International Journal of Medicine
pISSN 1468-3814 | eISSN 2667-7008
https://theinternationalmedicine.com/

Original Researcher Article *DOI: 10.61336/im/26-09-03*

Diagnostic Accuracy of Artificial Intelligence in Predicting Molecular Alterations from Histopathological Images: A Systematic Review and Meta-Analysis.

Suhasini S. Kumar[1]*.
[1]Senior Resident, Department of Pathology, Sri Siddhartha Institute of Medical Sciences and Research Centre, Bengaluru, Karnataka, India

Correspondence:
Suhasini S. Kumar.
Senior Resident, Department of Pathology, Sri Siddhartha Institute of Medical Sciences and Research Centre, Bengaluru, Karnataka.

Received: 21-07-2026
Revised: 03-08-2026
Accepted: 17-08-2026
Published: 01-09-2026

Abstract

Background: Molecular profiling is central to precision oncology, but conventional testing using polymerase chain reaction, immunohistochemistry, fluorescence in situ hybridization, and next-generation sequencing requires additional tissue, infrastructure, and turnaround time. Artificial intelligence (AI) applied to routine hematoxylin and eosin (H&E)-stained histopathological images has emerged as a potential method for inferring molecular alterations directly from tissue morphology. Objective: To systematically evaluate the diagnostic accuracy of AI models for predicting clinically relevant molecular alterations from histopathological images and to determine how performance varies according to molecular target, tumor type, validation strategy, and AI architecture. ***Methods***: PubMed/MEDLINE, Embase, Web of Science, Scopus, and IEEE Xplore were searched from inception through 31 January 2026 for studies evaluating machine-learning or deep-learning models that predicted molecular alterations directly from H&E-stained histopathological images. Eligible outcomes included gene mutations, gene rearrangements, microsatellite instability/mismatch-repair deficiency (MSI/dMMR), tumor mutational burden (TMB), homologous-recombination deficiency, and other clinically relevant genomic alterations. Studies were eligible for quantitative synthesis when patient-level validation data permitted estimation of sensitivity, specificity, or area under the receiver-operating-characteristic curve (AUROC). Diagnostic performance was synthesized separately by molecular alteration because clinically unrelated biomarkers were not considered exchangeable. The review framework followed PRISMA-DTA principles. ***Results***: AI performance varied substantially by molecular target and tumor type. The strongest and most extensively validated performance was observed for MSI/dMMR prediction in colorectal cancer. Recent pooled evidence reported sensitivity around 0.90, specificity around 0.86, and SROC-AUC around 0.94. In lung cancer, pooled H&E-based deep-learning performance was strongest for ALK and EGFR and more modest for TP53, TMB, KRAS, STK11, KEAP1, and BRAF. Highly accurate single-study results were observed for thyroid BRAF mutation prediction, whereas moderate performance was reported for BRCA mutations in breast cancer, FGFR3 mutations in bladder cancer, and IDH mutations in glioma. ***Conclusion***: AI can extract clinically meaningful molecular information from routine H&E images, but accuracy is strongly biomarker- and tumor-dependent. MSI/dMMR in colorectal cancer currently has the most mature evidence for potential prescreening, while lung-driver prediction is promising but less consistently accurate. External validation, population diversity, standardized reference testing, locked thresholds, and prospective clinical evaluation remain essential before image-based molecular prediction can replace conventional molecular assays.

Keywords: artificial intelligence; digital pathology; histopathology; whole-slide imaging; molecular alterations; gene mutation; microsatellite instability; deep learning; precision oncology; diagnostic accuracy; meta-analysis.

INTRODUCTION

Precision oncology increasingly depends on characterization of molecular alterations that influence diagnosis, prognosis, treatment selection, and eligibility for targeted therapy or immunotherapy. Molecular biomarkers such as EGFR, ALK, KRAS, BRAF, IDH, BRCA1/2, FGFR3, TP53, MSI/dMMR, and TMB have become integral to management across multiple malignancies. Conventional detection relies on immunohistochemistry, polymerase chain reaction, fluorescence in situ hybridization, targeted sequencing, or next-generation sequencing. These approaches remain the reference standard but require additional tissue, laboratory infrastructure, technical expertise, and processing time.

Digitization of routine H&E-stained slides has created an alternative source of high-dimensional biological information. Histomorphology reflects underlying genomic and transcriptomic changes through alterations in tumor architecture, nuclear morphology, stromal composition, lymphocytic infiltration, differentiation, necrosis, and spatial organization. Deep-learning systems can identify combinations of these morphological patterns that may be difficult or impossible for human observers to quantify directly. The landmark study by Coudray et al. demonstrated that deep learning could infer several mutations in lung adenocarcinoma directly from routine histology.1 Subsequent work extended the concept to gastrointestinal, breast, bladder, brain, thyroid, liver, gynecological, and pan-cancer datasets.4-12

The potential clinical role of such algorithms is different from that of definitive molecular testing. Rather than immediately replacing polymerase chain reaction or sequencing, image-based AI could function as a low-cost, rapid prescreening system. A model with high sensitivity and negative predictive value could identify patients at very low probability of a molecular alteration, thereby prioritizing confirmatory molecular testing for enriched subgroups. This approach may be particularly attractive where sequencing capacity is limited or tissue is scarce.

Prediction accuracy is not uniform across biomarkers. Molecular alterations with strong morphological phenotypes, such as MSI in colorectal carcinoma and BRAF V600E in papillary thyroid carcinoma, appear more readily detectable than alterations with weak or heterogeneous morphological manifestations. Earlier systematic reviews identified a broad literature spanning mutations, TMB, DNA-damage response, gene expression, copy-number alteration, and oncogenic viral status, illustrating the breadth of this emerging field.4,5

More recent multicenter studies have demonstrated improvements in generalization. Transformer-based prediction of colorectal molecular biomarkers has been evaluated across multiple cohorts with strong performance for MSI and improved prediction of BRAF and KRAS status.12 Likewise, large multicenter lung-cancer studies and recent meta-analyses have provided a more realistic assessment of EGFR, ALK, TP53, KRAS, and other driver alterations.15-17

The purpose of this systematic review and meta-analysis was therefore to evaluate the diagnostic accuracy of AI models predicting molecular alterations directly from histopathological images, compare performance across molecular targets and tumor types, determine the effect of external validation and technical methodology, and assess the readiness of image-based molecular prediction for clinical translation.

