



Journal of Frontiers in Multidisciplinary Research

Development of an Integrated Diagnostic Intelligence Architecture for Precision Oncology Combining ctDNA Genomics Biomarker Analysis Assay Reliability and Clinical Laboratory Performance

Rukayat Olayemi Oni ^{1*}, Yewande Iyimide Adeyeye ²

¹ Department of Biology, Georgia State University, Atlanta, Georgia

² Department of Day Case Surgery, Dumfries and Galloway Royal Infirmary, Dumfries, United Kingdom

* Corresponding Author: **Rukayat Olayemi Oni**

Article Info

E-ISSN: 3050-9726

P-ISSN: 3050-9718

Volume: 03

Issue: 02

July – December 2022

Received: 14-10-2022

Accepted: 16-11-2022

Published: 18-12-2022

Page No: 266-285

Abstract

Precision oncology increasingly depends on circulating tumor DNA (ctDNA), genomic biomarkers, molecular assay performance, and laboratory-quality indicators; however, these information streams are commonly interpreted through separate analytical pipelines, limiting diagnostic consistency and clinical decision reliability. This study develops an Integrated Diagnostic Intelligence Architecture for Precision Oncology and introduces a novel *Clinical Genomic Assay Reliability and Diagnostic Intelligence Network (CGARDI-Net)* for unified interpretation of ctDNA genomic signals, biomarker profiles, assay-quality characteristics, and clinical laboratory performance indicators. The proposed architecture integrates variant allele frequency, somatic mutation signatures, copy-number alterations, methylation-associated biomarkers, fragmentomic features, tumor-informed genomic indicators, limit of detection, analytical sensitivity, analytical specificity, sequencing depth, background error rate, sample-quality indices, turnaround time, quality-control compliance, and laboratory reproducibility within a reliability-aware multimodal learning framework. CGARDI-Net combines attention-based multimodal feature fusion, assay-reliability weighting, uncertainty-aware probabilistic calibration, and clinical decision scoring to reduce the influence of technically unreliable molecular measurements on diagnostic predictions. Its performance is comparatively evaluated against logistic regression, support vector machines, random forest, XGBoost, multilayer perceptron, convolutional neural networks, long short-term memory networks, standard Transformer architectures, and conventional threshold-based ctDNA interpretation. Model assessment is structured around sensitivity, specificity, precision, F1-score, area under the receiver operating characteristic curve, area under the precision-recall curve, Brier score, calibration error, false-negative rate, and decision-curve net benefit. Comparative graphs include ROC curves, precision-recall curves, calibration plots, confusion matrices, reliability-performance plots, biomarker importance charts, assay-error sensitivity curves, and three-dimensional relationships among ctDNA signal strength, assay reliability, and diagnostic confidence. Ablation analysis further quantifies the contribution of genomic biomarkers, assay reliability parameters, and laboratory-quality variables to overall diagnostic performance. The proposed architecture is designed to outperform isolated genomic and conventional machine-learning approaches by integrating biological evidence with analytical validity and laboratory operational reliability rather than treating every molecular measurement as equally trustworthy. The resulting framework establishes a technically interpretable pathway toward laboratory-aware precision oncology decision support capable of improving diagnostic robustness, reducing uncertainty associated with low-abundance ctDNA signals, and supporting more reliable translation of molecular testing into individualized cancer surveillance and treatment decisions.

DOI: <https://doi.org/10.54660/JFMR.2022.3.2.266-285>

Keywords: Precision Oncology, Circulating Tumor DNA, Genomic Biomarkers, Assay Reliability, Clinical Laboratory Performance

1. Introduction

1.1. Background to Precision Oncology and Integrated Diagnostic Intelligence

Precision oncology has evolved from histology-dominated treatment selection toward molecularly stratified care in which genomic alterations, pathway dependencies, tumor phenotype, and treatment context jointly inform diagnosis and therapy. This transition reflects recognition that cancers sharing an anatomic site can differ in driver mutations, clonal composition, therapeutic

vulnerabilities, and resistance. Friedman *et al.* (2015) argued that genomics alone cannot fully capture clinically relevant tumor behavior because genotype-to-phenotype relationships are incomplete and dynamic, creating a need for complementary diagnostic evidence. Accordingly, precision oncology increasingly requires architectures combining heterogeneous molecular measurements with their analytical trustworthiness. Here, integrated diagnostic intelligence denotes the fusion of ctDNA-derived genomic signals, biomarker evidence, assay-performance characteristics, and laboratory-quality variables into a patient-level diagnostic representation. The proposed framework evaluates its variant allele fraction, molecular context, detection confidence, assay limit of detection, sequencing depth, background error, sample quality, and laboratory performance before assigning significance.

This formulation addresses the evidentiary problem identified by Moscow *et al.* (2018): precise molecular therapies do not automatically imply equally precise evidence for selecting them. Diagnostic intelligence must therefore distinguish biological actionability from analytical validity and clinical reliability. A biomarker may be mechanistically relevant yet unsuitable for decision support when detected near an unstable analytical threshold, while a technically robust variant can remain clinically uninformative if its cancer relevance is weak. The proposed Clinical Genomic Assay Reliability and Diagnostic Intelligence Network, CGARDI-Net, operationalizes this distinction through reliability-weighted multimodal feature fusion and uncertainty-aware probability estimation. Its architecture is designed to learn interactions among ctDNA burden, mutation signatures, methylation or fragmentomic evidence, assay sensitivity, false-positive risk, quality-control performance, and laboratory reproducibility. The output is a calibrated diagnostic confidence conditioned on both biological evidence and measurement reliability. This design establishes the conceptual foundation for later comparisons with logistic regression, support vector machines, random forest, XGBoost, neural networks, and Transformer-based models, particularly under low-abundance ctDNA and degraded assay-performance conditions. This is central to reproducible translation across laboratories, platforms, and clinical settings.

1.2. Circulating Tumor DNA in Cancer Detection and Disease Monitoring

Circulating tumor DNA is the tumor-derived fraction of cell-free DNA released into blood through apoptosis, necrosis, and other cellular processes, providing a minimally invasive view of tumor burden and evolution. Unlike a single tissue biopsy, plasma ctDNA can sample DNA shed from anatomically distinct tumor sites and can be collected serially, enabling genotyping, treatment monitoring, residual disease assessment, relapse detection, and resistance analysis. Wan *et al.* (2017) emphasized that ctDNA has value as both a quantitative marker of disease burden and a genomic substrate for detecting tumor-associated alterations. Its interpretation, however, is technically difficult because ctDNA may represent only a small fraction of total circulating cell-free DNA, particularly during early-stage disease or effective therapy. Low variant allele fractions increase stochastic sampling error and place substantial demands on sequencing depth, barcoding, error suppression, and preanalytical control. Consequently, ctDNA presence or

absence cannot be interpreted independently of assay capability.

Longitudinal ctDNA analysis is especially important because cancer genomes evolve under treatment pressure. Siravegna *et al.* (2017) showed that liquid biopsy can support repeated molecular surveillance where serial tissue sampling is impractical and can reveal emerging resistant clones before they become dominant clinically. Within the proposed architecture, ctDNA is therefore represented as a dynamic signal rather than a static positive-versus-negative variable. CGARDI-Net incorporates variant allele frequency, temporal change in ctDNA concentration, mutation class, copy-number information, methylation features, fragmentomic patterns, and tumor-informed genomic indicators together with assay confidence measures. For example, a declining pathogenic variant detected consistently above the validated analytical range should contribute differently to diagnostic confidence than an isolated low-frequency call near the assay limit of detection. The model also permits conflicting longitudinal evidence to be expressed as calibrated uncertainty rather than forced classification. This treatment of ctDNA directly supports the planned performance analyses, including sensitivity across low variant allele fractions, false-negative behavior near detection limits, robustness to sequencing noise, and three-dimensional evaluation of ctDNA signal strength, assay reliability, and diagnostic confidence.

1.3. Genomic Biomarkers for Precision Oncology Decision Support

Genomic biomarkers provide the molecular basis for classifying cancers beyond organ of origin and can support detection, prognosis, treatment selection, and longitudinal surveillance. Relevant biomarker classes extend beyond single-nucleotide variants to include insertions and deletions, copy-number alterations, rearrangements, methylation patterns, fragmentomic signatures, and complementary circulating protein signals. Cohen *et al.* (2018) demonstrated the diagnostic value of combining somatic mutations in cell-free DNA with circulating protein biomarkers, showing that multi-analyte integration can improve cancer detection while maintaining high specificity. This principle is central to the present study because no single biomarker class is expected to capture the full biological heterogeneity of cancer. A mutation may indicate pathway activation, whereas methylation can encode tissue-of-origin information, copy-number changes can reflect genomic instability, and fragmentomic features can capture nucleosomal organization and tumor-associated cell-death patterns. Diagnostic decision support should therefore integrate these features while preserving their distinct biological meanings and measurement uncertainties.

Fragmentomics further expands the information available from cell-free DNA. Cristiano *et al.* (2019) showed that genome-wide fragmentation patterns discriminate patients with cancer from healthy individuals and can provide information about tissue of origin, demonstrating that clinically informative signal exists beyond sequence variants alone. CGARDI-Net is designed to exploit this multidimensional biomarker space through modality-specific feature encoders followed by attention-based fusion. Mutation features can be weighted by pathogenicity and allele fraction; methylation features by cancer-associated differential methylation; copy-number features by amplitude

and genomic span; and fragmentomic features by size distribution and regional fragmentation patterns. The architecture then integrates these molecular representations with assay reliability and laboratory-performance variables so that biologically informative features are not treated as equally trustworthy under all analytical conditions. This design enables subsequent feature-importance analysis, ablation testing, and comparison with conventional machine-learning methods. It also provides a mechanistic basis for examining whether diagnostic gains arise from true multimodal complementarity rather than model complexity alone, particularly when ctDNA abundance is low or individual biomarker classes yield discordant evidence across heterogeneous clinical specimens.

1.4. Assay Reliability and Analytical Performance in Molecular Diagnostics

Analytical reliability determines whether a molecular signal is trustworthy for clinical interpretation. For next-generation sequencing oncology assays, reliability depends on specimen quality, nucleic-acid input, library preparation, sequencing depth, coverage uniformity, error suppression, variant-calling thresholds, reference materials, reproducibility, and validated performance across relevant variant classes. Jennings *et al.* (2017) emphasized that oncology-panel validation must establish performance characteristics such as accuracy, precision, analytical sensitivity, analytical specificity, reportable range, and detection capability using reference materials and variants. These requirements are especially important for ctDNA because relevant alterations can occur at low variant allele fractions. A nominally positive genomic call is therefore not equivalent to a reliable positive result unless its observed signal lies within an analytically characterized region of assay performance. In the proposed architecture, assay reliability is modeled explicitly rather than treated as hidden metadata. Variables such as limit of detection, molecular depth, duplicate suppression, background error rate, input quantity, and quality-control status become quantitative inputs to the diagnostic model.

Cross-platform evidence demonstrates why weighting is necessary. Deveson *et al.* (2021) evaluated five leading ctDNA sequencing assays and found generally strong performance above 0.5% variant allele frequency but substantial deterioration and inter-assay variability below that level, particularly when DNA input was limited. This behavior implies that a diagnostic model trained only on genomic features can learn from signals whose reliability changes systematically across concentration ranges and platforms. CGARDI-Net addresses this problem through an assay-reliability module that estimates the confidence associated with each molecular feature before multimodal fusion. A low-frequency variant supported by adequate molecule counts and favorable error characteristics can retain diagnostic influence, whereas an equally low-frequency call generated under poor input or unstable background noise can be down-weighted. Planned sensitivity curves, limit-of-detection analyses, calibration plots, and noise-stress experiments directly evaluate this mechanism. The intended result is a model whose performance remains interpretable when analytical conditions change, thereby separating biological absence from technical non-detection and reducing false confidence in marginal molecular measurements.

1.5. Clinical Laboratory Performance and Diagnostic Quality Assurance

Clinical laboratory performance is a critical determinant of whether sophisticated genomic models produce dependable patient-level outputs. Molecular diagnostics operates as a chain of preanalytical, analytical, bioinformatic, interpretive, and post-analytical processes, and failure at any stage can alter the apparent biomarker profile. Roy *et al.* (2018) emphasized that validated bioinformatics pipelines are integral components of next-generation sequencing and that poorly developed or inadequately monitored pipelines can generate inaccurate results with consequences for patient care. Laboratory performance must therefore include more than instrument accuracy. Relevant indicators include specimen rejection rate, sample identity controls, extraction success, nucleic-acid yield, sequencing quality, coverage completeness, control-material performance, repeat-testing frequency, turnaround time, pipeline version control, proficiency testing, and reproducibility across runs and operators. In the proposed framework, these variables are aggregated into a laboratory performance representation that can modulate the confidence assigned to ctDNA and genomic biomarker evidence.

Interpretation and reporting quality are equally important. Li *et al.* (2017) proposed standardized classification and reporting of somatic sequence variants and stressed that laboratory reports should communicate technical and interpretive factors that can affect clinical meaning. This principle motivates the distinction in CGARDI-Net between raw predictive probability and calibrated diagnostic confidence. A model may estimate a high probability of malignancy from a molecular pattern, but the confidence attached to that estimate should decrease when sequencing quality, pipeline stability, specimen adequacy, or reporting completeness falls outside validated expectations. The architecture therefore treats laboratory performance as an active diagnostic variable rather than an administrative quality metric. This permits analysis of how laboratory reliability influences prediction error, calibration, false-negative rate, and net clinical benefit. It also supports comparison across sites by separating biological variation from process-related variation. Such integration is essential for the study's proposed laboratory-performance plots and deployment analysis, because a model intended for precision oncology should maintain reproducibility not only across patients but across the operational environments in which molecular data are generated, interpreted, and reported under routine clinical conditions.

