diagnosis based on Artificial Intelligence techniques is largely linked to the quality, heterogeneity, and representativeness aspects of the input data used.

Publisher: SARC Publisher

The fact is, many studies are still conducted on single-center or single-ethnicity populations, leading to overoptimistic discrimination capabilities but poor transferability. For example, according to (Li *et al*., 2025), the predominance of European or East Asian populations was common in most lung cancer imaging datasets, which created transferability problems across ethnicities. Similar sampling problems have also occurred in gastric cancer or breast cancer models, with poor diversity of the study populations, leading to poor predictions on unseen datasets according to (Lee *et al*., 2021; Mendes *et al*., 2025). Small sample sizes and class imbalance further encourage overfitting, particularly when deep-learning architectures are trained without adequate augmentation or regularization.

To address these limitations, there has been a focus on learning transfer, data augmentation, or algorithms involving feature selection. Y. Jiang *et al*. (2025) demonstrated the effectiveness of starting convolutional neural networks with weights from ImageNet, leading to the reduction of overfitting and the stability of AUC measurements for lung cancer image detection. Data-level adjustment techniques, including rotations, flipping, or the addition of noise, have become common in imaging, while the application of LASSO or Boruta feature selection algorithms is widening in the context of molecular models to remove redundancy or noise from the information (Blyuss *et al*., 2020; Chen *et al*., 2025). EHR-based algorithms like LungFlag (Henriksen *et al*., 2025) exemplify the value of structured, routinely collected laboratory data for improving representativeness and reducing bias. Despite these improvements, transparent documentation of preprocessing, normalization, and feature-selection criteria remains inconsistent across publications, hindering reproducibility and meta-analytic synthesis.

Validation Paradigms

Validation strategy is a critical determinant of the clinical reliability of AI models in oncology, as performance estimates depend largely on the rigor of the approach used. The dominant method in the literature remains internal validation, in which model stability is assessed using k-fold cross-validation or bootstrapping. Despite their computational efficiency, these procedures often produce overly optimistic estimates when they use the same data distribution for both training and testing. Jiang *et al*. (2025) cautioned that such internal assessments may present an inflated perception of performance due to optimism bias. Malik *et al*. (2024) made the same point, saying that internal validation alone may not be enough to ensure clinical transferability.

Temporal validation remains uncommon despite its ability to evaluate model robustness over time and evolving clinical practices, as it tests the model on patients enrolled at later time points. Henriksen *et al*. (2025) validated the LungFlag model using a national registry across successive years, while Lee *et al*. (2021) demonstrated temporal generalization of a transcriptomic signature in gastric-cancer cohorts. These examples show that temporal validation offers valuable information regarding the longitudinal stability of model performance.

External or multicenter validation, the gold standard for checking generalizability, tests models on datasets that are completely separate from those of other institutions, regions, or health systems. This approach exposes performance variation arising from scanner differences, staining protocols, and population heterogeneity that internal validation cannot detect. Guo *et al*. (2022) and Xu *et al*. (2024) validated lung-cancer cfDNA classifiers across multiple centers, while T. Wu *et al*. (2025) showed that harmonized imaging protocols substantially reduced inter-site variability. Regulatory agencies increasingly recognize the importance of this type of evidence: the U.S. FDA (2021) now requires proof of AI/ML device performance in “real-world clinical settings” distinct from development data, and the EMA mandates validation across multiple independent sites for high-risk algorithms.

Collectively, these validation frameworks form a hierarchy of clinical evidence. Internal validation provides rapid feedback during model development but frequently overestimates performance. Temporal validation assesses longitudinal stability and resistance to data drift. External multicenter validation, though resource-intensive, remains essential for establishing true generalizability and meeting regulatory expectations for clinical translation. Ensuring that the validation strategy aligns with the model’s intended use for developmental, investigational, or clinical purposes will enhance reproducibility, transparency, and trust in AI-based cancer prediction.