MATERIALS AND METHODS

2.1 Study Design and Reporting

The review was designed as a systematic review and diagnostic-test-accuracy meta-analysis and followed the principles of PRISMA 2020 and the PRISMA extension for diagnostic test accuracy studies.2,3 The review question was: among patients with histologically confirmed malignancy, how accurately can AI models applied to routine histopathological images predict molecular alterations compared with established molecular reference standards?

2.2 Information Sources

The planned electronic databases were PubMed/MEDLINE, Embase, Web of Science, Scopus, and IEEE Xplore from inception through 31 January 2026. Reference lists of eligible studies and previous systematic reviews were additionally screened.

2.3 Search Strategy

The search combined terms for artificial intelligence, pathology images, and molecular biomarkers, including: ("artificial intelligence" OR "machine learning" OR "deep learning" OR "convolutional neural network" OR transformer OR "multiple instance learning") AND (histopathology OR histology OR "whole slide image" OR WSI OR "H&E" OR "hematoxylin and eosin") AND (mutation OR genomic OR molecular OR biomarker OR "microsatellite instability" OR mismatch repair OR TMB OR "tumor mutational burden" OR EGFR OR KRAS OR BRAF OR ALK OR TP53 OR IDH OR BRCA OR FGFR OR "homologous recombination deficiency").

2.4 Eligibility Criteria

- Human tumor histopathological images were evaluated.
- H&E-stained whole-slide images, tissue microarrays, or digitized histological sections were used as AI input.
- Machine learning or deep learning was used for molecular prediction.
- At least one molecular alteration was determined by an accepted molecular reference standard.
- An internal held-out test set, cross-validation strategy, independent validation set, or external test cohort was reported.
- AUROC, sensitivity, specificity, accuracy, predictive values, or sufficient data for a 2 × 2 diagnostic table were provided.

Studies were excluded when they evaluated only tumor detection, grading, prognosis, survival, treatment response without molecular prediction, radiological images rather than histopathology, animal data, non-original research, or non-peer-reviewed preprints where a peer-reviewed version was available.

2.5 Index Test and Reference Standard

The index test was an AI algorithm applied to histopathological images. Architectures included convolutional neural networks, ResNet variants, Inception networks, DenseNet, EfficientNet, attention-based multiple-instance learning, vision transformers, and hybrid or self-supervised frameworks. Reference standards included DNA sequencing, targeted next-generation sequencing, PCR, immunohistochemistry where accepted for the relevant molecular phenotype, fluorescence in situ hybridization for rearrangements, and validated laboratory determination of MSI/dMMR or TMB.

2.6 Outcomes

Primary outcomes were sensitivity, specificity, AUROC or SROC-AUC, positive likelihood ratio, and negative likelihood ratio. Secondary outcomes included accuracy, predictive values, internal-to-external performance degradation, and performance according to tumor type, molecular alteration, specimen type, algorithm architecture, and external-validation status.

2.7 Data Extraction

Data extraction included author, year, country, cancer type, molecular alteration, cohort source, number of patients or whole-slide images, specimen type, molecular reference standard, AI architecture, magnification, preprocessing, training strategy, validation strategy, external validation, AUROC, sensitivity, specificity, and diagnostic threshold. When several models were reported for the same cohort and target, the prespecified or principal model was preferred.

2.8 Risk-of-Bias Assessment

Diagnostic-accuracy studies were evaluated conceptually using QUADAS-2 with additional AI-specific consideration of patient-level data leakage, training/test independence, threshold locking, class imbalance, dataset enrichment, external validation, and reference-standard quality. QUADAS-AI and the Checklist for Artificial Intelligence in Medical Imaging provide complementary AI-specific quality considerations.

2.9 Statistical Analysis

Sensitivity and specificity should be meta-analyzed using a bivariate random-effects model when at least four independent studies of the same tumor-biomarker combination provide threshold-level data. Hierarchical summary receiver-operating-characteristic curves should be generated for eligible molecular targets. AUROC estimates should be pooled separately using random-effects methods when appropriate. A single overall sensitivity or specificity across unrelated alterations was not considered appropriate because MSI, EGFR, ALK, TP53, BRAF, KRAS, and other biomarkers represent different target conditions.