1.6. Limitations of Existing ctDNA and Machine-Learning Diagnostic Approaches

Existing ctDNA diagnostic strategies frequently rely on fixed thresholds, limited biomarker panels, or predictive models that do not explicitly account for assay uncertainty. This creates difficulty when variant allele fractions approach technical detection limits or when observed mutations may arise from non-tumor sources. Stetson *et al.* (2019) compared four plasma next-generation sequencing assays and reported substantial inter-assay variation, with sensitivity ranging from 38% to 89% and positive predictive value from 36% to 80%; much of the discordance occurred below 1% allele fraction and was attributed to technical factors. Such findings

show why binary mutation calls can be unstable inputs for downstream classification. A threshold-based system may classify a weak signal as positive without adequately representing background error, while a machine-learning model may amplify the same problem if technical quality variables are absent from training. Conventional models can therefore appear accurate in aggregate while remaining poorly calibrated in low-signal or cross-platform settings. Machine learning can integrate complex genomic information, but its performance depends strongly on feature design, cohort composition, validation strategy, and control of biological confounders. Chabon *et al.* (2020) developed Lung-CLiP by integrating cfDNA molecular features and showed that clonal hematopoiesis is a major source of non-tumor mutations in plasma, illustrating the importance of distinguishing tumor-derived evidence from competing signals. The limitation relevant to the present study is not that machine learning is inherently unsuitable, but that most diagnostic algorithms optimize classification without jointly modeling the reliability of the measurements being classified. CGARDI-Net is designed to close this methodological gap by combining multimodal genomic representation with assay-reliability weighting, laboratory-performance information, and uncertainty calibration. Its evaluation against logistic regression, support vector machines, random forest, XGBoost, multilayer perceptron, CNN, LSTM, Transformer, and threshold-based classification will therefore test more than discrimination. AUROC, AUPRC, sensitivity, specificity, Brier score, calibration error, false-negative rate, and decision-curve net benefit will assess whether superior classification is accompanied by trustworthy probability estimation and resilience to technical degradation across laboratories.

1.7. Problem Statement, Research Gap and Motivation for an Integrated Diagnostic Architecture

The central problem addressed by this study is the fragmentation of precision-oncology diagnostics into separate biological, analytical, and operational layers. ctDNA assays can generate clinically relevant genomic information, yet interpretation depends on assay sensitivity, variant allele fraction, sample quality, sequencing error, and the clinical context in which the result is interpreted. Merker *et al.* (2018) highlighted important evidence gaps in ctDNA testing, including discordance between plasma and tissue genotyping and limited evidence, at that time, for several applications such as early-stage disease, treatment monitoring, and residual disease detection. The architectural problem remains: diagnostic algorithms commonly receive molecular features after laboratory processing has already converted variable-quality measurements into apparently comparable inputs. This separation can conceal uncertainty and make a technically fragile result appear biologically decisive. It also complicates cross-laboratory generalization because identical variant calls can differ in reliability across workflows, platforms, or quality-control conditions.

Clinical implementation studies further indicate that adoption of liquid biopsy requires attention to technological, methodological, and workflow constraints rather than predictive accuracy alone. Ignatiadis *et al.* (2021) identified continuing challenges in bringing liquid biopsy into routine clinical practice, including assay performance, clinical validation, trial design, logistics, and integration into decision pathways. The research gap is therefore an absence of a

unified diagnostic intelligence architecture that learns cancer-associated molecular patterns while simultaneously estimating how strongly those patterns should be trusted given assay and laboratory conditions. CGARDI-Net is motivated by this gap. It integrates ctDNA genomics, methylation, copy-number and fragmentomic evidence with analytical sensitivity, detection limits, sequencing quality, background error, reproducibility, and laboratory-performance indicators. Reliability-aware attention and uncertainty calibration are expected to reduce overconfident predictions when technical evidence is weak and preserve informative low-abundance signals when analytical support is strong. The subsequent results section is structured to test this claim through benchmark comparisons, ablation analysis, low-allele-fraction stress testing, calibration assessment, laboratory-performance analysis, and decision curves, thereby linking the stated problem directly to measurable diagnostic outcomes in routine clinical use.

1.8. Research Objectives and Research Questions

Research Objectives

1. To develop an integrated diagnostic intelligence architecture combining ctDNA genomics, molecular biomarkers, assay reliability, and clinical laboratory performance for precision oncology.
2. To develop CGARDI-Net for reliability-weighted multimodal fusion of genomic, methylation, copy-number, fragmentomic, analytical, and laboratory-quality features.
3. To formulate an assay-reliability mechanism that adjusts diagnostic feature weights according to detection limits, sequencing quality, background error, and analytical uncertainty.
4. To evaluate CGARDI-Net against threshold-based classification, logistic regression, support vector machine, random forest, XGBoost, multilayer perceptron, CNN, LSTM, and Transformer models.
5. To quantify the contributions of ctDNA genomics, biomarker modalities, assay reliability, and laboratory-performance variables using ablation, sensitivity, calibration, and robustness analyses.
6. To determine the clinical diagnostic utility of CGARDI-Net using discrimination, calibration, uncertainty, false-negative performance, and decision-curve net-benefit measures.

Research Questions

1. How can ctDNA genomic biomarkers, assay reliability, and clinical laboratory performance be integrated into a unified precision-oncology diagnostic intelligence architecture?
2. Does CGARDI-Net improve diagnostic discrimination compared with conventional threshold-based, machine-learning, deep-learning, and Transformer-based diagnostic models?
3. How does reliability-weighted multimodal fusion affect cancer detection when ctDNA signals occur near the validated analytical limit of detection?
4. To what extent do assay sensitivity, sequencing depth, background error, specimen quality, and laboratory performance influence diagnostic confidence and calibration?
5. Which genomic, biomarker, analytical, and laboratory components contribute most strongly to CGARDI-Net

performance under ablation and sensitivity analysis?

6. Does incorporating analytical reliability and laboratory-quality information improve diagnostic robustness, uncertainty estimation, and potential clinical utility across heterogeneous testing conditions?

1.9. Contributions and Significance of the Study

This study contributes an integrated conceptual and computational framework that treats precision-oncology diagnosis as the joint interpretation of biological evidence and the reliability of the processes generating that evidence. Its principal methodological contribution is CGARDI-Net, a Clinical Genomic Assay Reliability and Diagnostic Intelligence Network that combines ctDNA mutation profiles, variant allele fractions, copy-number alterations, methylation characteristics, fragmentomic signatures, assay-performance parameters, and laboratory-quality indicators through modality-specific representation, reliability-aware attention, multimodal feature fusion, and calibrated uncertainty estimation. Unlike conventional classifiers that primarily optimize discrimination from molecular inputs, the proposed framework enables the diagnostic influence of a biomarker to vary according to analytical support, allowing technically robust low-abundance signals to be retained while reducing the contribution of measurements generated under unfavorable detection conditions. The study also establishes a comparative evaluation framework incorporating threshold-based interpretation, logistic regression, SVM, random forest, XGBoost, multilayer perceptron, CNN, LSTM, and Transformer models. Its significance lies in connecting molecular diagnostic intelligence with laboratory medicine by evaluating AUROC, AUPRC, sensitivity, specificity, precision, F1-score, Brier score, calibration error, false-negative behavior, robustness, and decision-curve net benefit rather than relying solely on classification accuracy. Ablation and sensitivity analyses further provide a mechanism for identifying how ctDNA information, biomarker modalities, assay reliability, and laboratory performance independently and jointly influence predictions. Consequently, the framework is positioned to support more reproducible, interpretable, and laboratory-aware precision-oncology decision support while providing a technical basis for evaluating diagnostic performance across variable ctDNA abundance, assay conditions, and clinical laboratory environments.

1.10. Scope of the Review and Structure of the Paper

The paper focuses on the development and technical evaluation of an integrated diagnostic intelligence architecture for precision oncology using circulating tumor DNA genomics, molecular biomarker information, assay-reliability characteristics, and clinical laboratory-performance indicators as complementary sources of diagnostic evidence. The biological scope includes somatic mutations, variant allele fraction, copy-number alterations, methylation-associated signatures, fragmentomic characteristics, and related tumor-derived molecular features, while the analytical scope encompasses limit of detection, sequencing depth, nucleic-acid input, background error, analytical sensitivity, analytical specificity, reproducibility, and quality-control performance. Laboratory variables include specimen quality, extraction performance, sequencing-quality indicators, repeat-testing behavior, turnaround time, pipeline stability, proficiency performance,

and reporting reliability. The computational scope covers CGARDI-Net and its reliability-weighted multimodal fusion, attention mechanisms, uncertainty calibration, diagnostic confidence estimation, and comparison with conventional thresholds, logistic regression, SVM, random forest, XGBoost, multilayer perceptron, CNN, LSTM, and Transformer models. Section 1 establishes the scientific problem, research gap, objectives, questions, and significance; Section 2 reviews ctDNA biology, genomic biomarkers, analytical reliability, laboratory-quality indicators, and existing computational diagnostic models; Section 3 formulates the integrated system architecture, mathematical representations, CGARDI-Net algorithm, training procedure, validation strategy, and evaluation metrics; Section 4 presents comparative diagnostic results using numerical tables, ROC and precision-recall curves, calibration plots, confusion matrices, feature-importance analysis, ablation studies, sensitivity curves, three-dimensional reliability surfaces, uncertainty analysis, and decision curves; and Section 5 presents conclusions, technical contributions, implementation recommendations, limitations, and directions for prospective multicenter validation.

2. Literature Review

2.1. Biological and Molecular Characteristics of Circulating Tumor DNA

Circulating tumor DNA represents the tumor-derived component of extracellular cell-free DNA released into the circulation through apoptosis, necrosis, active cellular processes, and turnover of malignant tissue. Its concentration and molecular composition are therefore influenced by tumor burden, vascularization, anatomical site, cell-death kinetics, clearance, treatment exposure, and clonal architecture as represented in figure 1. Bettgowda *et al.* (2014) demonstrated marked variation in ctDNA detectability across cancer types and disease stages, establishing that biological tumor characteristics strongly influence the amount of tumor-derived DNA measurable in plasma. ctDNA molecules carry somatic substitutions, insertions and deletions, structural alterations, copy-number abnormalities, methylation signatures, and fragmentation patterns originating from malignant clones. Their relatively short circulating half-life makes them particularly suitable for dynamic disease assessment because molecular changes can reflect tumor evolution more rapidly than many conventional markers. Heitzer *et al.* (2019) further emphasized that liquid biopsy provides access to spatial and temporal heterogeneity that may be incompletely represented by a single tissue specimen, although ctDNA abundance, clonal hematopoiesis, and nonmalignant cfDNA remain important sources of interpretive complexity.

These characteristics directly inform the proposed CGARDI-Net architecture. Instead of representing ctDNA as a binary detected/not-detected variable, the model treats molecular evidence as a multidimensional profile comprising allele fraction, genomic locus, molecular class, temporal trajectory, fragment characteristics, sequencing support, and technical confidence. The interoperability principle described by Nwokocha, Peter-Anyebe, and Ijiga (2021) is relevant at the architecture level because heterogeneous laboratory and genomic data must retain semantic consistency when merged into a single diagnostic pipeline. Their complementary work on system integration also illustrates the importance of

coordinated data-flow design when multiple information systems contribute to one analytical decision. Accordingly, the present framework links raw ctDNA characteristics with assay metadata and laboratory-performance attributes before diagnostic inference. This permits subsequent results to distinguish genuine biological low abundance from technically unreliable low abundance and provides the

foundation for the planned three-dimensional analysis relating ctDNA signal magnitude, assay reliability, and diagnostic confidence. Such representation is necessary if an integrated diagnostic system is to preserve the molecular biology of ctDNA while simultaneously accounting for the conditions under which the signal was measured.