Publisher: SARC Publisher

Performance Metrics

The evaluation of models is usually done depending on discrimination measures, AUC, accuracy, sensitivity, specificity, or F1 score. Among the studies, the value of AUC for controlled experiments ranges between 0.75 and 0.98, showcasing the discriminative ability of current AI systems (Gong & Bang, 2025; Lee *et al*., 2021; Lu *et al*., 2020). However, discrimination is only part of the data, as there is much more beyond that, especially with the application to real-world problems requiring predictive validity. Only a few studies, for example, (Li *et al*., 2025) and (Chen *et al*., 2025), have reported their calibration of agreement of predicted probability with outcomes or have relied on Hosmer & Lemeshow tests or calibration plots, even if only to show the lack of calibration.

New generations of evaluation tools promote an integrated perspective incorporating decision curve analysis and net benefit analysis to formally assess the value of the predictive tool. Such methods provide the connection between the accuracy of predictions for the statistician, on one side, and the value of predictions for the clinician, on the other, but are rarely extended beyond isolated examples of high impact.

Standardizing the simultaneous reporting of discrimination, calibration, and decision-curve metrics along with clear confidence intervals would markedly improve the interpretive validity and comparability of AI-based predictive models across cancer types.

Model Interpretability and Transparency

A major drawback in the application of AI in cancer diagnosis is integrating predictive capabilities with explainability. Deep learning models are "black boxes," which are detrimental from the interpretation perspective, making the application questionable for the scientific community to fully appreciate the efforts of the researchers involved in the study. However, the application of explainable AI (XAI) has led to the development of intensive solutions for the previously described problem. SHAP visualization was integrated with pancreatic cancer prediction models involving proteomic variables, according to Chen *et al*. (2025), while Li *et al*. (2025) embedded attention mechanisms into transformer networks to elucidate gene-gene interactions within multi-omic data. Mendes *et al*. (2025) combined random-forest feature-importance maps with CNN embeddings for mammographic risk prediction, improving both transparency and interpretability.

To emphasize the importance of these newer reporting guidelines, the study by Gong & Bang (2025)expanded the concept of explainability in real-time applications by integrating confidence visualization into edge-AI colonoscopy platforms, allowing the clinician to evaluate the degree of prediction confidence in real time. However, the valuable work conducted in these areas, although encouraging, is largely represented by the lack of crucial information on hyperparameter adjustment, feature ranking, or cross-validation strategies in the published models; hence, the need for reporting in accordance with TRIPOD-AI/PROBAST-AI.

GAPS, LIMITATIONS, AND FUTURE DIRECTIONS

Notwithstanding the tremendous success, predictive modeling based on AI for the early diagnosis of cancer is currently facing some inherent challenges, the biggest one being the lack of common calibration or reporting structures. Most papers on predictive AI models have yet to follow the TRIPOD-AI or even the PROBAST-AI reporting structure, hence the lack of commonality in validation reporting methods, confidence levels, or even calibration curves, even if the headline results, AUC, or sensitivity are outstanding (Henriksen *et al*., 2025; Lu *et al*., 2020)

The biggest hurdle is, of course, external validation. Most deep learning models, especially those applicable to lung, breast, or gastric malignancy, are often validated on the minutest possible demographics, usually abstracted from single-institution datasets. The problem is well elucidated by the study conducted by D. Liu *et al*. (2021) and Henriksen *et al*. (2025), wherein the efficiency of the algorithms is often hampered by their applicability on different or diverse populations, hence the burning need for multi-ethnic, across-country validation studies.

A second limitation is the current lack of biological explainability for existing AI solutions. Although some recent advances in the explainability of AIs, including SHAP values or attention heatmaps, have enhanced explainability, the current state-of-the-art is largely the result of mathematical or computational abstractions, which, by their nature, lack direct connection with biologically grounded processes. The inclusion of pathway-level, or molecular, prior knowledge into

Publisher: SARC Publisher

current AI solutions, for example, the proteomics model presented by Chen *et al*. (2025) and the cross-attention multi-omic framework of Li *et al*. (2025), could bridge this gap by linking computational signals to oncogenic processes. Such biologically informed models enhance both interpretability and clinical trust yet remain the exception rather than the norm.