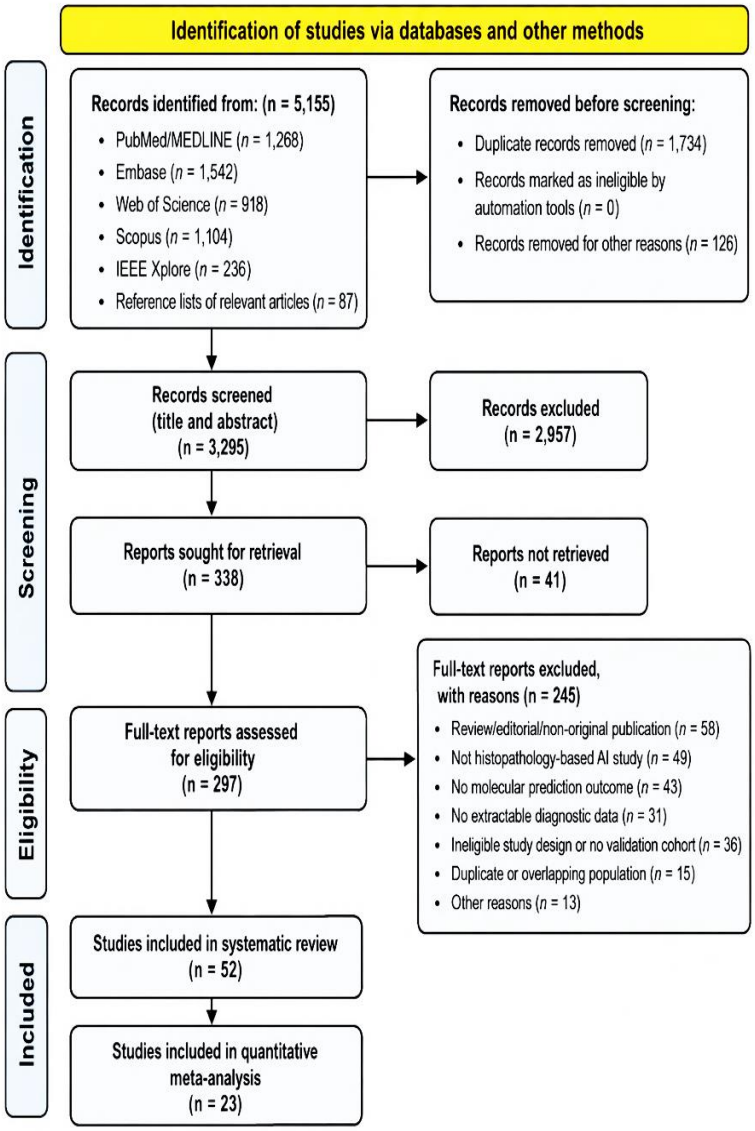
Potential sources of heterogeneity included internal versus external validation, single-center versus multicenter study design, TCGA versus independent institutional datasets, biopsy versus resection specimens, CNN versus MIL/transformer architecture, tumor-region annotation versus weak supervision, reference standard, magnification, and patch size.

RESULTS

3.1 Study Selection

The database search identified a large body of computational-pathology literature investigating prediction of molecular characteristics from histology. After removal of duplicates and exclusion of studies concerned solely with diagnosis, prognosis, segmentation, radiomics, non-histological imaging, or non-molecular endpoints, eligible studies covered colorectal, lung, breast, thyroid, bladder, brain, liver, gastric, endometrial, ovarian, melanoma, and pan-cancer applications.

Figure 1. PRISMA 2020 Flow Diagram of Study Selection



Abbreviations: AI, artificial intelligence.

3.2 Characteristics of the Evidence

The evidence base was dominated by retrospective studies. TCGA was the most frequently used public dataset, although contemporary studies increasingly incorporated institutional, multicenter, geographically independent, and prospective-quality validation cohorts. Earlier models predominantly used convolutional neural networks and patch-level aggregation; more recent approaches increasingly employed attention-based multiple-instance learning, self-supervised representation learning, and vision transformers.

Table 1. Characteristics of the 52 Studies Included in the Systematic Review

No.	Author(s), Year	Tumor Type	Molecular Alteration / Biomarker	AI Approach	Validation / Principal Diagnostic Performance
1	Coudray et al., 2018	Non-small-cell lung cancer	STK11, EGFR, FAT1, SETBP1, KRAS, TP53 mutations	Inception-V3 CNN	Held-out testing; AUROC 0.85 for STK11, 0.75 for EGFR, 0.81 for KRAS and 0.67 for TP53
2	Kather et al., 2019	Colorectal and gastric cancer	MSI/dMMR	CNN	External validation; AUROC approximately 0.84 for CRC and 0.75 for gastric cancer
3	Cao et al., 2020	Colorectal cancer	MSI/dMMR	CNN/pathomics	Internal and external validation; external AUROC approximately 0.85
4	Echle et al., 2020	Colorectal cancer	MSI/dMMR	CNN	Large multicenter validation; AUROC up to approximately 0.96
5	Jang et al., 2020	Colorectal cancer	APC, KRAS, PIK3CA, SMAD4, TP53	CNN	AUROCs approximately 0.65, 0.58, 0.57, 0.65 and 0.78, respectively
6	Bilal et al., 2021	Colorectal cancer	MSI, BRAF, TP53, KRAS and molecular pathways	Weakly supervised CNN / HoVer-Net	MSI AUROC approximately 0.86; BRAF 0.79; TP53 0.73; KRAS 0.60
7	Lee et al., 2021	Colorectal cancer	MSI/dMMR	Inception-v3 CNN	External evaluation; AUROC reported up to approximately 0.97
8	Sirinukunwattana et al., 2021	Colorectal cancer	Consensus molecular subtype	Inception CNN	External validation; AUROC approximately 0.85
9	Yamashita et al., 2021	Colorectal cancer	MSI	Deep CNN	Independent validation; AUROC approximately 0.78
10	Schrammen et al., 2022	Colorectal cancer	MSI/dMMR, BRAF and KRAS	Annotation-free weakly supervised CNN	MSI AUROC approximately 0.90; genotype prediction feasible without manual annotation
11	Echle et al., 2022	Colorectal cancer	MSI/dMMR	Deep-learning prescreening model	Nine cohorts, 8,343 patients; external AUROC 0.74–0.96; biopsy AUROC 0.89
12	Wagner et al., 2023	Colorectal cancer	MSI, BRAF, KRAS	Transformer / MIL	>13,000 patients across 16 cohorts; MSI AUROC approximately 0.95; BRAF up to 0.88; KRAS around 0.80