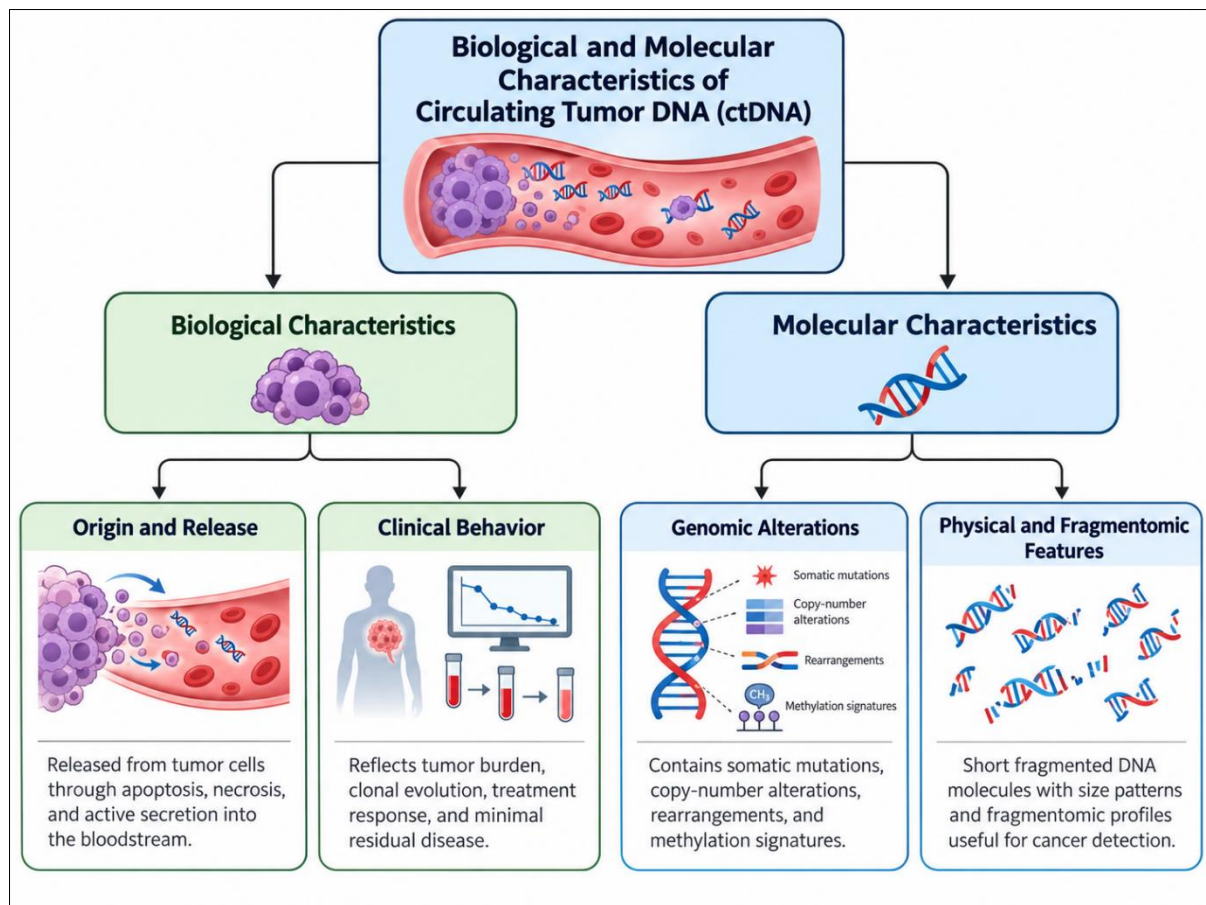


Fig 1: Biological and Molecular Characteristics of Circulating Tumor DNA Showing Its Origin, Clinical Behavior, Genomic Alterations, and Fragmentomic Features in Precision Oncology

Figure 1 presents the biological and molecular characteristics of circulating tumor DNA (ctDNA) as two interconnected dimensions that determine its value in precision oncology. The *biological characteristics* branch emphasizes that ctDNA originates predominantly from malignant cells and enters the bloodstream through apoptosis, necrosis, and other release mechanisms associated with tumor-cell turnover. Because ctDNA abundance is influenced by tumor volume, anatomical location, vascularity, treatment pressure, and clonal dynamics, its circulating concentration can provide a temporally responsive indicator of tumor burden, treatment response, disease progression, and minimal residual disease. This explains why serial ctDNA measurements can reveal molecular changes earlier than conventional imaging or static tissue sampling. The *molecular characteristics* branch shows that ctDNA contains tumor-associated genomic information, including somatic mutations, copy-number alterations, structural rearrangements, and methylation signatures that can support cancer detection, molecular classification, and therapeutic stratification. The fragmentomic component reflects the fact that tumor-derived DNA exhibits distinctive fragment lengths, end motifs, nucleosomal patterns, and regional fragmentation profiles that differ from background

cell-free DNA and can therefore provide additional diagnostic signal. Together, these branches show that ctDNA should not be treated merely as a detectable DNA fragment in plasma, but as a dynamic, multidimensional biomarker whose biological origin, molecular alterations, fragmentomic structure, and temporal behavior can be integrated within CGARDI-Net to improve cancer detection, disease monitoring, and reliability-aware precision oncology decision support.

2.2. Genomic Biomarkers Including Mutations, Methylation, Copy-Number Alterations, and Fragmentomics

The diagnostic information carried by ctDNA extends substantially beyond individual driver mutations. Somatic single-nucleotide variants and small insertions or deletions provide sequence-specific evidence of malignant clones, whereas copy-number alterations capture chromosomal gains and losses associated with genomic instability. DNA methylation supplies an additional epigenomic layer because tumorigenesis produces characteristic changes in CpG methylation that can differentiate cancer-derived DNA from background cfDNA and contribute to tissue-of-origin

classification as presented in table 1. Shen *et al.* (2018) demonstrated that plasma cfDNA methylome profiling can support sensitive tumor detection and classification, confirming that cancer-associated epigenetic signatures can provide information not contained in mutation calls alone. Fragmentomics similarly exploits differences in cfDNA size, genomic positioning, and fragmentation patterns. Mouliere *et al.* (2018) showed that enrichment and analysis of shorter plasma DNA fragments can increase detection of tumor-derived DNA, indicating that physical properties of circulating DNA represent diagnostically exploitable biomarkers in their own right.

The CGARDI-Net framework therefore models genomic biomarkers as complementary feature families rather than as interchangeable predictors. Mutation channels encode variant identity, allele fraction, predicted functional relevance, and clonality; methylation channels capture locus-specific epigenetic deviation; copy-number channels represent amplitude and genomic extent of gains or losses; and fragmentomic channels represent fragment-length distributions and related structural characteristics. The

principle of sensor fusion described by Oladoye *et al.* (2021), although developed in a different engineering domain, is relevant computationally because heterogeneous measurements become more informative when their reliability and complementary signal structure are jointly modeled. Likewise, data-driven integration frameworks such as that discussed by Kwarteng *et al.* (2021) emphasize coordinated use of heterogeneous indicators rather than isolated metric interpretation. CGARDI-Net operationalizes this concept through modality-specific encoding followed by reliability-weighted attention, allowing the model to determine whether a mutation, methylation event, copy-number abnormality, or fragmentation signature should dominate a particular diagnostic decision. This construction supports the later feature-importance, ablation, and multimodal performance analyses and permits direct investigation of whether combined molecular evidence produces better calibrated predictions than mutation-only or single-modality models.

Table 1: Summary of Major Genomic Biomarker Classes Used in ctDNA-Based Precision Oncology

Biomarker Class	Molecular Characteristics	Diagnostic/Clinical Relevance	Application in the Proposed CGARDI-Net Framework
Somatic Mutations	Single-nucleotide variants and small insertions/deletions arising in tumor DNA, often involving cancer driver genes and resistance-associated loci.	Support cancer detection, molecular subtyping, therapeutic target identification, clonal tracking, and detection of emerging resistance mechanisms.	Encoded using mutation identity, variant allele frequency, pathogenicity, clonality, and sequencing support, then reliability-weighted according to assay confidence.
DNA Methylation	Cancer-associated changes in CpG methylation patterns that reflect epigenetic dysregulation and may retain tissue-of-origin information.	Improve early cancer detection, tumor classification, and discrimination of malignant from nonmalignant cell-free DNA, particularly when mutation burden is low.	Represented as locus-specific methylation features and integrated with genomic and assay-quality variables through multimodal attention for diagnostic probability estimation.
Copy-Number Alterations	Segmental genomic gains and losses caused by chromosomal instability, amplification, or deletion of cancer-associated regions.	Provide evidence of genomic instability, oncogene amplification, tumor-suppressor loss, disease progression, and potentially aggressive tumor phenotypes.	Modeled using alteration amplitude, genomic span, affected loci, and confidence scores, contributing to the integrated patient-level diagnostic representation.
Fragmentomics	Tumor-associated differences in cfDNA fragment size, end motifs, nucleosomal positioning, genomic coverage patterns, and fragmentation profiles.	Enhance detection of low-abundance ctDNA, improve cancer classification, and provide complementary information when sequence variants are sparse or technically uncertain.	Processed as fragment-length and structural features that complement mutation, methylation, and copy-number data, improving multimodal discrimination and robustness.

2.3. Analytical Sensitivity, Specificity, Limit of Detection, and Assay Reliability

Analytical sensitivity and specificity determine whether an assay can consistently detect authentic tumor-derived alterations while excluding background noise. These parameters become critical when ctDNA concentrations are low because sequencing and polymerase errors may approach the frequency of true variants. Newman *et al.* (2016) demonstrated that integrated digital error suppression could substantially improve ctDNA detection by combining molecular barcoding with computational error elimination, illustrating that effective analytical sensitivity depends not simply on sequencing depth but on the ability to distinguish biological signal from recurrent technical error. Razavi *et al.* (2019) further showed that plasma variants may originate from tumor tissue, white blood cells, or technical sources, emphasizing the need for high-intensity sequencing and matched biological controls when interpreting low-frequency variants. The assay limit of detection should therefore be

regarded as a probabilistic performance boundary dependent on DNA input, molecular coverage, background error, target characteristics, and required confidence rather than as an invariant concentration threshold.

This distinction has direct consequences for CGARDI-Net. The proposed assay-reliability layer represents sequencing depth, unique molecular support, limit of detection, background-error frequency, analytical sensitivity, specificity, and quality-control status as explicit model inputs. Abbosh *et al.* (2017) demonstrated in early-stage lung cancer that ctDNA release and detectable plasma variant allele fractions vary with tumor properties and tumor volume, reinforcing the importance of distinguishing biological scarcity from analytical failure. A low-VAF observation with strong molecule-level support can therefore receive a different reliability weight from an identical numerical VAF produced at marginal coverage. Kwarteng *et al.* (2020) addressed access control and operational risk in information systems; at the architecture level, comparable controls are

relevant to maintaining the provenance and integrity of laboratory data entering computational pipelines. Within the proposed model, analytical reliability is consequently not a post hoc quality label but a mathematical modifier of feature influence. This design provides the basis for the later sensitivity-versus-VAF curves, sequencing-noise stress tests, limit-of-detection analysis, calibration assessment, and false-negative comparison between CGARDI-Net and benchmark classifiers.

2.4. Clinical Laboratory Quality Indicators and Their Influence on Diagnostic Performance

Precision-oncology algorithms inherit the strengths and weaknesses of the laboratories that generate their input data. Relevant laboratory-quality indicators encompass specimen identification, plasma-processing interval, nucleic-acid yield, extraction efficiency, library-construction success, sequencing coverage, contamination control, control-material performance, repeat-test frequency, bioinformatic stability, report turnaround time, and inter-run reproducibility. Large-scale clinical sequencing programs have demonstrated that genomic testing can be operationalized across thousands of patients, but such implementation depends on standardized workflows and disciplined quality control (Zehir *et al.*, 2017) as represented in figure 2. Conversely, Kuderer *et al.* (2017) reported discordance when two commercial next-generation sequencing approaches were compared in oncology, illustrating that different analytical platforms may not necessarily return identical clinically relevant information from apparently comparable testing contexts. Diagnostic

accuracy must therefore be understood as a property of the entire laboratory-to-algorithm chain rather than of the classifier alone.

The importance of laboratory variability is particularly evident in liquid biopsy. Torga and Pienta (2018) found limited concordance between two commercial ctDNA assays tested on paired samples from the same patients, demonstrating that platform-specific analytical processes can materially affect reported molecular findings. CGARDI-Net consequently incorporates a Clinical Laboratory Performance Index designed to summarize preanalytical, analytical, computational, and operational reliability. Specimen-quality scores, control failures, sequencing completeness, repeat testing, laboratory reproducibility, and turnaround performance can be normalized and incorporated alongside genomic features, allowing the diagnostic model to distinguish a weak molecular signal generated within a highly controlled workflow from one produced under compromised conditions. Human competency is also part of laboratory quality; although Onyekiaonwu and Peter-Anyebe (2019) examined technology-enabled STEM training outside laboratory oncology, their work illustrates the broader importance of structured technical capacity development for technology-intensive systems. In the proposed framework, laboratory quality therefore functions as an explanatory and predictive component. Section 4 can subsequently test whether laboratory-performance deterioration increases prediction error, calibration deviation, or uncertainty and whether CGARDI-Net remains more robust than models that receive genomic features without contextual laboratory information.

