Investigating the Promise of AI-Assisted Liquid Biopsy for Early Cancer Detection - Ahona Siddique.pdf

by Sanaul Haque

Submission date: 13-Oct-2025 10:52PM (UTC+0700)

Submission ID: 2779951126

File name: Investigating_the_Promise_of_AI-Assisted_Liquid_Biopsy_for_Early_Cancer_Detection_-_Ahona_Siddique.pdf (243.19K)

Word count: 3130

Character count: 19602

Investigating the Promise of AI-Assisted Liquid Biopsy for Early Cancer Detection: A Proposed In-Silico Framework of Computational Diagnostic Strategies

---

Full Legal Name: Upashana Siddique Ahona

Institution: Uttara High School and College

Category: Junior

Grade: 12th

Country: Bangladesh

Submission Date: 30th September 2025

Registered Email Address: ahona.siddique5116@gmail.com

Research Topic: Investigating the Promise of AI-Assisted Liquid Biopsy for Early Cancer Detection: A Proposed In-Silico Framework of Computational Diagnostic Strategies

Investigating the Promise of AI-Assisted Liquid Biopsy for Early Cancer Detection: A Proposed In-Silico Framework of Computational Diagnostic Strategies

Research Aims: This research examines the potential of Artificial Intelligence (AI) to improve early cancer detection using liquid biopsy data. This study particularly aims to:

1. Design a computational framework to distinguish cancer versus normal datasets using public datasets (gene expressions, miRNA, methylation)
2. Compare different biomarkers to identify which provides the most reliable detection
3. Assess AI's performance in early detection by evaluating AI sensitivity at low tumor fractions.
4. Discuss its broader implications in universal cancer screening, including population level challenges

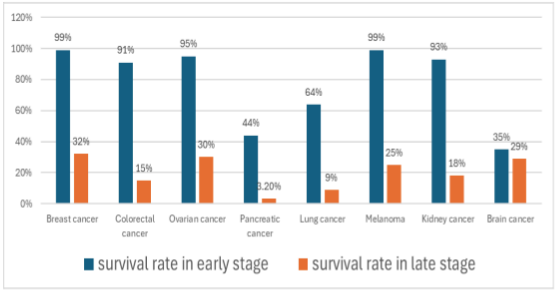
Research Questions:

1. Can supervised AI models distinguish cancer from normal samples with AUROC above 0.9?
2. Which biomarker (methylation, miRNA, gene expression or combined) provides the highest predictive accuracy?
3. How sensitive is AI in detecting cancer when tumor signals are diluted to very low fractions, as in early-stage cancer?
4. What challenges occur when scaling from small datasets to population level screening?

Introduction: Cancer continues to be one of the leading causes of mortality with 9.7 million deaths globally in 2022 according to WHO.[1] Despite decades of research and treatment advancements, traditional diagnostics remain invasive, harsh, and focused on late-stage detection. There is a desperate need for patient-friendly, precise methods of detection and treatment. Liquid biopsy is an emerging non-invasive diagnostic tool, though promising it faces challenges. This study explores the usage of AI in overcoming those limitations.

The crisis of Late Detection: It is one of the greatest challenges of oncology. Many cancers, such as ovarian, pancreatic or brain cancers have vague early symptoms like bloating, back pain or ear disturbances, which are often ignored or misdiagnosed. By the time it is discovered it's already late, treatment options and survival rates drop sharply.

The graph below compares the survival rates in late and early stages across different cancers. [2]-[9]



Limitations of Traditional Biopsy: While conventional biopsy is considered the standard for cancer diagnosis, it has many limitations. It is an invasive procedure that requires tissue extraction, which can cause discomfort and carry risk of complications, and requires access to specialized equipment and expertise. Biopsies are also expensive and time consuming, making them impractical for repeated monitoring or large-scale population screening. [2] Due to these reasons, physicians may hesitate to recommend them unless patients exhibit significant symptoms, potentially delaying diagnosis until the cancer is more advanced. On the contrary, if cancer could be detected through a minimally invasive method, such as a slightly more expensive blood test (liquid biopsy), doctors would be more likely to suggest testing even for subtle symptoms, enabling earlier detection and increasing the likelihood of successful treatment.