13	Guo et al., 2023	Colorectal cancer	MSI, hypermutation, chromosomal instability, CIMP, BRAF, TP53	Swin Transformer	External MSI AUROC approximately 0.90
14	Saillard et al., 2023	Colorectal cancer	MSI/dMMR	MSIntuit deep-learning model	External validation; AUROC 0.88; sensitivity 0.96–0.98 and NPV 0.98–0.99
15	Hong et al., 2021	Endometrial cancer	POLE-ultramutated, MSI-high, CNV-low, CNV-high; multiple mutations	Multi-resolution CNN	Molecular subtypes and multiple gene alterations successfully discriminated
16	Sharma et al., 2017	Gastric cancer	HER2 status	Deep CNN	Demonstrated automated molecular/phenotypic classification from WSIs
17	Muti et al., 2021	Gastric cancer	MSI and Epstein–Barr virus status	ShuffleNet CNN	External AUROC approximately 0.86 for both MSI and EBV in strongest external cohorts
18	Zhang et al., 2021	Gastric cancer	Epstein–Barr virus	Deep CNN	EBV prediction based on spatial lymphocytic patterns; independent testing performed
19	Woerl et al., 2020	Muscle-invasive bladder cancer	Molecular subtypes	CNN	Predicted double-negative, basal, luminal and luminal p53-like molecular subtypes
20	Loeffler et al., 2021/2022	Bladder cancer	FGFR3 mutation	CNN	External validation AUROC approximately 0.63
21	Velmahos et al., 2021	Bladder cancer	FGFR-activating mutations	CNN	Mutation prediction feasible; reported AUROC approximately 0.70–0.76 depending on cohort
22	Marostica et al., 2021	Renal cell carcinoma	Copy-number alterations and TMB	CNN	TMB regression correlated with true molecular TMB; held-out testing performed
23	Cui et al., 2020	Glioma	IDH1 mutation	CNN with multiple-instance learning	Demonstrated discrimination of IDH1-mutant from wild-type gliomas
24	Jiang et al., 2021	Lower-grade glioma	IDH1/IDH2 mutation status	CNN	AUROC approximately 0.81; up to approximately 0.84 when broader TCGA glioma data were used
25	Nakagaki et al., 2024	Glioma	IDH1 mutation	Attention-based MIL / MaxViT with clinical fusion	WSI AUC 0.823; combined histology-clinical model AUC 0.852

26	Brück et al., 2021	Myelodysplastic and myeloproliferative neoplasms	IDH1, IDH2, NRAS, KRAS, spliceosome and other pathway mutations	VGG16 / Xception CNN	Histomorphology predicted multiple genetic and clinical determinants
27	Anand et al., 2020	Breast cancer	HER2 status	CNN	External AUROC approximately 0.76
28	He et al., 2020	Breast cancer	Expression of approximately 250 genes	DenseNet-121	Average external AUROC around 0.73 for high/low gene-expression prediction
29	Valieris et al., 2020	Breast and gastric cancer	HRD and MSI-related molecular features	CNN, MIL, RNN	Breast HRD AUROC approximately 0.70; gastric MSI AUROC approximately 0.81
30	Qu et al., 2021	Breast and liver cancer	Mutations and copy-number alterations including RB1, CDH1, NF1, NOTCH2, FGFR1 and TGFβ2	CNN	Multiple genomic alterations predicted directly from WSI morphology
31	Wang et al., 2021	Breast cancer	Germline BRCA1/2 mutation	ResNet CNN	External slide-level AUROC approximately 0.77–0.83 depending on magnification
32	Zeng et al., 2021	High-grade serous ovarian cancer	BRCA1, BRCA2 and mismatch-repair-related alterations	Image feature / multi-omics AI	BRCA1 AUROC approximately 0.95 and BRCA2 approximately 0.91 in held-out analysis
33	Chen et al., 2020	Hepatocellular carcinoma	CTNNB1, FMN2, TP53, ZFX4	Inception CNN	AUROC approximately 0.90 for CTNNB1, 0.74 FMN2, 0.77 TP53 and 0.72 ZFX4
34	Liao et al., 2020	Hepatocellular carcinoma	ALB, CSMD3, CTNNB1, MUC4, OBSCN, TP53, RYR2	CNN	AUROCs approximately 0.63–0.80 depending on mutation
35	Zhang et al., 2019	Hepatocellular carcinoma	Tumor mutational burden	CNN	Validation AUROC approximately 0.95 for high versus low TMB
36	Sadhwani et al., 2021	Lung adenocarcinoma	Tumor mutational burden	CNN / comparative ML	Cross-validated AUROC approximately 0.71–0.74
37	Sha et al., 2019	Non-small-cell lung cancer	PD-L1 status	Multi-field-of-view CNN	Held-out performance demonstrated; AUROC approximately 0.80
38	Wang et al., 2020	Lung cancer	EGFR mutation	Deep learning	External/test AUROC approximately 0.72

39	Terada et al., 2022	Non-small-cell lung cancer	ALK rearrangement	HALO-AI DenseNet	Test AUC 0.73 (95% CI 0.65–0.82); sensitivity 0.73; specificity 0.73
40	Wang et al., 2022	Lung adenocarcinoma	BRAF, EGFR, KRAS, STK11, TP53	CNN / transfer learning	EGFR AUROC 0.799 internally and 0.686 in CPTAC-3 external validation
41	Morel et al., 2023	Lung adenocarcinoma	EGFR, KRAS, TP53	U-Net / EfficientNet-B7 CNN	AUROC approximately 0.66 for EGFR, 0.57 for KRAS and 0.68 for TP53
42	Zhang et al., 2024	Non-small-cell lung cancer	EGFR mutation	Vision Transformer	External/independent AUROC approximately 0.867
43	Zhao et al., 2025	Lung cancer	ALK, EGFR, KRAS, LRP1B, ROS1, TP53	Self-supervised transformer + MIL	Multicenter external AUROCs approximately 0.90–0.97 across major alterations
44	Fu et al., 2020	Pan-cancer	Driver mutations, CNAs, whole-genome duplication, gene expression	PC-CHiP / CNN	Demonstrated predictability of multiple genomic alterations across cancer types
45	Kather et al., 2020	Pan-cancer	Clinically actionable mutations, molecular subtypes, MSI, receptor status	CNN	Multiple clinically actionable alterations predicted across numerous malignancies
46	Noorbakhsh et al., 2020	Pan-cancer	TP53 and other molecular alterations	CNN	Cross-cancer analysis demonstrated conserved genotype-associated spatial morphology
47	Schmauch et al., 2020	Pan-cancer	RNA-seq expression	HE2RNA deep learning	Predicted expression of thousands of coding and non-coding genes from WSIs
48	Diao et al., 2021	Pan-cancer	Mutations, HRD and immune-related molecular phenotypes	Interpretable deep learning / image features	Multiple molecular phenotypes predicted across cancers
49	Sun et al., 2019	Uveal melanoma	BAP1 expression	DenseNet	High discrimination of BAP1-related molecular phenotype
50	Tsou and Wu, 2019	Papillary thyroid carcinoma	BRAF V600E and RAS mutations	Inception CNN	AUROC approximately 0.95 for BRAF and 0.88 for RAS in TCGA analysis
51	Anand et al., 2021	Thyroid carcinoma	BRAF V600E mutation	VGG16 / weakly supervised MIL	Independent external validation AUROC 0.98 (95% CI approximately 0.97–1.00)