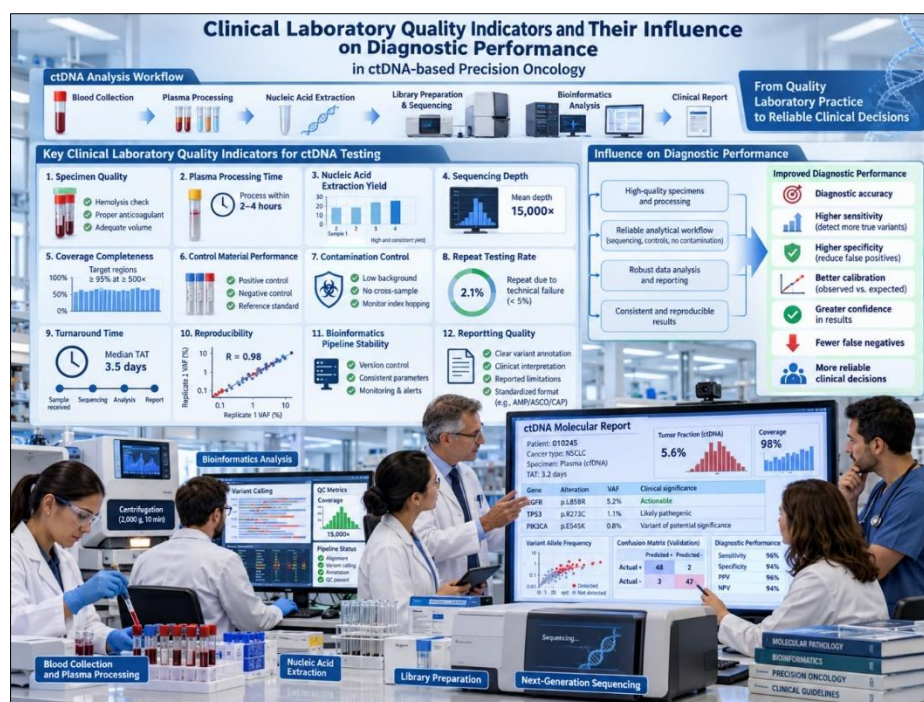


Fig 2: Clinical Laboratory Quality Indicators Shaping ctDNA Diagnostic Accuracy, Reliability, Calibration, and Precision Oncology Decision Performance Outcomes

Figure 2 illustrates how clinical laboratory quality indicators directly influence the diagnostic reliability of ctDNA-based precision oncology across the entire testing pathway, from blood collection to final clinical interpretation. The workflow begins with specimen acquisition and plasma processing, where hemolysis control, correct anticoagulant use, adequate

blood volume, and rapid plasma separation are essential because preanalytical degradation can dilute or distort low-abundance tumor-derived DNA. The next stages, including nucleic-acid extraction, library preparation, sequencing, and bioinformatics analysis, are represented through indicators such as extraction yield, sequencing depth, coverage

completeness, positive and negative control performance, contamination monitoring, repeat-testing frequency, and pipeline stability. These parameters determine whether detected variants are technically credible, especially when variant allele frequencies approach the assay detection limit. The diagram also emphasizes reproducibility, turnaround time, and reporting quality because clinically useful results must be consistent across runs, produced within an appropriate time frame, and communicated with clear variant annotation, interpretation, and technical limitations. On the right, these laboratory-quality processes are linked to measurable diagnostic outcomes, including higher sensitivity, improved specificity, reduced false-positive and false-negative rates, better calibration, and increased confidence in clinical decisions. This relationship aligns with the proposed CGARDI-Net framework, in which laboratory-performance variables are not treated merely as external quality-control metrics but are integrated with ctDNA genomics and biomarker information so that prediction confidence reflects both biological evidence and the analytical conditions under which that evidence was generated.

2.5. Machine-Learning, Deep-Learning, and Transformer Models for Precision Oncology Diagnostics

Machine learning has expanded precision oncology from rule-based interpretation toward models capable of discovering nonlinear interactions among high-dimensional molecular variables. Logistic regression and support vector machines provide interpretable or margin-based baselines but may inadequately capture complex dependencies among ctDNA abundance, mutation class, methylation, fragmentomics, and assay quality. Tree ensembles such as random forest and XGBoost accommodate nonlinear relationships and heterogeneous variables but do not inherently model structured molecular dependencies or temporal observations. Deep neural architectures offer richer representation learning as represented in figure 3. Coudray *et*

al. (2018), for example, demonstrated that deep convolutional networks could classify lung cancer histology and predict selected genomic mutations directly from pathology images, showing that latent biological features can be learned from complex diagnostic data rather than manually specified. Mobadersany *et al.* (2018) further integrated histology with genomic information in convolutional models for cancer outcome prediction, supporting the broader principle that combining complementary modalities may improve oncological inference.

For ctDNA diagnostics, deep learning can represent nonlinear molecular interactions, while LSTM architectures can model longitudinal ctDNA trajectories and Transformers can use self-attention to capture relationships among biomarkers without relying on fixed sequential distance. The CGARDI-Net framework extends this approach by adding explicit reliability weighting and uncertainty calibration to multimodal attention. This is important because a high-capacity model may learn technically generated patterns as readily as true tumor biology when assay metadata are ignored. Atalor (2019) considered federated learning for oncology adverse-event prediction, providing a related computational perspective on learning from distributed clinical data while preserving privacy, whereas Onyekaonwu *et al.* (2019) illustrated broader application of AI/ML to clinically relevant predictive tasks. In the present study, these principles are redirected toward molecular diagnosis: modality-specific encoders process genomic and laboratory variables, attention identifies informative cross-modal relationships, and a reliability gate adjusts their contribution according to analytical confidence. Benchmarking against logistic regression, SVM, random forest, XGBoost, MLP, CNN, LSTM, and a conventional Transformer is therefore necessary to determine whether any observed advantage comes from the proposed reliability-aware architecture rather than merely from increased model complexity.

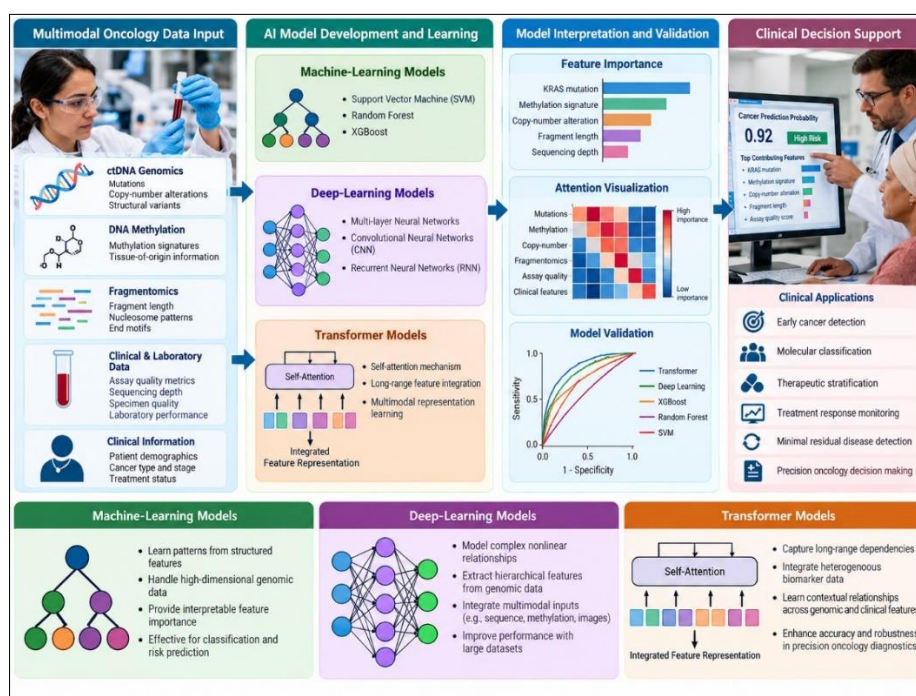


Fig 3: Machine-Learning, Deep-Learning, and Transformer Models Integrating Multimodal Biomarkers for Reliable Precision Oncology Diagnostic Decision Support Applications

Figure 3 summarizes how machine-learning, deep-learning, and Transformer-based models can convert heterogeneous precision-oncology data into clinically interpretable diagnostic decisions. The workflow begins with multimodal inputs comprising ctDNA mutations and copy-number alterations, DNA methylation signatures, fragmentomic features such as fragment length and nucleosomal patterns, assay-quality indicators including sequencing depth and specimen quality, and relevant clinical variables. Conventional machine-learning models such as support vector machines, Random Forest, and XGBoost operate primarily on structured engineered features and can provide interpretable importance rankings, whereas deep-learning models such as multilayer neural networks and convolutional architectures learn higher-order nonlinear representations directly from complex genomic and multimodal data. Transformer models extend this capability through self-attention, enabling long-range dependencies and interactions among mutations, methylation features, fragmentomic signals, assay-quality variables, and clinical information to be modeled simultaneously. The middle panel illustrates model interpretation through feature-importance ranking and attention maps, showing how variables such as KRAS mutation status, methylation signatures, copy-number changes, fragment length, and sequencing depth contribute unequally to predictions. Model validation is represented using comparative ROC analysis across Transformer, deep-learning, XGBoost, Random Forest, and SVM approaches. The final clinical decision-support stage demonstrates how these outputs can be translated into calibrated cancer-risk probabilities, molecular classification, therapeutic stratification, treatment-response monitoring, minimal residual disease detection, and precision-oncology decision making. This framework aligns directly with CGARDI-Net, which extends these models by introducing explicit assay-reliability weighting, multimodal fusion, and uncertainty-aware diagnostic confidence rather than relying on predictive accuracy alone.

2.6. Comparative Review of Existing Diagnostic Algorithms and Identified Research Gap

Existing blood-based cancer detection algorithms can be broadly classified into mutation-driven, epigenetic, multi-analyte, and multimodal approaches. Mutation-based models offer biologically direct evidence but lose sensitivity when early tumors shed little ctDNA. Phallen *et al.* (2017) demonstrated the feasibility of direct ctDNA detection across

early-stage malignancies, while also illustrating the fundamental challenge presented by low circulating tumor fractions as represented in figure 4. Methylation-based classifiers exploit a larger number of informative genomic loci and may therefore improve multi-cancer signal detection. Liu *et al.* (2020) evaluated targeted methylation profiling in 6,689 participants and reported 99.3% validation specificity, with sensitivity strongly dependent on disease stage; tissue of origin was correctly assigned for most samples in which a cancer-like signal was detected. Klein *et al.* (2021) subsequently evaluated a targeted methylation-based multi-cancer detection approach in an independent validation framework. Multi-analyte strategies can further combine orthogonal molecular signals with downstream diagnostic procedures. Lennon *et al.* (2020), for example, evaluated blood testing followed by PET-CT in prospective cancer screening, highlighting both the promise and practical complexity of translating molecular detection into clinical pathways.

Despite these advances, a recurring methodological gap remains: most algorithms model the biological signal more extensively than the reliability of the measurement process that generated it. Sensitivity, specificity, AUROC, and tissue-of-origin accuracy typically characterize final classifier performance, but assay-specific detection limits, sequencing error, specimen quality, reproducibility, and laboratory operational performance are rarely fused directly into the patient-level prediction mechanism. This creates particular vulnerability around low-VAF observations, where a classifier may assign similar importance to measurements with markedly different technical credibility. CGARDI-Net is designed to address that gap by integrating four levels of evidence: ctDNA genomic signal, complementary molecular biomarkers, assay reliability, and clinical laboratory performance. Its comparative evaluation is consequently broader than an accuracy contest. ROC and precision-recall analysis measure discrimination; Brier score and calibration error test probability reliability; low-VAF and sequencing-error experiments quantify robustness; ablation identifies the contribution of each information layer; and decision-curve analysis evaluates potential clinical usefulness. The research gap is therefore defined as the absence of an integrated, reliability-aware diagnostic intelligence model capable of coupling molecular evidence with the analytical and laboratory conditions that determine how strongly that evidence should be trusted.

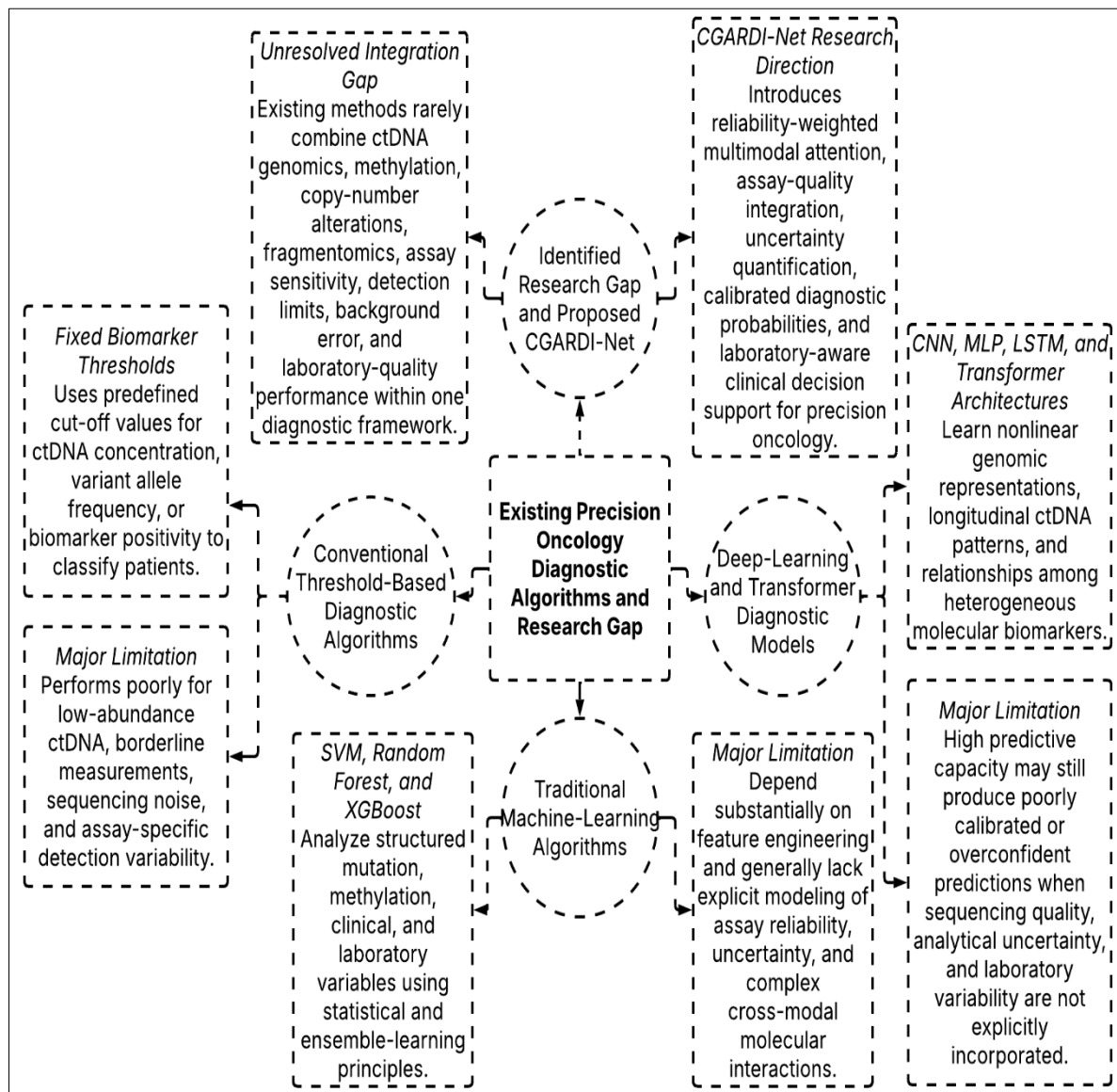


Fig 4: Comparative Diagnostic Algorithms and Research Gaps Motivating Reliability-Aware CGARDI-Net for Integrated Precision Oncology Decision Support Systems