The promise of Liquid Biopsy: Liquid biopsy presents a shift in how cancer can be detected and monitored. Unlike traditional biopsies, it offers the possibility of detecting cancer from a simple blood draw by sequencing ctDNA and other biomarkers. [5] This approach enables repeated minimally invasive testing that can track disease progression, treatment response and recurrence in real time. [6] Most importantly, it has the potential to detect cancer at earlier more treatable stages. Its scalability and accessibility make it a stronger candidate for a universal screening test.

Importance of the Research: While liquid biopsy is promising, its application in universal screening remains limited. This study proposes a simulation-based framework (in-silico) to evaluate how AI can improve its performance. The study integrates public genomic datasets with AI models to demonstrate the reduction of false positives and detection of low-level signals. Tying this approach to population-level challenges underscores both the potential and the limitations of liquid biopsy as a universal screening test.

Research Goals: Unlike earlier studies that focus on high tumor fractions or single omics datasets, this study simulates low abundance tumor signals across multiple omics to mimic real world scenarios. By assessing multiple AI models with calibration and interpretability analyses, this provides a more clinically relevant evaluation.

Literature Review: Liquid Biopsy in Cancer Diagnostic

Recent Advances and Current Landscape: Liquid biopsy has emerged as a major breakthrough in cancer diagnosis, offering a noninvasive alternative to tissue biopsy. Recent studies have shown its potential in early detection, monitoring treatment responses and assessing MRDs (minimal residual disease). A new groundbreaking blood test can predict breast cancer recurrence up to three years prior the tumor can be detected in scans, using whole genome sequencing to track 1800 mutations, offering a greater sensitivity than earlier methods. [10] Advancements in multi cancer early detection tests have shown promising results while minimizing false positives and negatives. [11] Recent works of 2024 has made significant progress in understanding how ctDNA tumor fractions (TF) affects both diagnosis accuracy and prognosis. One study showed that low TF can lead to negative liquid biopsy results which are unreliable without tissue biopsy confirmation,

underlining how the absence of ctDNA doesn't always mean disease free. [12] Another study in non-small cell lung cancer found that ctDNA with very low variant allele frequencies (as low as 0.00002%) though challenging to detect, were predictive of relapse and survival outcomes. [13] It is essential to advance strategies to overcome the limitations of recent research.

Biomarkers in Liquid Biopsy: The success of liquid biopsy depends on the detection of various biomarkers present in bodily fluids. Notable biomarkers include circulating tumor DNA (ctDNA), circulating tumor cells (CTCs), extracellular vesicles (EVs), exosomes and microRNA (miRNA). These biomarkers offer valuable insights on the tumor's progression, genetic mutations, and mechanisms of treatment resistance. For instance, ctDNA analysis can detect genetic alterations linked to specific cancer types, thereby aiding the selection of targeted therapy. [17] However, the detection of these biomarkers requires cutting edge technologies and methodologies to ensure accuracy and reliability.

Challenges and Limitations in Prior Research: Despite its potential, liquid biopsy faces several challenges that hamper its widespread clinical use. One critical issue is the false-positive and false-negative results which may lead to misdiagnosis and inappropriate treatment decisions. [14] Another significant issue is the low abundance of certain biomarkers, for example, ctDNA in early-stage cancer, poses difficulties in detection. The heterogeneity of genetic mutations across different types of cancer further complicates the development of universal liquid biopsy tests since each cancer possess distinct molecular signatures in the blood stream. [15] The challenge of isolating and analyzing biomarkers highlights the need for standardized reproducible protocols to ensure reliability.