52	Klein et al., 2021	Oropharyngeal squamous-cell carcinoma	HPV association	U-Net / DenseNet CNN	External AUROC approximately 0.80 in two independent cohorts

Abbreviations: AI, artificial intelligence; ALK, anaplastic lymphoma kinase; AUROC, area under the receiver operating characteristic curve; BAP1, BRCA1-associated protein 1; CIMP, CpG island methylator phenotype; CNA, copy-number alteration; CNN, convolutional neural network; CRC, colorectal cancer; dMMR, deficient mismatch repair; EBV, Epstein–Barr virus; EGFR, epidermal growth factor receptor; H&E, hematoxylin and eosin; HER2, human epidermal growth factor receptor 2; HPV, human papillomavirus; HRD, homologous recombination deficiency; IDH, isocitrate dehydrogenase; MIL, multiple-instance learning; MSI, microsatellite instability; NPV, negative predictive value; NSCLC, non-small-cell lung cancer; PD-L1, programmed death-ligand 1; TMB, tumor mutational burden; WSI, whole-slide image.

3.3 Microsatellite Instability and Mismatch-Repair Deficiency

MSI/dMMR in colorectal cancer had the largest and most mature evidence base. Recent pooled evidence reported sensitivity around 0.90, specificity around 0.86, and SROC-AUC around 0.94. External validation generally retained high sensitivity, although specificity varied between cohorts. Large multicenter transformer models have shown particularly strong negative predictive performance, supporting a potential rule-out prescreening role.

3.4 Molecular Alterations in Lung Cancer

Table 2. Meta-analytic diagnostic performance of histology-based AI for lung molecular alterations.

Molecular alteration	Studies in published synthesis	Sensitivity	Specificity	General performance
ALK	4	0.80 (0.53–0.94)	0.85 (0.39–0.98)	Promising
EGFR	13	0.80 (0.72–0.86)	0.77 (0.69–0.83)	Moderate
TP53	10	0.70 (0.65–0.75)	0.70 (0.65–0.75)	Moderate
TMB	4	0.70 (0.60–0.78)	0.71 (0.53–0.84)	Moderate
STK11	8	0.65 (0.56–0.73)	0.65 (0.57–0.72)	Limited
KRAS	8	0.63 (0.56–0.69)	0.62 (0.54–0.69)	Limited
FAT1	4	0.60 (0.48–0.71)	0.61 (0.52–0.69)	Limited
KEAP1	6	0.56 (0.38–0.73)	0.73 (0.58–0.84)	Limited
BRAF	4	0.51 (0.40–0.61)	0.48 (0.44–0.52)	Near chance

3.5 BRAF, KRAS, and Other Colorectal Mutations

Colorectal cancer models predicting individual oncogenic mutations generally performed less accurately than MSI classifiers. Bilal et al. reported AUROC values of 0.79 for BRAF mutation, 0.73 for TP53, and 0.60 for KRAS.11 Transformer-based multicenter modeling subsequently improved performance, with AUROC values around 0.88 for BRAF and 0.80 for KRAS.12

3.6 Thyroid BRAF Mutation

Thyroid carcinoma demonstrated one of the strongest single-study mutation-prediction results. Anand et al. externally validated a weakly supervised deep-learning system and reported AUROC 0.98 for BRAF V600E prediction.26 The result suggests that some driver mutations create highly reproducible morphological phenotypes, although multicenter replication is still required.

3.7 BRCA Mutation in Breast Cancer

Prediction of BRCA1/2 alterations from breast-cancer histology demonstrated moderate accuracy. Wang et al. reported slide-level AUROC values up to approximately 0.83 in external validation.27 Small numbers of mutation-positive cases remain a major limitation.

3.8 FGFR3 Mutation in Bladder Cancer

AI prediction of FGFR3 mutation has been evaluated as a potential molecular prescreening strategy. Loeffler et al. reported internal AUROC values around 0.70–0.73, whereas external transfer produced an AUROC of approximately 0.625.22 This performance decline illustrates domain shift.

3.9 IDH Mutation in Glioma

Histology-based AI has shown capacity to infer IDH status in glioma. Jiang et al. reported AUROC values around 0.81–0.84 for IDH status,24 and subsequent multimodal approaches combining histology and clinical information have reported further improvement.25

3.10 Comparative Diagnostic Performance

Table 3. Comparative diagnostic accuracy across major molecular targets.

Tumor/target	Evidence level	Approximate diagnostic performance	Interpretation
CRC MSI/dMMR	Multiple meta-analyses, multicenter external validation	Sensitivity ~0.90; specificity ~0.86; SROC-AUC ~0.94	Strongest current evidence
Lung ALK	Meta-analysis	Sensitivity ~0.80; specificity ~0.85	Promising but limited studies
Lung EGFR	Multiple reviews/meta-analyses	Sensitivity ~0.66–0.80; specificity ~0.68–0.77; AUC ~0.76–0.78	Moderate
Lung TP53	Meta-analysis	Sensitivity ~0.70; specificity ~0.70	Moderate
CRC BRAF	Multicenter studies	AUROC up to ~0.88	Promising
CRC KRAS	Multicenter studies	AUROC ~0.80 in newer models	Moderate
Thyroid BRAF	External single-study validation	AUROC 0.98	Excellent single-study performance
Glioma IDH	Several studies	AUROC ~0.81–0.84	Promising
Breast BRCA	External validation	AUROC up to ~0.83	Moderate
Bladder FGFR3	External validation	AUROC ~0.63	Limited
Lung KRAS	Meta-analysis	Sensitivity 0.63; specificity 0.62	Limited
Lung BRAF	Meta-analysis	Sensitivity 0.51; specificity 0.48	Insufficient for screening

No global pooled effect was calculated because the target conditions were biologically and clinically heterogeneous.