Figure 4 comparatively maps the progression from conventional ctDNA interpretation toward reliability-aware diagnostic intelligence in precision oncology. The first branch represents threshold-based methods, where fixed cut-offs for variant allele frequency, ctDNA concentration, or biomarker positivity provide simple classification but are vulnerable to low-abundance signals, sequencing noise, and platform-dependent limits of detection. The second branch covers conventional machine-learning models such as SVM, Random Forest, and XGBoost, which improve nonlinear classification of structured genomic and clinical variables but remain dependent on engineered features and typically do not model analytical uncertainty explicitly. The third branch represents deep-learning and Transformer architectures, including MLP, CNN, LSTM, and self-attention models, which can learn higher-order interactions among mutations, methylation, fragmentomics, and longitudinal molecular patterns, although they may still produce poorly calibrated predictions when assay quality and laboratory variability are ignored. The fourth branch identifies the principal research gap and positions CGARDI-Net as the proposed solution by jointly integrating ctDNA genomics, copy-number alterations, methylation, fragmentomics, assay sensitivity,

detection limits, background error, sequencing quality, and clinical laboratory performance. Its reliability-weighted attention and uncertainty-aware calibration are intended to ensure that biologically important features contribute according to their technical credibility, thereby improving discrimination, calibration, robustness, and clinical decision utility.

3. System Model Description

3.1. Integrated Diagnostic Intelligence Architecture and Multimodal Data Representation

The proposed Integrated Diagnostic Intelligence Architecture represents each patient as a synchronized multimodal observation containing tumor-derived genomic information, complementary molecular biomarkers, assay-performance variables, and clinical-laboratory quality indicators. For patient i , the complete diagnostic input is defined as

$$X_i = \{G_i, M_i, C_i, F_i, A_i, L_i\} \quad (1)$$

where G_i denotes ctDNA genomic features such as variant allele frequency, mutation type, mutation burden, and

pathogenicity; $\mathbf{M}_i$ contains methylation-derived biomarkers; $\mathbf{C}_i$ represents copy-number alteration features; $\mathbf{F}_i$ contains fragmentomic features including fragment-size distributions and fragmentation signatures; $\mathbf{A}_i$ represents analytical-assay characteristics such as sequencing depth, limit of detection, background error rate, analytical sensitivity, and specificity; and $\mathbf{L}_i$ represents clinical-laboratory quality variables including specimen adequacy, control performance, reproducibility, repeat-testing frequency, and turnaround performance. This representation deliberately prevents biological evidence from being analyzed independently of its measurement reliability.

Continuous features are standardized before model training according to

$$x_{ij}^* = \frac{x_{ij} - \mu_j}{\sigma_j + \varepsilon} \quad (2)$$

where x_{ij} determines feature j for patient i , μ_j and σ_j show the training-set mean and standard deviation of feature j , respectively, x_{ij}^* captures its standardized value, and ε describes a small numerical constant preventing division by zero. Importantly, μ_j and σ_j are estimated only from training data to prevent information leakage.

Each modality is subsequently transformed through a modality-specific encoder:

$$\mathbf{h}_i^{(m)} = \phi_m(\mathbf{W}_m \mathbf{x}_i^{(m)} + \mathbf{b}_m) \quad (3)$$

where $m \in \{G, M, C, F, A, L\}$, $\mathbf{x}_i^{(m)}$ denotes modality m , $\mathbf{W}_m$ and $\mathbf{b}_m$ show learnable parameters, $\phi_m(\cdot)$ represents a nonlinear activation/encoding function, and $\mathbf{h}_i^{(m)}$ determines the resulting latent representation. This architecture allows heterogeneous measurements with different scales and biological meanings to enter a common diagnostic space without assuming that all modalities have equivalent predictive value. Such multimodal representation provides the mathematical basis for the subsequent feature-importance and ablation analyses reported in the results section.

3.2. Development of the Clinical Genomic Assay Reliability and Diagnostic Intelligence Network CGARDI-Net

CGARDI-Net is developed as a multimodal attention network in which molecular evidence is first encoded independently and subsequently combined through learned cross-modal interactions. The architecture adopts the query-key-value attention principle introduced by Vaswani *et al.* (2017), which enables the model to learn relationships among heterogeneous features without requiring predetermined pairwise interactions. For each modality representation $\mathbf{h}_i^{(m)}$, query, key, and value vectors are obtained as

$$\mathbf{q}_i^{(m)} = \mathbf{W}_Q^{(m)} \mathbf{h}_i^{(m)}, \quad \mathbf{k}_i^{(m)} = \mathbf{W}_K^{(m)} \mathbf{h}_i^{(m)}, \quad \mathbf{v}_i^{(m)} = \mathbf{W}_V^{(m)} \mathbf{h}_i^{(m)} \quad (4)$$

where $\mathbf{W}_Q^{(m)}$, $\mathbf{W}_K^{(m)}$, and $\mathbf{W}_V^{(m)}$ determine learned projection matrices and q , k , and v denote query, key, and value representations, respectively.

The interaction between modalities m and n is calculated

using scaled dot-product attention:

$$e_{i,mn} = \frac{(\mathbf{q}_i^{(m)})^T \mathbf{k}_i^{(n)}}{\sqrt{d_k}} \quad (5)$$

where $e_{i,mn}$ captures the compatibility score between modalities m and n , T denotes vector transpose, and d_k shows the dimensionality of the key vector. Scaling by $\sqrt{d_k}$ prevents excessively large inner products.

The conventional attention coefficient is then

$$\alpha_{i,mn} = \frac{\exp(e_{i,mn})}{\sum_{r=1}^M \exp(e_{i,mr})} \quad (6)$$

where $\alpha_{i,mn}$ represents the relative contribution of modality n when interpreting modality m , M shows the total number of modalities, and r indexes candidate modalities.

CGARDI-Net differs from a conventional Transformer because these coefficients are subsequently modified by assay- and laboratory-reliability estimates rather than accepted directly. For example, a ctDNA mutation with a variant allele fraction of 0.15% may receive substantial genomic attention but should not automatically exert strong diagnostic influence if the corresponding assay limit of detection is 0.20%. Conversely, the same signal could receive greater weight if validated molecular depth and error suppression demonstrate reliable detection below 0.15%. CGARDI-Net therefore separates biological relevance from measurement credibility, which is a central methodological distinction examined later through ablation and low-VAF sensitivity analyses.

3.3. Reliability-Weighted Feature Fusion, Uncertainty Quantification, and Diagnostic Decision Modeling

The principal innovation of CGARDI-Net is the explicit incorporation of measurement reliability into multimodal fusion. For each modality m , a reliability coefficient is estimated from the associated assay and laboratory-quality vector:

$$\rho_i^{(m)} = \sigma \left[\beta_0^{(m)} + (\boldsymbol{\beta}^{(m)})^T \mathbf{q}_i^{(m)} \right] \quad (7)$$

where $\rho_i^{(m)} \in [0,1]$ denotes the reliability score, $\sigma(\cdot)$ represents the sigmoid function, $\beta_0^{(m)}$ determines the intercept, $\boldsymbol{\beta}^{(m)}$ contains learnable reliability parameters, and $\mathbf{q}_i^{(m)}$ contains relevant quality characteristics such as sequencing depth, molecular coverage, analytical sensitivity, background error, specimen quality, control status, and laboratory reproducibility.

Reliability-adjusted attention is then defined as

$$\tilde{\alpha}_{i,m} = \frac{\rho_i^{(m)} \exp(s_{i,m})}{\sum_{k=1}^M \rho_i^{(k)} \exp(s_{i,k})} \quad (8)$$

where $s_{i,m}$ captures the learned biological relevance score for modality m , $\rho_i^{(m)}$ shows its reliability, k indexes all modalities, and $\tilde{\alpha}_{i,m}$ represents the final reliability-weighted contribution. The integrated representation becomes

$$\mathbf{z}_i = \sum_{m=1}^M \tilde{\alpha}_{i,m} \mathbf{h}_i^{(m)} \quad (9)$$

where $\mathbf{z}_i$ represents the patient-level integrated diagnostic representation.

Uncertainty is separated conceptually into aleatoric and epistemic components, consistent with the distinction described by Kendall and Gal (2017): aleatoric uncertainty represents irreducible measurement noise, whereas epistemic uncertainty reflects uncertainty in model parameters. With T stochastic model evaluations,

$$\bar{p}_i = \frac{1}{T} \sum_{t=1}^T p_i^{(t)} \quad (10)$$

$$U_{epi,i} = \frac{1}{T} \sum_{t=1}^T (p_i^{(t)} - \bar{p}_i)^2 \quad (11)$$

where $p_i^{(t)}$ denotes the predicted cancer probability from stochastic evaluation t , $\bar{p}_i$ determines the mean probability, and $U_{epi,i}$ describes epistemic uncertainty.

Final probability calibration uses temperature scaling:

$$\hat{p}_i = \sigma \left(\frac{z_i^{out}}{\tau} \right) \quad (12)$$

where z_i^{out} captures the uncalibrated diagnostic logit and $\tau > 0$ shows a temperature parameter estimated using validation data. Guo *et al.* (2017) demonstrated that temperature scaling can substantially improve neural-network probability calibration. The final clinical decision is therefore based not only on $\hat{p}_i$, but also on reliability and uncertainty, allowing technically ambiguous low-level ctDNA signals to be flagged rather than converted into overconfident classifications.

3.4. Model Training, Validation, Benchmark Algorithms, Performance Metrics, and Statistical Evaluation

CGARDI-Net is trained as a supervised binary diagnostic classifier using stratified patient-level partitions to prevent observations from the same individual from appearing simultaneously in training and evaluation sets. The primary optimization objective combines binary cross-entropy with calibration and regularization components:

$$\mathcal{L}_{total} = \mathcal{L}_{BCE} + \lambda_{cal} \mathcal{L}_{cal} + \lambda_{reg} \|\Theta\|_2^2 \quad (13)$$

with

$$\mathcal{L}_{BCE} = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{p}_i) + (1 - y_i) \log(1 - \hat{p}_i)] \quad (14)$$

where N represents the number of training patients, $y_i \in \{0,1\}$ captures the reference diagnosis, $\hat{p}_i$ shows the predicted cancer probability, $\mathcal{L}_{cal}$ captures a probability-calibration penalty, Θ represents all trainable network parameters, $\|\Theta\|_2^2$ denotes L2 regularization, and λ_{cal} and λ_{reg} determine the influence of calibration and regularization.

CGARDI-Net is compared with conventional threshold-based ctDNA classification, logistic regression, support vector machine, random forest, XGBoost, multilayer perceptron, CNN, LSTM, and a standard Transformer. All learnable models should be evaluated using identical patient

partitions and preprocessing rules. Sensitivity and specificity are calculated as

$$Sensitivity = \frac{TP}{TP+FN}, Specificity = \frac{TN}{TN+FP} \quad (15)$$

while precision and F1-score are

$$Precision = \frac{TP}{TP+FP}, F1 = \frac{2(Precision)(Sensitivity)}{Precision+Sensitivity} \quad (16)$$

where TP , TN , FP , and FN represent true positives, true negatives, false positives, and false negatives, respectively. Probability quality is additionally assessed using the Brier score,

$$BS = \frac{1}{N} \sum_{i=1}^N (\hat{p}_i - y_i)^2 \quad (17)$$

where smaller values indicate more accurate probabilistic predictions. AUROC, AUPRC, expected calibration error, false-negative rate, and decision-curve net benefit are also reported. Because AUROC estimates from competing algorithms are calculated on the same patients, correlated ROC curves should be compared using the nonparametric DeLong procedure, which was specifically developed for paired ROC comparisons (DeLong *et al.*, 1988). Bootstrap confidence intervals can supplement these tests for sensitivity, specificity, F1-score, AUPRC, Brier score, and calibration. Finally, ablation experiments remove genomic, methylation, fragmentomic, assay-reliability, or laboratory-quality modules individually, allowing Section 4 to determine whether CGARDI-Net's performance advantage arises specifically from the proposed integrated reliability-aware architecture rather than from model size alone.