Ethical considerations, along with concerns about fairness, also make the process of deploying more complex. There is the possibility of existing health inequalities being exacerbated by the inherent biases in the data used to train the models, especially if there is a paucity of data from minority communities or developing areas (Mendes *et al*., 2025)

None of these problems can be easily solved without the adoption of an open science ethos, or without the application of ethical structures of AI governance.

Looking ahead, several promising areas of innovation may help fill these gaps. Federated learning approaches that support the distributed learning of predictive models from different centers without requiring the transfer of private data will help improve the privacy afforded to patients even while retaining the benefits of statistical inference on large datasets (Gong & Bang, 2025). The development of edge-based AI solutions with real-time processing capabilities integrated with endoscopic or imaging procedures is an example of how translation may be realistically achieved to improve efficiency while offering greater time-saving efficiency. Interpretable machine learning architectures that bring together the strengths of traditional statistical inference with the strengths of deep learning-based representations are being advanced as promising lines of innovation aimed toward developing interpretable but highly efficient predictive models for health outcomes (Chen *et al*., 2025). Expanding datasets to encompass underrepresented populations and harmonizing acquisition protocols across centers will be essential to building globally generalizable AI systems. Ultimately, overcoming these barriers will demand not only algorithmic ingenuity but also rigorous data stewardship, methodological standardization, and sustained collaboration among clinicians, biostatisticians, and computational scientists.

CONCLUSION

AI predictive models are revolutionizing the landscape of early cancer diagnosis by integrating radiologic, molecular, and clinical aspects into holistic risk prediction platforms. In lung, breast, colorectal, gastric, and other cancers, these predictive models show strong discriminative power, with immense potential to supplement or even replace the existing diagnostic methods in these areas. However, successful translational validation is hampered by the lack of calibration, validation, or interpretation of these AI models. The realization of the full potential of predictive oncology with regards to its clinical capabilities will, therefore, depend on some matters, such as the validation of predictive oncology, reporting, multicenter validation, heterogeneity, matters relating to fairness, transparency, interpretation, validation, and approval, among others. The future of predictive oncology, therefore, relies on the intersection of federated learning, real-time edge AI, and the interpretable framework, otherwise known as the future frontier, artificial intelligence for the early detection of cancer while dealing with fairness, transparency, and validation.

REFERENCES

1. Ahmad, I., & Alqurashi, F. "Early cancer detection using deep learning and medical imaging: A survey." *Critical Reviews in Oncology/Hematology* 204 (2024): 104528.
2. Ben Ammar, M., Ayachi, F. L., Cardoso de Paiva, A., Ksantini, R., & Mahjoubi, H. "Harnessing deep learning for early breast cancer diagnosis: a review of datasets, methods, challenges, and future directions." *International Journal of Computing and Digital Systems* 16.1 (2024): 1643-1661.
3. Bi, Y., Li, D., Pang, R., Du, C., Li, D., Zhao, X., & Lv, H. "Diagnosis methods for pancreatic cancer with the technique of deep learning: a review and a meta-analysis." *Frontiers in Oncology* 15 (2025): 1597969.
4. Blyuss, O., Zaikin, A., Cherepanova, V., Munblit, D., Kiseleva, E. M., Prytomanova, O. M., & Crnogorac-Jurcevic, T. "Development of PancRISK, a urine biomarker-based risk score for stratified screening of pancreatic cancer patients." *British journal of cancer* 122.5 (2020): 692-696.
5. Chen, Y., Long, T., Wang, M., Liu, S., Lv, Z., Jiang, Y., & Liu, M. "Prospective cohort study