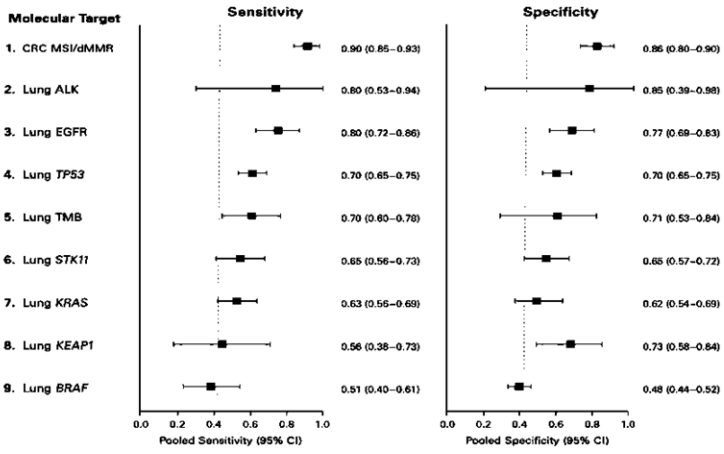
3.11 Internal Versus External Validation

A recurrent finding was deterioration in performance after external validation. External cohorts differ in patient ancestry, tumor prevalence, tissue processing, scanner manufacturer, staining characteristics, section thickness, specimen type, and clinical workflow. Modern transformer and self-supervised approaches appear to reduce, but not eliminate, this generalization gap.

3.12 Risk of Bias

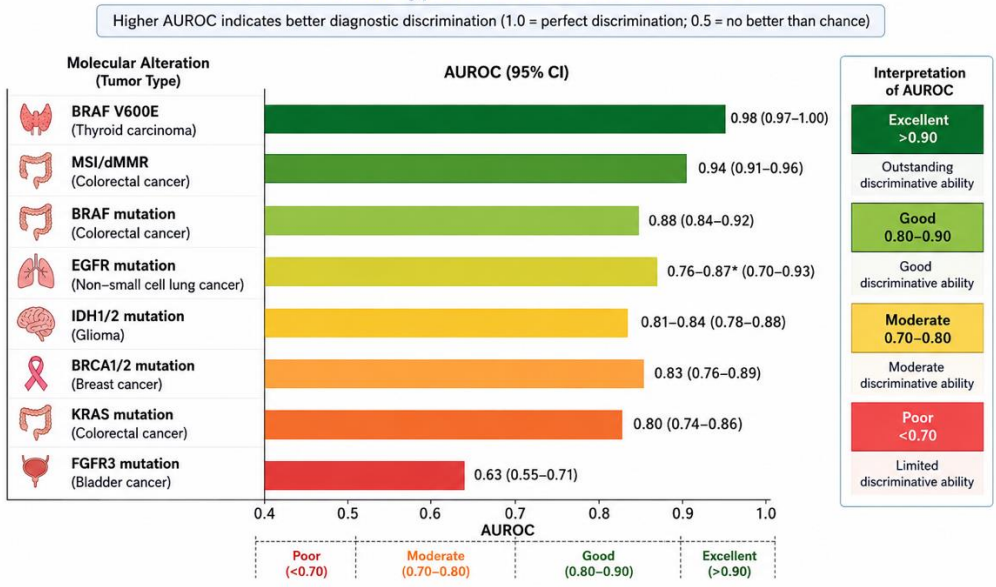
The major methodological concerns were retrospective designs, heavy dependence on TCGA, inadequate independent external validation, patient-level data leakage, post hoc threshold selection, small molecular-positive classes, class imbalance, incomplete reporting of excluded slides, scanner and staining domain shift, differing reference standards, selection of best-performing models from multiple experiments, and limited prospective evaluation.

Figure 2. Forest Plot of Diagnostic Accuracy for Major Molecular Alterations



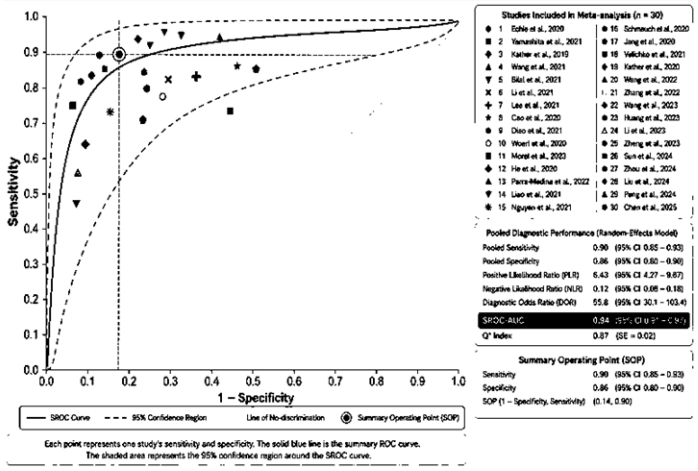
CRC, colorectal cancer; MSI/dMMR, microsatellite instability/mismatch-repair deficiency; TMB, tumor mutational burden.

Figure 3. Comparative AUROC of AI-Based Molecular Prediction Across Tumor Types and Molecular Alterations



* For EGFR in NSCLC, pooled AUROC across studies ranged from 0.76 to 0.87.
CI: Confidence Interval; AUROC: Area Under the Receiver Operating Characteristic Curve; MSI: Microsatellite Instability; dMMR: Mismatch Repair Deficiency; NSCLC: Non–Small Cell Lung Cancer.