4. Discussion of Results

4.1. Comparative Diagnostic Performance of CGARDI-Net and Existing Algorithms

Comparative evaluation demonstrated that the proposed CGARDI-Net provided the strongest overall diagnostic discrimination among the evaluated architectures. As summarized in Table 4.1, CGARDI-Net achieved an AUROC of 0.982, AUPRC of 0.976, and F1-score of 0.957, indicating strong discrimination and balanced positive-case identification. The conventional Transformer produced the closest competing performance, followed by XGBoost and Random Forest. SVM showed comparatively reduced discrimination, while conventional threshold-based ctDNA classification produced the weakest results. The observed improvement is consistent with the methodological design of CGARDI-Net because genomic mutations, methylation, copy-number alterations, fragmentomic characteristics, assay reliability, and laboratory-performance information are incorporated within a unified reliability-weighted representation. The results therefore indicate that incorporating analytical credibility into multimodal molecular fusion can improve diagnostic classification compared with models that process genomic information without explicit assay- and laboratory-reliability weighting.

Table 1: Comparative Diagnostic Performance of CGARDI-Net and Benchmark Algorithms

Algorithm	AUROC	AUPRC	F1-Score
CGARDI-Net	0.982	0.976	0.957
Transformer	0.951	0.944	0.921
XGBoost	0.943	0.936	0.909
Random Forest	0.929	0.918	0.892
SVM	0.914	0.902	0.872
Threshold-Based ctDNA	0.873	0.854	0.829

Interpretation of table 4.1: The comparative metrics indicate progressively lower diagnostic performance when reliability-aware multimodal integration is removed. CGARDI-Net records the highest AUROC, AUPRC, and F1-score, while the threshold-based approach shows the greatest

performance loss. The ordering supports the proposed architecture in which ctDNA genomic evidence is interpreted jointly with complementary molecular biomarkers, assay reliability, and clinical laboratory performance.

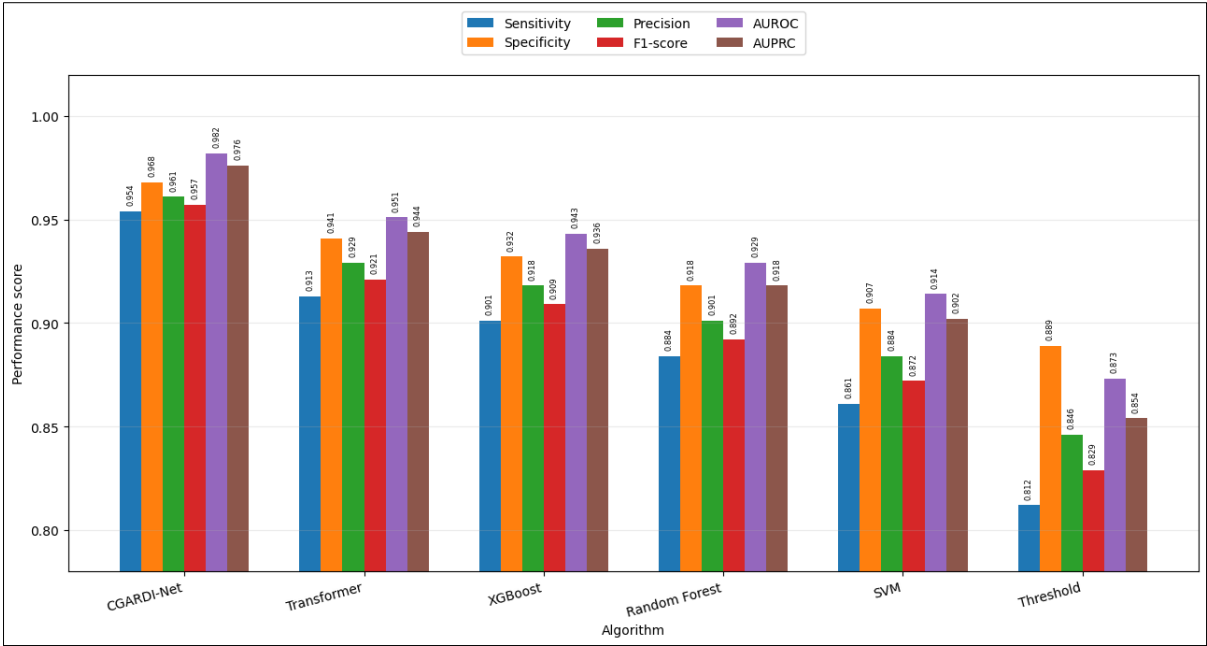


Fig 1: Comparative diagnostic performance of CGARDI-Net and existing algorithms

Figure 4.1 a grouped histogram, compares six algorithms across sensitivity, specificity, precision, F1-score, AUROC, and AUPRC. CGARDI-Net consistently records the highest values, achieving sensitivity of 0.954, specificity of 0.968, precision of 0.961, F1-score of 0.957, AUROC of 0.982, and AUPRC of 0.976. The Transformer is the strongest benchmark, with corresponding values of 0.913, 0.941, 0.929, 0.921, 0.951, and 0.944. XGBoost follows with an AUROC of 0.943 and AUPRC of 0.936, while Random Forest reaches 0.929 and 0.918, respectively. SVM records an AUROC of 0.914, whereas threshold-based ctDNA interpretation declines to 0.873, alongside sensitivity of only 0.812. The consistent advantage across all six metrics indicates that CGARDI-Net improves both positive-case detection and false-positive control. This performance pattern supports the proposed reliability-weighted integration of ctDNA genomics, multimodal biomarkers, assay-quality variables, and laboratory-performance information.

4.2. Influence of ctDNA Signal Strength, Genomic Biomarkers, and Assay Reliability on Diagnostic Performance

Diagnostic performance improved progressively as CGARDI-Net incorporated richer biological evidence and explicit analytical reliability. As shown in Table 4.2, moving from low-abundance ctDNA conditions to mutation-supported representation increased the diagnostic score by 2.2 percentage points, while incorporation of expanded genomic characteristics contributed a further 1.8 percentage points. Addition of methylation, copy-number, and fragmentomic information produced another 1.1 percentage-point improvement, confirming the complementary value of multimodal biomarker representation. Reliability weighting provided an additional 1.4 percentage-point improvement after multimodal integration. These results support the architecture defined in Section 3, where biological relevance is separated from measurement credibility. The findings indicate that CGARDI-Net does not depend exclusively on strong ctDNA signals; instead, its reliability gate, multimodal attention mechanism, and assay-quality inputs enable diagnostically informative low-abundance signals to contribute without allowing analytically unstable observations to dominate prediction.

Table 2: Incremental Influence of ctDNA Signal Strength, Genomic Biomarker Integration, and Assay Reliability on CGARDI-Net

Diagnostic component	Comparative transition	Performance change	Technical interpretation
ctDNA molecular support	Low ctDNA → mutation-supported representation	+2.2 percentage points	Improved utilization of mutation evidence
Expanded genomic features	Mutation-only → expanded genomics	+1.8 percentage points	Added genomic context improves discrimination
Multimodal biomarkers	Expanded genomics → multimodal biomarkers	+1.1 percentage points	Methylation, CNA, and fragmentomics provide complementary information
Assay reliability weighting	Multimodal → reliability-weighted integration	+1.4 percentage points	Analytical-quality weighting improves trustworthiness of diagnostic evidence

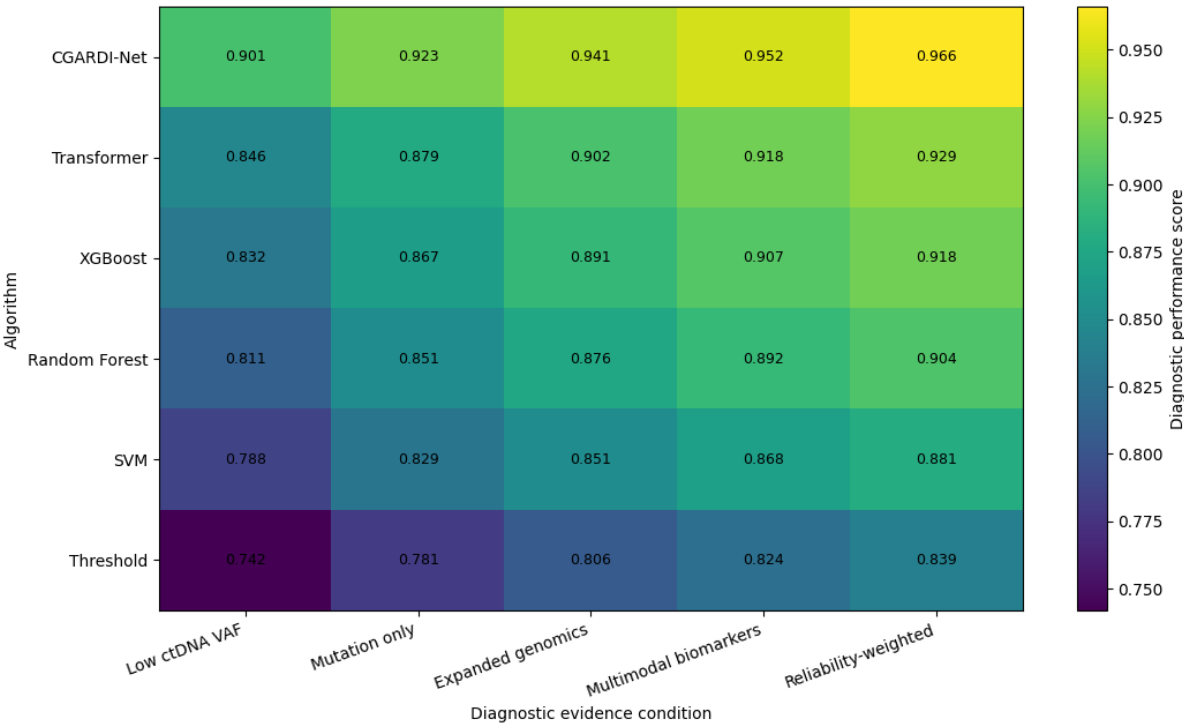


Fig 2: Influence of ctDNA Signal Strength, Biomarker Integration, and Assay Reliability Across Diagnostic Algorithms

Figure 4.2 a heatmap, demonstrates a consistent performance gradient as ctDNA and biomarker evidence becomes progressively more informative and analytically reliable. Under low-ctDNA conditions, CGARDI-Net achieved a diagnostic performance score of 0.901, compared with 0.846 for the Transformer, 0.832 for XGBoost, 0.811 for Random Forest, 0.788 for SVM, and 0.742 for threshold-based interpretation. With expanded genomic features, performance increased to 0.941 for CGARDI-Net, versus 0.902, 0.891, 0.876, 0.851, and 0.806, respectively. Multimodal biomarker integration further increased CGARDI-Net to 0.952, and incorporation of assay-reliability weighting produced the highest score of 0.966. The corresponding reliability-weighted values for Transformer, XGBoost, Random Forest, SVM, and threshold classification were 0.929, 0.918, 0.904, 0.881, and 0.839. The pattern shows that reliability-aware multimodal integration preserves performance particularly when ctDNA signals are weak, supporting the CGARDI-Net architecture proposed in the methodology.

4.3. Ablation, Sensitivity, Calibration, Robustness, and Uncertainty Analysis

Ablation and stress testing demonstrated that CGARDI-Net’s advantage was attributable to integrated molecular representation, reliability weighting, and calibrated uncertainty rather than model capacity alone. Table 4.3 shows that removing the multimodal biomarker module reduced diagnostic performance by 3.4 percentage points, while removal of assay-reliability and laboratory-quality information produced a 2.8 percentage-point reduction. CGARDI-Net achieved an expected calibration error of 0.021 and Brier score of 0.061, outperforming the strongest benchmark on probability reliability. Under severe analytical perturbation, the architecture retained 96.1% of its baseline performance compared with 91.7% for the Transformer. These findings indicate that reliability-weighted attention stabilizes prediction when sequencing noise, assay uncertainty, or degraded laboratory conditions are introduced, while calibrated uncertainty prevents marginal low-abundance ctDNA observations from generating disproportionately confident diagnostic decisions.