Implications for Research Design and Technological Integration: The manual analysis of liquid biopsy samples is challenging due to low abundance and minuscule size of biomarkers, increasing the chances of missing critical indicators. [15] Incorporating artificial intelligence in the analysis process has the potential to revolutionize this field. AI algorithms can be trained to detect subtle patterns and anomalies in complex datasets, enhancing the sensitivity and specificity of liquid biopsy assays. Recent studies have highlighted the effectiveness of AI in improving the detection of CTCs, demonstrating its capability to outperform traditional biopsies in case of accuracy and efficiency. [16] However, the success of AI integration depends on the quality and diversity of training datasets, along with continuous optimization of algorithms that enables adaptation to evolving clinical scenarios. [16] This study involves both research and technological application of liquid biopsy. By integrating AI models with multi omics data and evaluating performances in dilution scenarios, this study emphasizes the potential of computational tools in case of enhancing, sensitivity, reliability and interpretability. Rigorous model training, calibration, feature assessment are essential for advancing liquid biopsy into a reliable and scalable screening approach.

Research Methodologies: Given the challenges of low abundance biomarkers and the risk of mislabeling in manual interpretation, this study adopts a mixed method approach that combines

quantitative in-silico experiments with qualitative assessments of model interpretability to ensure more accurate and reliable insights.

Quantitative Phase: This study is conducted through an entirely in-silico approach using publicly available multi omics datasets from The Cancer Genome Atlas (TCGA) and The International Cancer Genome Consortium (ICGC), thereby eliminating the need for patient recruitment. [18] The dataset will consist of tumor samples and their matched normal controls including gene expression, miRNA and DNA methylation profiles. Prior to downstream analysis, data will undergo rigorous preprocessing, which includes label correction, elimination of low variance features, and normalization to ensure reliable comparability across samples and platforms. To prevent data leakage samples will be split at the patient level with 60% allocated for training, 20% for validation and 20% reserved for testing. Feature sets will be structured to enable analysis of single omics datasets as well as integrated multi-omics profiles, allowing for a comprehensive assessment of biomarker contributions. [20]

Qualitative Phase: While this study is primarily computational, a qualitative perspective is included via simulated liquid biopsy scenarios. Tumor signals will be artificially diluted within normal profiles to simulate the characteristics of low-abundance biomarkers in early-stage cancer detection, using fractional admixtures of 1%, 0.5% and 0.1%.[12][15] This strategy evaluates the model's ability to identify subtle molecular signatures under realistic conditions, offering better insights into practical challenges like false negatives due to low tumor fractions. Additionally, interpretive analyses including feature importance via logistic regression coefficients, permutation importance via random forest , and SHAP values [22] for gradient boosting models will be conducted to provide a qualitative understanding of which biomarkers have the most influence on prediction.

Data Analysis: The analytical strategy combines both classical and modern machine learning techniques to evaluate predictive performance across biomarker types and dilution scenarios. [16] Logistic regression will serve as a baseline model, providing interpretable coefficients that directly links features to cancer status. Ensemble techniques including random forests and gradient boosting methods like XGBoost or LightGBM will be applied to capture complex multi omics datasets. Random forests effectively handle complex, non-linear patterns and reduce overfitting, making them ideal for heterogenous multi-omics analyses. Gradient boosting methods are expected to excel in detecting subtle signals from low abundance biomarkers because of their iterative error minimization process. This boosts sensitivity in challenging dilution scenarios. Model performance will be assessed using standard metrics including area under the receiver operating characteristic curve (AUROC), area under the precision-recall curve (AUPRC), sensitivity, specificity and overall accuracy. Calibration curves and Brier scores will be calculated to evaluate the reliability of these probabilistic predictions. [23] This study will compare model performance at low tumor fractions to determine robustness and the potential of AI to enhance sensitivity in challenging detection scenarios.

Project Practicalities: Several practical considerations will guide the execution and interpretation of this study. The size of the dataset is a limiting factor, while TCGA and ICGC provides rich resources, clinical implementation requires substantially larger and diverse cohorts to validate generalizability across populations. The low prevalence of cancer in screening populations emphasizes the importance of high specificity to minimize false positives and the resulting clinical and financial burden. [11] The discussion of translational feasibility will consider economic aspects, sequencing costs, confirmatory imaging and follow-up procedures. [24] Future expansions may involve population level simulations to estimate sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) providing a basis for large-scale implementation. [23] This methodology thus provides a rigorous and reproduceable framework for AI-assisted liquid biopsy in early-stage cancer detection, addressing both computational and translational aspects.