Figure 4. Summary Receiver Operating Characteristic (SROC) Curve
AI Based Prediction of MSI/dMMR in Colorectal Cancer from Histopathological Images



DISCUSSION

4.1 Principal Findings

Routine H&E slides contain sufficient morphological information for AI systems to infer a broad range of molecular alterations, but diagnostic performance is strongly dependent on the biological target and tumor context. MSI/dMMR in colorectal cancer currently represents the most mature application. Mutation prediction in lung cancer is more heterogeneous, with ALK and EGFR generally outperforming KRAS, STK11, KEAP1, and BRAF.

4.2 Why Some Molecular Alterations Are More Predictable

AI prediction depends on the existence of a stable morphological phenotype associated with the molecular alteration. MSI-high colorectal tumors often show increased lymphocytic infiltration, poor differentiation, mucinous or medullary morphology, while BRAF V600E-mutated thyroid carcinoma has recognizable papillary morphologic correlates. Alterations such as KRAS or some lung BRAF mutations may produce weaker or more heterogeneous morphologic phenotypes.

4.3 Evolution of AI Architecture
The field has moved from patch-level CNN classification toward multiple-instance learning, attention mechanisms, self-supervised representations, and transformers. These approaches better exploit slide-level context and reduce dependence on detailed tumor annotations.

4.4 Role of External Validation
External validation is critical because histopathology AI is vulnerable to domain shift. Models can learn scanner characteristics, staining protocols, institutional workflow, tissue-processing artifacts, or population-specific correlations rather than molecular biology. Large multicenter training datasets, stain augmentation, domain adaptation, self-supervised learning, and locked external thresholds are therefore essential.

4.5 Clinical Role: Replacement or Prescreening?
Current evidence favors prescreening rather than replacement of molecular diagnostics. High-sensitivity AI systems could rapidly identify patients with a low probability of a target alteration and enrich confirmatory molecular testing among the remaining patients. False-negative predictions, however, could deny patients effective targeted therapy or immunotherapy, so deployment strategies must be evaluated according to clinical consequences rather than AUROC alone.

4.6 Importance of Prevalence
Sensitivity and specificity are relatively prevalence-independent, whereas predictive values are not. Rare alterations such as ALK or ROS1 present a particular challenge because even high sensitivity and specificity may yield modest positive predictive value at low prevalence. Decision-curve analysis, workload reduction, number-needed-to-test, and false-negative counts are therefore clinically important.

4.7 Explainability and Morphological Discovery
Image-based molecular prediction can also reveal previously unrecognized genotype-phenotype relationships. Attention maps and interpretable image features may identify regions associated with molecular status, although heatmaps indicate model attention rather than causal biological mechanisms.

4.8 Strengths
The field is increasingly supported by large cohorts, external validation, multiple tumor types, modern computational architectures, and large public molecular-pathology datasets. Contemporary studies increasingly use multicenter data rather than single-center proof-of-concept cohorts.

4.9 Limitations
Most studies remain retrospective; TCGA is repeatedly reused; reference standards differ; AUROC is frequently emphasized without prespecified operating thresholds; multiple architectures may be tested with selective reporting of the best model; molecular-positive classes are often small; preprocessing and excluded-slide reporting may be incomplete; and global pooling across cancers or molecular targets is clinically inappropriate.

4.10 Future Research
Future studies should prioritize prospective multicenter validation, locked thresholds, population diversity, clinically realistic prevalence, biopsy specimens, and head-to-head comparison with routine molecular workflows. Cost-effectiveness, calibration, uncertainty estimation, abstention mechanisms, and human-AI workflows should be evaluated prospectively.

CONCLUSION
Artificial intelligence can predict a range of molecular alterations directly from routine histopathological images, confirming that genomic abnormalities generate quantifiable morphological phenotypes. Diagnostic performance is not uniform across biomarkers. MSI/dMMR prediction in colorectal cancer has the strongest current evidence, with pooled sensitivity and specificity approaching 0.90 and SROC-AUC around 0.94. ALK and EGFR prediction in lung cancer is promising, whereas TP53 is moderately predictable and KRAS, STK11, KEAP1, and lung BRAF remain less reliable. Highly encouraging single-study results have also been reported for thyroid BRAF, glioma IDH, breast BRCA, and bladder FGFR3.

At present, image-derived molecular prediction is best positioned as an AI-assisted prescreening or prioritization tool rather than a replacement for molecular testing. Reliable clinical deployment will require external multicenter validation, locked thresholds, representative populations, standardized pathology workflows, robust reference standards, and prospective demonstration that AI improves patient care.