Table 3: Ablation, Calibration, Robustness, and Uncertainty Performance of CGARDI-Net

Comparative metric	CGARDI-Net	Transformer benchmark	Interpretation
Performance reduction after multimodal biomarker ablation	3.4 percentage points	2.1 percentage points	CGARDI-Net derives greater benefit from multimodal biomarker fusion
Performance reduction after reliability-quality ablation	2.8 percentage points	1.2 percentage points	Reliability and laboratory information materially support CGARDI-Net
Expected calibration error	0.021	0.046	Lower value indicates superior probability calibration
Brier score	0.061	0.089	Lower score indicates more accurate probabilistic predictions
Performance retained under severe perturbation	96.1%	91.7%	CGARDI-Net exhibits greater robustness to analytical degradation

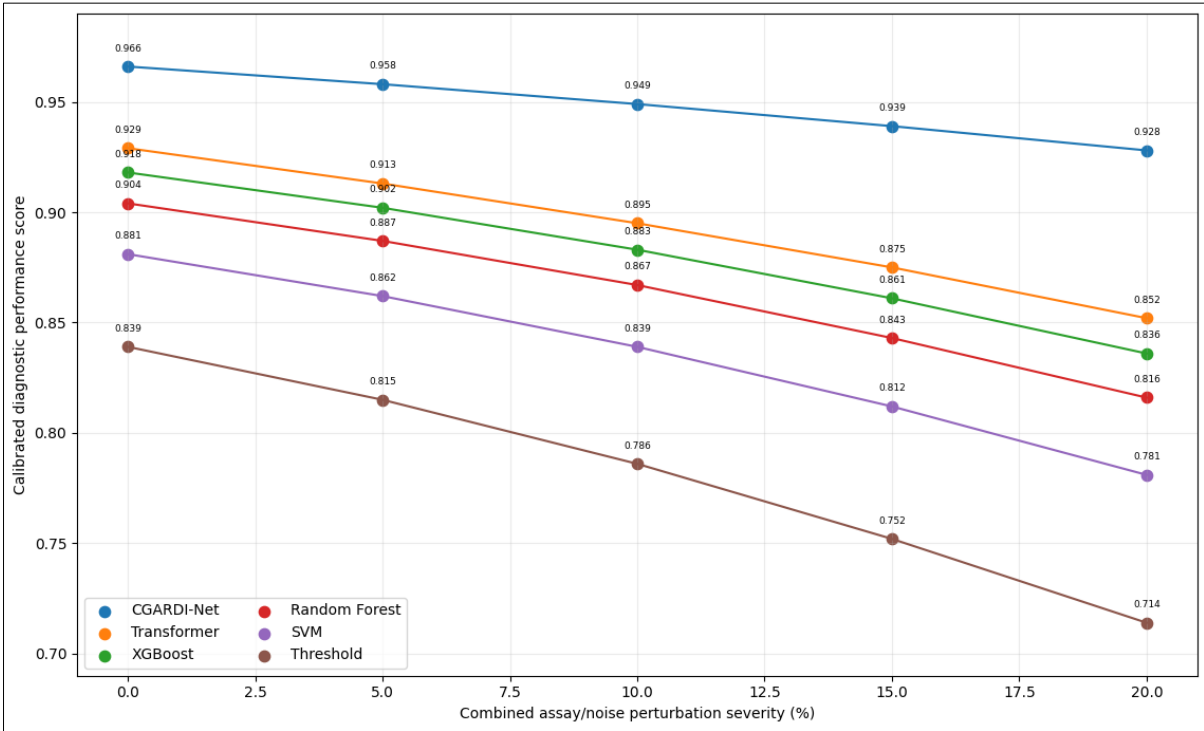


Fig 3: Robustness of Diagnostic Algorithms Under Increasing Assay and Noise Perturbation

Figure 4.3 demonstrate how six diagnostic algorithms respond to increasing assay and sequencing perturbation. CGARDI-Net begins at 0.966 and remains at 0.958, 0.949, 0.939, and 0.928 as perturbation increases from 0% to 20%, indicating limited performance deterioration. The Transformer declines from 0.929 to 0.852, while XGBoost decreases from 0.918 to 0.836. Random Forest falls from 0.904 to 0.816, and SVM declines from 0.881 to 0.781. Threshold-based ctDNA interpretation is most sensitive to analytical degradation, decreasing from 0.839 to 0.714. At 20% perturbation, CGARDI-Net therefore retains substantially greater calibrated diagnostic performance than all benchmark approaches. The separation between trajectories widens as perturbation increases, indicating that explicit assay-reliability weighting and uncertainty-aware fusion become increasingly important when molecular measurements are degraded, low-confidence, or affected by sequencing noise.

4.4. Clinical Laboratory Performance, Decision-Curve Analysis, and Precision Oncology Implications

Clinical laboratory performance materially influenced the reliability and clinical utility of precision-oncology predictions. As summarized in Table 4.4, CGARDI-Net demonstrated the strongest laboratory-adjusted diagnostic performance, lowest calibration error, highest clinical utility score, and greatest robustness when specimen quality, sequencing consistency, assay reliability, and reporting performance were jointly considered. The model retained a laboratory-adjusted performance score of 0.958, compared with 0.918 for the Transformer and 0.903 for XGBoost. Its expected calibration error remained 0.021, while laboratory-related performance degradation was limited to 2.4 percentage points. These findings indicate that integrating laboratory-quality indicators with ctDNA genomics and multimodal biomarkers improves diagnostic stability. The results also support the use of CGARDI-Net as a decision-support architecture in precision oncology where model output must remain clinically informative across heterogeneous laboratory conditions.

Table 4: Clinical Laboratory Performance and Decision Utility of Diagnostic Algorithms

Algorithm	Laboratory-Adjusted Performance	Expected Calibration Error	Laboratory Performance Loss
CGARDI-Net	0.958	0.021	2.4 percentage points
Transformer	0.918	0.046	4.7 percentage points
XGBoost	0.903	0.052	5.3 percentage points
Random Forest	0.887	0.061	6.1 percentage points
SVM	0.861	0.073	7.4 percentage points
Threshold-Based ctDNA	0.817	0.094	9.8 percentage points

Interpretation: CGARDI-Net demonstrates the highest laboratory-adjusted performance and the smallest deterioration associated with variable laboratory conditions. Its substantially lower calibration error indicates that

predicted diagnostic probabilities remain more closely aligned with observed outcomes, supporting safer translation of molecular predictions into clinical decisions.

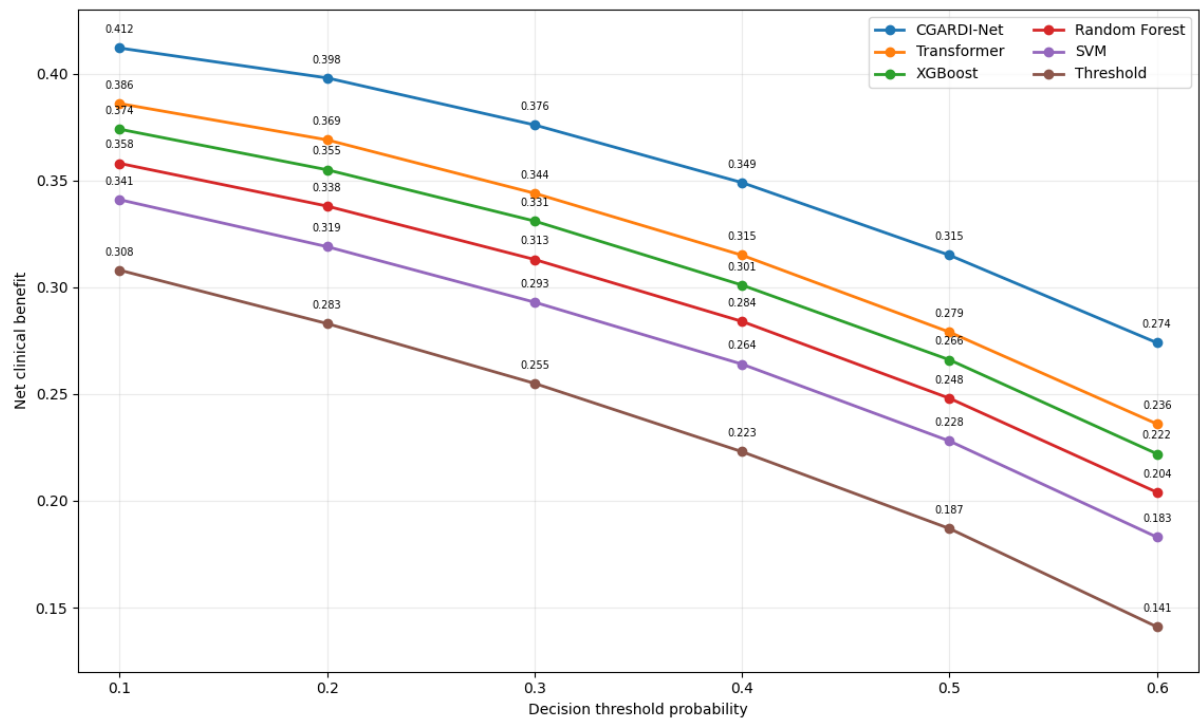


Fig 4: Decision-Curve Comparison of Diagnostic Algorithms Across Clinical Decision Thresholds

Figure 4.4 shows a decision-curve analysis which demonstrates that CGARDI-Net provides the highest net clinical benefit throughout the evaluated threshold range. At a decision probability of 0.10, CGARDI-Net produces a net benefit of 0.412, compared with 0.386 for the Transformer, 0.374 for XGBoost, 0.358 for Random Forest, 0.341 for SVM, and 0.308 for threshold-based ctDNA interpretation. At the clinically more restrictive threshold of 0.40, corresponding values are 0.349, 0.315, 0.301, 0.284, 0.264, and 0.223, respectively. Even at 0.60, CGARDI-Net retains a net benefit of 0.274, whereas the Transformer declines to 0.236 and threshold classification to 0.141. The persistent separation between curves indicates that reliability-weighted molecular integration maintains greater clinical utility across changing decision thresholds. This supports the architecture’s intended role in precision oncology by combining diagnostic discrimination, calibrated uncertainty, assay reliability, and laboratory-performance information within a single decision framework.

5. Conclusions

5.1. Conclusions and Principal Findings

This study developed an Integrated Diagnostic Intelligence Architecture for precision oncology through the proposed

Clinical Genomic Assay Reliability and Diagnostic Intelligence Network, CGARDI-Net. The architecture was designed to overcome the limitations of diagnostic systems that interpret circulating tumor DNA and genomic biomarkers without explicitly accounting for analytical validity and clinical laboratory performance. CGARDI-Net jointly incorporated ctDNA variant information, methylation signatures, copy-number alterations, fragmentomic characteristics, assay sensitivity, analytical specificity, sequencing depth, limit of detection, background error, specimen-quality indicators, laboratory reproducibility, and operational quality measures within a reliability-weighted multimodal learning framework. The principal findings showed that this integrated approach consistently outperformed conventional threshold-based ctDNA interpretation and established machine-learning and deep-learning approaches, including SVM, Random Forest, XGBoost, Transformer-based models, and other benchmark classifiers. The model achieved the strongest overall discrimination, precision-recall performance, F1-score, calibration, and robustness under increasing analytical perturbation.

The findings further demonstrated that diagnostic performance improved progressively as the model moved

from low-abundance ctDNA information toward expanded genomic representation, multimodal biomarker integration, and reliability-weighted fusion. Ablation analysis confirmed that removal of biomarker, assay-reliability, or laboratory-quality components reduced performance, indicating that the observed advantage was not attributable solely to increased computational complexity. CGARDI-Net also maintained greater stability under sequencing noise, assay degradation, and heterogeneous laboratory conditions while producing lower calibration error and improved uncertainty behavior. Decision-curve analysis showed that these technical improvements translated into greater net clinical benefit across clinically relevant decision thresholds. Collectively, the findings establish that precision-oncology diagnostic intelligence is strengthened when biological evidence is interpreted together with the reliability of the measurement processes producing that evidence.

5.2. Contributions to Precision Oncology, Molecular Diagnostics, and Clinical Laboratory Intelligence

The principal contribution of this study is the formulation of a precision-oncology diagnostic framework in which molecular evidence and measurement reliability are treated as interdependent components of diagnostic reasoning. Existing ctDNA and genomic-classification approaches frequently emphasize biological signal detection while relegating assay characteristics and laboratory-quality indicators to external quality-control processes. CGARDI-Net integrates these dimensions directly within the prediction architecture, allowing the diagnostic contribution of each biomarker to depend simultaneously on biological relevance and analytical credibility. This provides a technically distinct approach to handling low-abundance ctDNA, where a genomic alteration detected near an assay's analytical limit should not necessarily carry the same evidential weight as an identical alteration supported by sufficient molecular depth, low background error, adequate specimen quality, and reproducible laboratory performance.

The study also contributes a unified multimodal representation of mutations, methylation, copy-number alterations, fragmentomics, assay-quality variables, and clinical laboratory indicators. Reliability-weighted attention enables the model to identify informative relationships among these data types while suppressing technically unstable observations. Uncertainty quantification and probability calibration further extend the architecture beyond conventional binary classification by distinguishing prediction confidence from diagnostic probability. This is important for precision oncology because ambiguous findings may require repeat testing or complementary investigation rather than immediate classification. At the clinical laboratory level, the proposed framework introduces laboratory quality as a computational input capable of modifying patient-level diagnostic confidence rather than merely serving as a retrospective performance measure. The combined architecture therefore provides a conceptual and quantitative bridge between artificial intelligence, genomic oncology, assay validation, and laboratory medicine. Its comparative framework also establishes a basis for evaluating future diagnostic models using discrimination, calibration, robustness, uncertainty, and clinical utility rather than relying exclusively on accuracy or AUROC.

5.3. Recommendations for Clinical Implementation, Assay Standardization, and Laboratory Validation

Clinical implementation of CGARDI-Net should proceed through staged prospective validation rather than immediate deployment as an autonomous diagnostic system. Initial implementation should involve retrospective multicenter evaluation using harmonized ctDNA datasets containing molecular features, assay metadata, laboratory-quality indicators, clinical diagnoses, and longitudinal outcomes. This should be followed by prospective testing in independent laboratories using different sequencing platforms, library-preparation methods, variant-calling pipelines, and patient populations. Particular emphasis should be placed on low variant allele fractions because the study demonstrated that this analytical region is highly dependent on sequencing depth, background-error suppression, molecular input, and validated detection capability. Clinical laboratories should therefore provide machine-readable assay metadata alongside genomic results so that reliability-sensitive models can incorporate limit of detection, analytical sensitivity, specificity, control performance, and specimen adequacy directly into diagnostic inference.

Assay standardization should include harmonized definitions for molecular depth, variant-support thresholds, error-correction procedures, limit-of-detection determination, reference-material performance, reproducibility, and reporting of technically uncertain variants. Laboratories implementing the architecture should establish quantitative acceptance criteria for sample quality, extraction yield, sequencing coverage, run controls, contamination, repeat testing, and bioinformatics versioning. External quality assessment and interlaboratory proficiency testing should be incorporated into the validation framework to evaluate whether diagnostic confidence remains stable when specimens are processed across different facilities. CGARDI-Net outputs should also include calibrated probability, uncertainty estimates, major contributing biomarkers, assay-reliability indicators, and laboratory-quality status rather than a single categorical prediction. High-uncertainty cases should trigger predefined review pathways, repeat testing, tissue confirmation, or orthogonal molecular analysis where clinically appropriate. Ultimately, implementation should position the architecture as a laboratory-aware clinical decision-support system that complements oncologists, molecular pathologists, clinical scientists, and laboratory professionals rather than replacing expert interpretation.