Conclusion: This research in short strategically combines in-silico modelling with interpretability analysis to interrogate the feasibility of AI-assisted liquid biopsy. By integrating multi-omics data from TCGA/ICGC, bench marking models ranging from logistic regression to advanced gradient boosting, and stress-testing performance under dilution scenarios, this study ensures both methodological and translational value. This study not only clarifies which biomarkers and algorithms are most reliable but also accentuates the inherent limitations in small datasets and low abundance signals. Looking forward, this framework lays the foundation for population level simulations, helping bridge the gap between computational insights and clincal application. Ultimately this study points towards an AI-driven future where early cancer detection is more sensitive, reliable and cost-effective.

References:

[1] World Health Organization, "Global Cancer Burden growing, Amidst Mounting Need for Services," WHO, 2024. https://www.who.int/news/item/01-02-2024-global-cancer-burden-growing--amidst-mounting-need-for-services

[2] American Cancer Society, "Survival rates for breast cancer," www.cancer.org, Mar. 01, 2023. https://www.cancer.org/cancer/types/breast-cancer/understanding-a-breast-cancer-diagnosis/breast-cancer-survival-rates.html

[3] American Cancer Society, "Colorectal Cancer Survival Rates | Colorectal Cancer Prognosis," www.cancer.org, Mar. 01, 2023. https://www.cancer.org/cancer/types/colon-rectal-cancer/detection-diagnosis-staging/survival-rates.html

[4] Cancer Research UK, "Survival | Ovarian cancer | Cancer Research UK," www.cancerresearchuk.org, Nov. 28, 2024. https://www.cancerresearchuk.org/about-cancer/ovarian-cancer/survival

[5] American Cancer Society, "Survival Rates for Pancreatic Cancer," www.cancer.org, Feb. 05, 2024. https://www.cancer.org/cancer/types/pancreatic-cancer/detection-diagnosis-staging/survival-rates.html

[6] Cancer Treatment Centers of America, "Cancer Staging," Cancer Treatment Centers of America, Sep. 17, 2019. https://www.cancercenter.com/cancer-types/lung-cancer/stages

[7] R. Oricione, "Over 104,000 Americans Estimated to Be Diagnosed with Invasive Melanoma in 2025," Curemelanoma.org, Feb. 23, 2025. https://www.curemelanoma.org/blog/over-104-000-americans-estimated-to-be-diagnosed-with-invasive-melanoma-in-2025

[8] "Survival Rates for Kidney Cancer," www.cancer.org. https://www.cancer.org/cancer/types/kidney-cancer/detection-diagnosis-staging/survival-rates.html

[9] National Cancer Institute, "Cancer of the Brain and Other Nervous System - Cancer Stat Facts," SEER, 2018. https://seer.cancer.gov/statfacts/html/brain.html

[10] A. Gregory, "Predictive blood test hailed as 'incredibly exciting' breast cancer breakthrough," the Guardian, Jun. 2024. https://www.theguardian.com/society/article/2024/jun/02/predictive-blood-test-hailed-as-incredibly-exciting-breast-cancer-breakthrough

[11] N. Walter, J. Groth, and B. von und zu Zwerger, "Evaluation of an innovative multi-cancer early detection test: high sensitivity and specificity in differentiating cancer, inflammatory

conditions, and healthy individuals," Frontiers in Oncology, vol. 15, Mar. 2025, doi: https://doi.org/10.3389/fonc.2025.1520869.

[12] D. Rolfo et al., "Measurement of ctDNA Tumor Fraction Identifies Informative Negative Liquid Biopsy Results and Informs Value of Tissue Confirmation," Clinical Cancer Research, vol. 30, no. 11, pp. 2452–2460, Mar. 2024, doi: https://doi.org/10.1158/1078-0432.ccr-23-3321.

[13] N. Galant, Marcin Nicoś, B. Kuźnar-Kamińska, and P. Krawczyk, "Variant Allele Frequency Analysis of Circulating Tumor DNA as a Promising Tool in Assessing the Effectiveness of Treatment in Non-Small Cell Lung Carcinoma Patients," Cancers, vol. 16, no. 4, pp. 782–782, Feb. 2024, doi: https://doi.org/10.3390/cancers16040782.