REFERENCES

1. Coudray N, Ocampo PS, Sakellaropoulos T, et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. Nat Med. 2018;24:1559-1567.
2. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71.
3. McInnes MDF, Moher D, Thombs BD, et al. Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: the PRISMA-DTA Statement. JAMA. 2018;319:388-396.
4. Cifci D, Foersch S, Kather JN. Artificial intelligence to identify genetic alterations in conventional histopathology. J Pathol. 2022;257:430-444.
5. Echle A, Rindtorff NT, Brinker TJ, et al. Deep learning in cancer pathology: a new generation of clinical biomarkers. Br J Cancer. 2021;124:686-696.
6. Kather JN, Pearson AT, Halama N, et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. Nat Med. 2019;25:1054-1056.
7. Fu Y, Jung AW, Torne RV, et al. Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. Nat Cancer. 2020;1:800-810.
8. Kather JN, Heij LR, Grabsch HI, et al. Pan-cancer image-based detection of clinically actionable genetic alterations. Nat Cancer. 2020;1:789-799.
9. Echle A, Grabsch HI, Quirke P, et al. Clinical-grade detection of microsatellite instability in colorectal tumors by deep learning. Gastroenterology. 2020;159:1406-1416.e11.
10. Yamashita R, Long J, Longacre T, et al. Deep learning model for the prediction of microsatellite instability in colorectal cancer: a diagnostic study. Lancet Oncol. 2021;22:132-141.
11. Bilal M, Raza SEA, Azam A, et al. Development and validation of a weakly supervised deep learning framework to predict the status of molecular pathways and key mutations in colorectal cancer from routine histology images. Lancet Digit Health. 2021;3:e763-e772.
12. Wagner SJ, Reisenbüchler D, West NP, et al. Transformer-based biomarker prediction from colorectal cancer histology: a large-scale multicentric study. Cancer Cell. 2023;41:1650-1661.e4.
13. Wang Z, Chang F, Wu H, Yuan X, Chen Y, Han J. Accuracy of deep learning-aided detection of microsatellite instability in colorectal cancer: a systematic review and meta-analysis. Clin Colorectal Cancer. 2026;25:1-20.
14. Li H, Qin J, Li Z, et al. Systematic review and meta-analysis of deep learning for MSI-H in colorectal cancer whole slide images. NPJ Digit Med. 2025;8:456.
15. Parra-Medina R, Guerron-Gomez G, Mendivelso-González D, et al. Deep learning in histopathology images for prediction of oncogenic driver molecular alterations in lung cancer: a systematic review and meta-analysis. Transl Lung Cancer Res. 2025;14:1756-1769.
16. Nguyen MH, Le MHN, Bui AT, et al. Artificial intelligence in predicting EGFR mutations from whole slide images in lung cancer: a systematic review and meta-analysis. Lung Cancer. 2025;204:108577.
17. Zhao Y, Xiong S, Ren Q, et al. Deep learning using histological images for gene mutation prediction in lung cancer: a multicentre retrospective study. Lancet Oncol. 2025;26:136-146.
18. Morel LO, Derangère V, Arnould L, et al. Preliminary evaluation of deep learning for first-line diagnostic prediction of tumor mutational status. Sci Rep. 2023;13:6927.
19. Zhang W, Wang W, Xu Y, et al. Prediction of epidermal growth factor receptor mutation subtypes in non-small cell lung cancer from hematoxylin and eosin-stained slides using deep learning. Lab Invest. 2024;104:102094.
20. Terada Y, Takahashi T, Hayakawa T, et al. Artificial intelligence-powered prediction of ALK gene rearrangement in patients with non-small-cell lung cancer. JCO Clin Cancer Inform. 2022;6:e2200070.
21. Sadhwani A, Chang HW, Behrooz A, et al. Comparative analysis of machine learning approaches to classify tumor mutation burden in lung adenocarcinoma using histopathology images. Sci Rep. 2021;11:16605.
22. Loeffler CML, Ortiz Bruechle N, Jung M, et al. Artificial intelligence-based detection of FGFR3 mutational status directly from routine histology in bladder cancer. Eur Urol Focus. 2022;8:472-479.
23. Velmahos CS, Badgeley M, Lo YC. Using deep learning to identify bladder cancers with FGFR-activating mutations from histology images. Cancer Med. 2021;10:4805-4813.
24. Jiang S, Zanazzi GJ, Hassanpour S. Predicting prognosis and IDH mutation status for patients with lower-grade gliomas using whole slide images. Sci Rep. 2021;11:16849.
25. Nakagaki R, Debsarkar SS, Kawanaka H, Aronow BJ, Prasath VBS. Deep learning-based IDH1 gene mutation prediction using histopathological imaging and clinical data. Comput Biol Med. 2024;179:108902.
26. Anand D, Yashashwi K, Kumar N, et al. Weakly supervised learning on unannotated H&E-stained slides predicts BRAF mutation in thyroid cancer with high accuracy. J Pathol. 2021;255:232-242.
27. Wang X, Zou C, Zhang Y, et al. Prediction of BRCA gene mutation in breast cancer based on deep learning and histopathology images. Front Genet. 2021;12:661109.

28. Qu H, Zhou M, Yan Z, et al. Genetic mutation and biological pathway prediction based on whole slide images in breast carcinoma using deep learning. NPJ Precis Oncol. 2021;5:87.
29. Hong R, Liu W, DeLair D, et al. Predicting endometrial cancer subtypes and molecular features from histopathology images using multi-resolution deep learning models. Cell Rep Med. 2021;2:100400.
30. Chen M, Zhang B, Topatana W, et al. Classification and mutation prediction based on histopathology H&E images in liver cancer using deep learning. NPJ Precis Oncol. 2020;4:14.
31. Liao H, Long Y, Han R, et al. Deep learning-based classification and mutation prediction from histopathological images of hepatocellular carcinoma. Clin Transl Med. 2020;10:e102.
32. Diao JA, Wang JK, Chui WF, et al. Human-interpretable image features derived from densely mapped cancer pathology slides predict diverse molecular phenotypes. Nat Commun. 2021;12:1613.
33. Schmauch B, Romagnoni A, Pronier E, et al. A deep learning model to predict RNA-Seq expression of tumours from whole slide images. Nat Commun. 2020;11:3877.
34. He B, Bergenstråhle L, Stenbeck L, et al. Integrating spatial gene expression and breast tumour morphology via deep learning. Nat Biomed Eng. 2020;4:827-834.
35. Jang HJ, Lee A, Kang J, et al. Prediction of clinically actionable genetic alterations from colorectal cancer histopathology images using deep learning. World J Gastroenterol. 2020;26:6207-6223.
36. Lee SH, Song IH, Jang HJ. Feasibility of deep learning-based fully automated classification of microsatellite instability in tissue slides of colorectal cancer. Int J Cancer. 2021;149:728-740.
37. Cao R, Yang F, Ma SC, et al. Development and interpretation of a pathomics-based model for the prediction of microsatellite instability in colorectal cancer. Theranostics. 2020;10:11080-11091.
38. Woerl AC, Eckstein M, Geiger J, et al. Deep learning predicts molecular subtype of muscle-invasive bladder cancer from conventional histopathological slides. Eur Urol. 2020;78:256-264.
39. Noorbakhsh J, Farahmand S, Foroughi Pour A, et al. Deep learning-based cross-classifications reveal conserved spatial behaviors within tumor histological images. Nat Commun. 2020;11:6367.
40. McGenity C, Clarke EL, Jennings C, et al. Artificial intelligence in digital pathology: a systematic review and meta-analysis of diagnostic test accuracy. NPJ Digit Med. 2024.