Publisher: SARC Publisher

integrating plasma proteomics and machine learning for early risk prediction of prostate cancer." *International Journal of Surgery* 111.9 (2025): 6123-6134.
6. Crosby, D., Bhatia, S., Brindle, K. M., Coussens, L. M., Dive, C., Emberton, M., & Balasubramanian, S. "Early detection of cancer." *Science* 375.6586 (2022): eaay9040.
7. Fitzgerald, R. C., Antoniou, A. C., Fruk, L., & Rosenfeld, N. "The future of early cancer detection." *Nature medicine* 28.4 (2022): 666-677.
8. Gong, E. J., & Bang, C. S. "Edge Artificial Intelligence Device in Real-Time Endoscopy for the Classification of Colonic Neoplasms." *Diagnostics* 15.12 (2025): 1478.
9. Guo, W., Chen, X., Liu, R., Liang, N., Ma, Q., Bao, H., & He, J. "Sensitive detection of stage I lung adenocarcinoma using plasma cell-free DNA breakpoint motif profiling." *EBioMedicine* 81 (2022).
10. Hajjar, M., Albaradei, S., & Aldabbagh, G. "Machine learning approaches in multi-cancer early detection." *Information* 15.10 (2024): 627.
11. Henriksen, M. B., Hilberg, O., Juul, C., Thomsen, R., Wen, S. W. C., Borg, M., & Menuhin-Gruman, I. "Maximizing Lung Cancer Screening in high-risk population leveraging ML-developed risk-prediction algorithms: Danish Retrospective Validation of LungFlagTM." *Clinical Lung Cancer* (2025).
12. Javed, R., Abbas, T., Khan, A. H., Daud, A., Bukhari, A., & Alharbey, R. "Deep learning for lungs cancer detection: a review." *Artificial Intelligence Review* 57.8 (2024): 197.
13. Jiang, S., Bennett, D. L., Chen, S., Toriola, A. T., & Colditz, G. A. "Automated breast density assessment for full-field digital mammography and digital breast tomosynthesis." *Cancer Prevention Research* 18.1 (2025): 23-29.
14. Jiang, Y., Ebrahimpour, L., Després, P., & Manem, V. S. "A benchmark of deep learning approaches to predict lung cancer risk using national lung screening trial cohort." *Scientific reports* 15.1 (2025): 1736.
15. Kinar, Y., Kalkstein, N., Akiva, P., Levin, B., Half, E. E., Goldshtein, I., & Shalev, V. "Development and validation of a predictive model for detection of colorectal cancer in primary care by analysis of complete blood counts: a binational retrospective study." *Journal of the American Medical Informatics Association* 23.5 (2016): 879-890.
16. Lam, S. W. S., Lee, M. H., Dorosan, M., Altonji, S., Tan, H. K., & Lee, W. T. "Use of a Preliminary Artificial Intelligence-Based Laryngeal Cancer Screening Framework for Low-Resource Settings: Development and Validation Study." *JMIR Formative Research* 9.1 (2025): e66110.
17. Lee, I. S., Ahn, J., Kim, K., Okugawa, Y., Toiyama, Y., Hur, H., & Goel, A. "A blood-based transcriptomic signature for noninvasive diagnosis of gastric cancer." *British Journal of Cancer* 125.6 (2021): 846-853.
18. Li, X., Hua, Y., Liu, H., Zhou, J., Zhang, Y., & Chen, H. "Peripheral blood multimodal integration via cross-attention for cancer immune profiling." *BMC cancer* 25.1 (2025): 1523.
19. Liu, D., Sun, X., Liu, A., Li, L., Li, S., Li, J., & Jiao, W. "Predictive value of a novel Asian lung cancer screening nomogram based on artificial intelligence and epidemiological characteristics." *Thoracic Cancer* 12.23 (2021): 3130-3140.
20. Liu, L., Zhou, P., Hou, L., Kao, C., Zhang, Z., Wang, D., & Yu, Z. "Development and performance of female breast cancer incidence risk prediction models: a systematic review and meta-analysis." *Annals of medicine* 57.1 (2025): 2534522.
21. Lu, M. T., Raghu, V. K., Mayrhofer, T., Aerts, H. J., & Hoffmann, U. "Deep learning using chest radiographs to identify high-risk smokers for lung cancer screening CT: development and validation of a prediction model." *Annals of internal medicine* 173.9 (2020): 704.
22. Malik, M. M. U. D., Alqahtani, M. M., Hadadi, I., Kanbayti, I., Alawaji, Z., & Aloufi, B. A. "Molecular Imaging Biomarkers for Early Cancer Detection: A Systematic Review of Emerging Technologies and Clinical Applications." *Diagnostics* 14.21 (2024): 2459.
23. Mendes, J., Oliveira, B., Araújo, C., Galrão, J., Garcia, N. C., & Matela, N. "You get the best of both worlds? Integrating deep learning and traditional machine learning for breast cancer risk prediction." *Computers in Biology and Medicine* 187 (2025): 109733.
24. Metcalf, G. A. "MicroRNAs: circulating biomarkers for the early detection of imperceptible cancers via biosensor and machine-learning advances." *Oncogene* 43.28 (2024): 2135-2142.