[14] Y. Xu et al., "Liquid biopsy-based multi-cancer early detection: an exploration road from evidence to implementation," Science Bulletin, Jun. 2025, doi: https://doi.org/10.1016/j.scib.2025.06.030.

[15] S. Zafar et al., "Emerging biomarkers for early cancer detection and diagnosis: challenges, innovations, and clinical perspectives," European journal of medical research, vol. 30, no. 1, Aug. 2025, doi: https://doi.org/10.1186/s40001-025-03003-6.

[16] A. Tiwari, S. Mishra, and T.-R. Kuo, "Current AI technologies in cancer diagnostics and treatment," Molecular Cancer, vol. 24, no. 1, Jun. 2025, doi: https://doi.org/10.1186/s12943-025-02369-9.

[17] Y. Nakamura et al., "Targeted therapy guided by circulating tumor DNA analysis in advanced gastrointestinal tumors," Nature Medicine, Sep. 2024, doi: https://doi.org/10.1038/s41591-024-03244-8.

[18] K. Chaudhary, O. B. Poirion, L. Lu, and L. X. Garmire, "Deep Learning–Based Multi-Omics Integration Robustly Predicts Survival in Liver Cancer," Clinical Cancer Research, vol. 24, no. 6, pp. 1248–1259, Mar. 2018, doi: https://doi.org/10.1158/1078-0432.ccr-17-0853.

[19] B. Ryan, R. E. Marioni, and T. I. Simpson, "Multi-Omic Graph Diagnosis (MOGDx): A data integration tool to perform classification tasks for heterogeneous diseases," Bioinformatics, vol. 40, no. 9, Aug. 2024, doi: https://doi.org/10.1093/bioinformatics/btae523.

[20] S. Khadirnaikar, S. Shukla, and S. R. M. Prasanna, "Machine learning based combination of multi-omics data for subgroup identification in non-small cell lung cancer," Scientific Reports, vol. 13, no. 1, Mar. 2023, doi: https://doi.org/10.1038/s41598-023-31426-w.

[21] W. Zhang, H. Huang, L. Wang, B. D. Lehmann, and S. X. Chen, "An Integrative Multi-Omics Random Forest framework for Robust Biomarker Discovery," bioRxiv : the preprint server for biology, p. 2025.03.05.641533, Jun. 2025, doi: https://doi.org/10.1101/2025.03.05.641533.

[22] A. V. Ponce-Bobadilla, V. Schmitt, C. S. Maier, S. Mensing, and S. Stodtmann, "Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development," Clinical and Translational Science, vol. 17, no. 11, Oct. 2024, doi: https://doi.org/10.1111/cts.70056.

[23] G. S. Collins et al., "Evaluation of clinical prediction models (part 1): from development to external validation," BMJ, vol. 384, p. e074819, Jan. 2024, doi: https://doi.org/10.1136/bmj-2023-074819.

[24] M. J. Duffy and J. Crown, "A Personalized Approach to Cancer Treatment: How Biomarkers Can Help," Clinical Chemistry, vol. 54, no. 11, pp. 1770–1779, Nov. 2008, doi: https://doi.org/10.1373/clinchem.2008.110056.

Investigating the Promise of AI-Assisted Liquid Biopsy for Early Cancer Detection - Ahona Siddique.pdf

ORIGINALITY REPORT

20%	17%	16%	0%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.frontiersin.org Internet Source	2%
2	James Henry. "Population Health Management, Fit Lifecycles in Future Analytics", Open Science Framework, 2025 Publication	1%
3	www.medtecheurope.org Internet Source	1%
4	www.medrxiv.org Internet Source	1%
5	etasr.com Internet Source	1%
6	www.biorxiv.org Internet Source	1%
7	Borhan Mostafizi, Hamid RudsariKhosfekr, Ehsan Irajizad. "LOWERING CANCER-SPECIFIC MORTALITY RATES USING PROTEIN BIOMARKERS", Cold Spring Harbor Laboratory, 2025 Publication	1%
8	www.liveinhomecare.com Internet Source	1%
9	arxiv.org Internet Source	1%
10	trm-acceptance.guardanthealth.com Internet Source	1%