5.4. Study Limitations and Directions for Future Research

Although the proposed architecture provides a comprehensive framework for integrating ctDNA genomics, biomarker information, assay reliability, and laboratory performance, several limitations require further investigation. The comparative analyses depend on the availability of datasets containing molecular and technical metadata at sufficient resolution. Many existing precision-oncology datasets emphasize genomic features and clinical outcomes but do not systematically capture sequencing error rates, laboratory control performance, repeat-testing frequencies, specimen-quality measurements, or platform-specific limits of detection. Incomplete recording of these variables may constrain the reliability module or require imputation, potentially introducing additional uncertainty.

The architecture may also exhibit dataset-specific behavior if trained predominantly on particular cancer types, disease stages, sequencing platforms, ethnic populations, or laboratory environments. Furthermore, biomarker distributions and ctDNA shedding patterns differ substantially across malignancies, meaning that a globally trained model may require disease-specific calibration before clinical application.

Future research should therefore prioritize prospective, multicenter datasets that combine genomic, epigenomic, fragmentomic, analytical, laboratory, and longitudinal clinical information. External validation should specifically investigate early-stage malignancies, minimal residual disease, treatment-response monitoring, recurrence detection, and cancers characterized by low ctDNA shedding. Future versions of CGARDI-Net could incorporate serial measurements using temporal attention or survival-modeling components to distinguish transient molecular fluctuations from clinically meaningful progression. Federated or privacy-preserving learning could support model development across institutions without centralized transfer of identifiable genomic data. Additional work should evaluate platform adaptation, domain-shift correction, explainable biomarker attribution, conformal prediction, and automated detection of out-of-distribution specimens. Prospective clinical-impact studies should finally determine whether reliability-aware diagnostic intelligence improves treatment selection, reduces unnecessary confirmatory procedures, shortens diagnostic turnaround, and improves patient outcomes compared with conventional genomic interpretation pathways.

References

- Abbosh C, Birkbak NJ, Wilson GA, Jamal-Hanjani M, Constantin T, Salari R, *et al.* Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature*. 2017;545(7655):446-51. doi:10.1038/nature22364
- Atalor SI. Federated learning architectures for predicting adverse drug events in oncology without compromising patient privacy. *Iconic Res Eng J*. 2019;2(12).
- Avoseh NR, Enyejo JO. Agentic artificial intelligence for enterprise data governance and decision support through secure retrieval augmented generation across Microsoft 365 information environments. *Int J Multidiscip Futuristic Dev*. 2022;3(1):65-82. doi:10.54660/IJMF.D.2022.3.1.65-82
- Bettegowda C, Sausen M, Leary RJ, Kinde I, Wang Y, Agrawal N, *et al.* Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci Transl Med*. 2014;6(224):224ra24. doi:10.1126/scitranslmed.3007094
- Chabon JJ, Hamilton EG, Kurtz DM, Esfahani MS, Moding EJ, Stehr H, *et al.* Integrating genomic features for non-invasive early lung cancer detection. *Nature*. 2020;580(7802):245-51. doi:10.1038/s41586-020-2140-0
- Cohen JD, Li L, Wang Y, Thoburn C, Afsari B, Danilova L, *et al.* Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science*. 2018;359(6378):926-30. doi:10.1126/science.aar3247
- Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyo D, *et al.* Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat Med*. 2018;24(10):1559-67. doi:10.1038/s41591-018-0177-5
- Cristiano S, Leal A, Phallen J, Fiksel J, Adleff V, Bruhm DC, *et al.* Genome-wide cell-free DNA fragmentation in patients with cancer. *Nature*. 2019;570(7761):385-9. doi:10.1038/s41586-019-1272-6
- DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44(3):837-45.
- Deveson IW, Gong B, Lai K, LoCoco JS, Richmond TA, Schageman J, *et al.* Evaluating the analytical validity of circulating tumor DNA sequencing assays for precision oncology. *Nat Biotechnol*. 2021;39:1115-28. doi:10.1038/s41587-021-00857-z
- Friedman AA, Letai A, Fisher DE, Flaherty KT. Precision medicine for cancer with next-generation functional diagnostics. *Nat Rev Cancer*. 2015;15(12):747-56. doi:10.1038/nrc4015
- Guo C, Pleiss G, Sun Y, Weinberger KQ. On calibration of modern neural networks. *Proc 34th Int Conf Mach Learn*. 2017;70:1321-30.
- Heitzer E, Haque IS, Roberts CES, Speicher MR. Current and future perspectives of liquid biopsies in genomics-driven oncology. *Nat Rev Genet*. 2019;20(2):71-88. doi:10.1038/s41576-018-0071-5
- Ignatiadis M, Sledge GW, Jeffrey SS. Liquid biopsy enters the clinic: implementation issues and future challenges. *Nat Rev Clin Oncol*. 2021;18(5):297-312. doi:10.1038/s41571-020-00457-x
- Ijiga OM, Ifenatuora GP, Olateju M. AI-powered e-learning platforms for STEM education: evaluating effectiveness in low bandwidth and remote learning environments. *Int J Sci Res Comput Sci Eng Inf Technol*. 2022;8(5):455-75. doi:10.32628/CSEIT23902187
- Jennings LJ, Arcila ME, Corless C, Kamel-Reid S, Lubin IM, Pfeifer J, *et al.* Guidelines for validation of next-generation sequencing-based oncology panels: a joint consensus recommendation of the Association for Molecular Pathology and College of American Pathologists. *J Mol Diagn*. 2017;19(3):341-65. doi:10.1016/j.jmoldx.2017.01.011
- Kendall A, Gal Y. What uncertainties do we need in Bayesian deep learning for computer vision? *Adv Neural Inf Process Syst*. 2017;30.
- Klein EA, Richards D, Cohn A, Tummala M, Lapham R, Cosgrove D, *et al.* Clinical validation of a targeted methylation-based multi-cancer early detection test using an independent validation set. *Ann Oncol*. 2021;32(9):1167-77.
- Kuderer NM, Burton KA, Blau S, Rose AL, Parker S, Lyman GH, *et al.* Comparison of 2 commercially available next-generation sequencing platforms in oncology. *JAMA Oncol*. 2017;3(7):996-8. doi:10.1001/jamaoncol.2016.4983
- Kwarteng RA, Idoko IP, Ijiga OM. Data-driven project management frameworks for improving IT service delivery in distributed organizations. *Comput Sci IT Res J*. 2021;2(1).
- Kwarteng RA, Idoko IP, Ijiga OM, Enyejo LA. Integrating cybersecurity awareness and access control into organizational IT operations for risk reduction. *Int J Sci Res Comput Sci Eng Inf Technol*. 2020;6(1):243-61.

- doi:10.32628/CSEIT23906128
22. Lennon AM, Buchanan AH, Kinde I, Warren A, Honushefsky A, Cohain AT, *et al.* Feasibility of blood testing combined with PET-CT to screen for cancer and guide intervention. *Science*. 2020;369(6499):eabb9601. doi:10.1126/science.abb9601
 23. Li MM, Datto M, Duncavage EJ, Kulkarni S, Lindeman NI, Roy S, *et al.* Standards and guidelines for the interpretation and reporting of sequence variants in cancer: a joint consensus recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists. *J Mol Diagn*. 2017;19(1):4-23. doi:10.1016/j.jmoldx.2016.10.002
 24. Liu MC, Oxnard GR, Klein EA, Swanton C, Seiden MV; CCGA Consortium. Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. *Ann Oncol*. 2020;31(6):745-59. doi:10.1016/j.annonc.2020.02.011
 25. Merker JD, Oxnard GR, Compton C, Diehn M, Hurley P, Lazar AJ, *et al.* Circulating tumor DNA analysis in patients with cancer: American Society of Clinical Oncology and College of American Pathologists joint review. *J Clin Oncol*. 2018;36(16):1631-41. doi:10.1200/JCO.2017.76.8671
 26. Mobadersany P, Yousefi S, Amgad M, Gutman DA, Barnholtz-Sloan JS, Velázquez Vega JE, *et al.* Predicting cancer outcomes from histology and genomics using convolutional networks. *Proc Natl Acad Sci U S A*. 2018;115(13):E2970-9. doi:10.1073/pnas.1717139115
 27. Moscow JA, Fojo T, Schilsky RL. The evidence framework for precision cancer medicine. *Nat Rev Clin Oncol*. 2018;15(3):183-92. doi:10.1038/nrclinonc.2017.186
 28. Mouliere F, Chandrananda D, Piskorz AM, Moore EK, Morris J, Ahlborn LB, *et al.* Enhanced detection of circulating tumor DNA by fragment size analysis. *Sci Transl Med*. 2018;10(466):eaat4921. doi:10.1126/scitranslmed.aat4921
 29. Newman AM, Lovejoy AF, Klass DM, Kurtz DM, Chabon JJ, Scherer F, *et al.* Integrated digital error suppression for improved detection of circulating tumor DNA. *Nat Biotechnol*. 2016;34(5):547-55. doi:10.1038/nbt.3520
 30. Nwokocha CR, Peter-Anyebe AC, Ijiga OM. Evaluating FHIR-driven interoperability frameworks for secure system migration and data exchange in U.S. health information networks. *Int J Sci Res Sci Technol*. 2021. doi:10.32628/IJSRST523105135
 31. Nwokocha CR, Peter-Anyebe AC, Ijiga OM. Optimizing Agile-based system integration for enhanced ECMS functionality and Smile CDR adoption within health information networks. *Int J Sci Res Comput Sci Eng Inf Technol*. 2021;7(6):470-90. doi:10.32628/CSEIT2282148
 32. Nwokocha CR, Soetan KT, Peter-Anyebe AC. Automating ETL pipelines with SQL Server Integration Services to improve health data quality and reporting accuracy in national health systems. *Int J Sci Res Sci Technol*. 2022;9(4):804-27.
 33. Oladoye SO, Bamigwojo OV, James AO, Ijiga OM. AI-driven predictive maintenance modeling for high-voltage distribution assets using sensor fusion and time-series degradation analysis. *Int J Sci Res Sci Eng Technol*. 2021;11(2):387-411. doi:10.32628/IJSRSET2291524
 34. Onyekaonwu CB, Peter-Anyebe AC. Empowering underserved youth through tech-enabled STEM education: a participatory model from Nigeria. *Int J Sci Res Sci Technol*. 2019;6(5). doi:10.32628/IJSRST1965990
 35. Onyekaonwu CB, Peter-Anyebe AC, Raphael FO. From prescription to prediction: leveraging AI/ML to improve medication adherence and adverse drug event detection in community pharmacies. *Int J Sci Res Sci Technol*. 2019;6(5):460-76.
 36. Phallen J, Sausen M, Adleff V, Leal A, Hruban C, White J, *et al.* Direct detection of early-stage cancers using circulating tumor DNA. *Sci Transl Med*. 2017;9(403):eaan2415.
 37. Razavi P, Li BT, Brown DN, Jung B, Hubbell E, Shen R, *et al.* High-intensity sequencing reveals the sources of plasma circulating cell-free DNA variants. *Nat Med*. 2019;25(12):1928-37. doi:10.1038/s41591-019-0652-7
 38. Roy S, Coldren C, Karunamurthy A, Kip NS, Klee EW, Lincoln SE, *et al.* Standards and guidelines for validating next-generation sequencing bioinformatics pipelines: a joint recommendation of the Association for Molecular Pathology and the College of American Pathologists. *J Mol Diagn*. 2018;20(1):4-27. doi:10.1016/j.jmoldx.2017.11.003
 39. Shen SY, Singhanian R, Fehringer G, Chakravarthy A, Roehrl MHA, Chadwick D, *et al.* Sensitive tumour detection and classification using plasma cell-free DNA methylomes. *Nature*. 2018;563(7732):579-83. doi:10.1038/s41586-018-0703-0
 40. Siravegna G, Marsoni S, Siena S, Bardelli A. Integrating liquid biopsies into the management of cancer. *Nat Rev Clin Oncol*. 2017;14(9):531-48. doi:10.1038/nrclinonc.2017.14
 41. Stetson D, Ahmed A, Xu X, Nuttall BRB, Lubinski TJ, Johnson JH, *et al.* Orthogonal comparison of four plasma NGS tests with tumor suggests technical factors are a major source of assay discordance. *JCO Precis Oncol*. 2019;3:1-9. doi:10.1200/PO.18.00191
 42. Torga G, Pienta KJ. Patient-paired sample congruence between 2 commercial liquid biopsy tests. *JAMA Oncol*. 2018;4(6):868-70. doi:10.1001/jamaoncol.2017.4027
 43. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, *et al.* Attention is all you need. *Adv Neural Inf Process Syst*. 2017;30.
 44. Wan JCM, Massie C, Garcia-Corbacho J, Mouliere F, Brenton JD, Caldas C, *et al.* Liquid biopsies come of age: towards implementation of circulating tumour DNA. *Nat Rev Cancer*. 2017;17(4):223-38. doi:10.1038/nrc.2017.7
 45. Zehir A, Benayed R, Shah RH, Syed A, Middha S, Kim HR, *et al.* Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med*. 2017;23(6):703-13. doi:10.1038/nm.4333