Publisher: SARC Publisher

25. Nimmagadda, S. M., Suryanarayana, G., Kumar, P. V., Vamsi, G. S., Imla, D. A., & Harika, C. "Early Detection of Lung Cancer Using Deep Learning Techniques: A Comprehensive Review." *Archives of Computational Methods in Engineering* (2025): 1-32.
26. Roslidar, R., Rahman, A., Muharar, R., Syahputra, M. R., Arnia, F., Syukri, M., & Munadi, K. "A review on recent progress in thermal imaging and deep learning approaches for breast cancer detection." *IEEE access* 8 (2020): 116176-116194.
27. Su, X., Liu, Q., Gao, X., & Ma, L. "Evaluation of deep learning methods for early gastric cancer detection using gastroscopic images." *Technology and Health Care* 31.1_suppl (2023): 313-322.
28. Vosoughi, P., Naghib, S. M., & Takdehghan, G. "Machine learning in cancer prognostic and diagnostic biomarkers: A promising approach for early cancer detection." *Sensors and Actuators Reports* (2025): 100385.
29. Wang, X., & Su, C. "Deep learning for cancer detection based on genomic and imaging data: A comprehensive review." *Cancer Management and Research* (2025): 2089-2125.
30. Wu, T., Wang, Y., Cui, X., Xue, P., & Qiao, Y. "AI-Based Identification Method for Cervical Transformation Zone Within Digital Colposcopy: Development and Multicenter Validation Study." *JMIR cancer* 11.1 (2025): e69672.
31. Wu, Z., Wang, F., Cao, W., Qin, C., Dong, X., Yang, Z., & He, J. "Lung cancer risk prediction models based on pulmonary nodules: a systematic review." *Thoracic cancer* 13.5 (2022): 664-677.
32. Xu, S., Luo, J., Tang, W., Bao, H., Wang, J., Chang, S., & Wu, X. "Detecting pulmonary malignancy against benign nodules using noninvasive cell-free DNA fragmentomics assay." *ESMO open* 9.8 (2024): 103595.
33. Yao, I. Z., Dong, M., & Hwang, W. Y. "Deep learning applications in clinical cancer detection: a review of implementation challenges and solutions." *Mayo clinic proceedings: digital health* (2025): 100253.
34. Aftab, M., Mehmood, F., Zhang, C., Nadeem, A., Dong, Z., Jiang, Y., & Liu, K. "AI in oncology: transforming cancer detection through machine learning and deep learning applications." *arXiv preprint arXiv:2501.15489* (2025).
35. US Food and Drug Administration, "Artificial intelligence and machine learning in software as a medical device action plan." (2022).

Source of support: Nil; **Conflict of interest:** Nil.

Cite this article as:
Amevor, R., Aduampong, M. J. K. & Nketiah, T. A. “Evaluation of Ai-Based Predictive Models for Early Cancer Detection: Statistical Considerations and Validation Methods." *Jr. Inn. Sci.* 2.1 (2026): pp 1-10.

Publisher: SARC Publisher