11	Ravi Verma. "Microplastics: Environmental Ubiquity, Biological Fate, and Human Health Implications", International Journal of Sciences and Innovation Engineering, 2025 Publication	1 %
12	molecular-cancer.biomedcentral.com Internet Source	1 %
13	Huei-Ying Li, Chun-Chuan Chang, Yu-Hsuan Yang, Chi-Yuan Yao, Jason Chia-Hsun Hsieh, Shao-Hsuan Chang. "The Prospectives and Limitations of Liquid Biopsy Utilization for Clinical Practice in Taiwan", The Journal of Liquid Biopsy, 2025 Publication	<1 %
14	medical.olympusamerica.com Internet Source	<1 %
15	oaktrust.library.tamu.edu Internet Source	<1 %
16	journals.lww.com Internet Source	<1 %
17	mdpi-res.com Internet Source	<1 %
18	www.proteinatlas.org Internet Source	<1 %
19	San San Amelia Tai, Hooi Leong Loo, Athirah Bakhtiar, Paul Chi-Lui Ho, Lay Hong Chuah. "Antibody-conjugated polymer nanoparticles for brain cancer", Drug Delivery and Translational Research, 2025 Publication	<1 %
20	www.curemelanoma.org Internet Source	<1 %
	www.pfizer.com	

21	Internet Source	<1 %
22	aait.op.edu.ua Internet Source	<1 %
23	pure.fujita-hu.ac.jp Internet Source	<1 %
24	www.coursehero.com Internet Source	<1 %
25	www.nature.com Internet Source	<1 %
26	Nicola Fusco, Konstantinos Venetis, Francesco Pepe, Omshree Shetty et al. "International society of liquid biopsy (ISLB) perspective on minimal requirements for ctDNA testing in solid tumors", The Journal of Liquid Biopsy, 2025 Publication	<1 %
27	breast-cancer-research.biomedcentral.com Internet Source	<1 %
28	jeasd.uomustansiriyah.edu.iq Internet Source	<1 %
29	rima.ufrrj.br Internet Source	<1 %
30	www.aol.co.uk Internet Source	<1 %
31	www.auctoresonline.org Internet Source	<1 %
32	clinchem.aaccjnls.org Internet Source	<1 %
33	doctorpenguin.com Internet Source	<1 %

34	Douglas M. Ruden. "Multi-Biofluid Approaches for cftDNA and cftRNA Biomarker Detection: Advances in Early Cancer Detection and Monitoring", Current Issues in Molecular Biology, 2025 Publication	<1 %
35	Tilottama Roy, Subhajit Dutta, Sonali Paul. "Herbal Wealth of the Asia-Pacific - Genomics, Medicinal Uses, and Conservation", CRC Press, 2025 Publication	<1 %
36	knowledgecommons.lakeheadu.ca Internet Source	<1 %
37	sciendo.com Internet Source	<1 %
38	www.mdpi.com Internet Source	<1 %
39	Pushpa Choudhary, Sambit Satpathy, Arvind Dagur, Dhirendra Kumar Shukla. "Recent Trends in Intelligent Computing and Communication", CRC Press, 2025 Publication	<1 %
40	Roberto Borea, Erick F. Saldanha, Shivahamy Maheswaran, Eleonora Nicolo et al. "Cancer in a drop: Advances in liquid biopsy in 2024", Critical Reviews in Oncology/Hematology, 2025 Publication	<1 %
41	Roxane Axel Jacob, Leo Gaskin, Thomas Seidel, Ya Chen, Angelica Mazzolari, Johannes Kirchmair. "FAME 3R: A Fast, Compact, Flexible, and Practical Re-Design of the FAME 3 Model for Predicting Sites of Metabolism", American Chemical Society (ACS), 2025 Publication	<1 %

42	Sameen Zafar, Amna Hafeez, Hania Shah, Iqra Mutiullah et al. "Emerging biomarkers for early cancer detection and diagnosis: challenges, innovations, and clinical perspectives", European Journal of Medical Research, 2025 Publication	<1 %
43	air.unimi.it Internet Source	<1 %
44	eurjmedres.biomedcentral.com Internet Source	<1 %

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	Off		