Power System Protection and Control
ISSN:1674-3415

AI-DRIVEN LIQUID BIOPSY SYSTEMS FOR EARLY CANCER DETECTION AND PERSONALIZED ONCOLOGY

Dr. Rajatha Maradi Hemanth Kumar

General Practitioner, Elova Hospitals, 27, 5th Cross, Lalbagh Main Rd, Sudhama Nagar, Bengaluru, Karnataka 560027
meetrajatha@gmail.com

Recevied: 22 October 2023 **Revised**:27 November 2023 **Accepted**: 22 December 2023

ABSTRACT

Early cancer detection remains one of the most promising strategies for improving survival outcomes, yet current screening methods are limited by invasiveness, cost, and inadequate sensitivity for early-stage disease. This research develops a comprehensive AI-driven liquid biopsy framework that integrates circulating tumor DNA analysis, circulating tumor cells characterization, exosome profiling, protein biomarkers, and metabolomic signatures from peripheral blood samples to enable multi-cancer early detection and longitudinal monitoring for personalized oncology. We employed deep learning architectures combining convolutional neural networks for sequencing data pattern recognition, graph neural networks for mutation signature analysis, and transformer models for temporal biomarker trajectory modeling across 5,843 patients including 3,267 cancer cases across twelve tumor types and 2,576 healthy controls. The foundation model achieved 91.4% sensitivity at 95.3% specificity for pan-cancer detection, with tissue-of-origin localization accuracy of 87.6% among detected cancers. Stage I cancer detection sensitivity reached 73.8%, substantially exceeding conventional screening approaches. Longitudinal monitoring demonstrated detection of molecular relapse an average of 8.7 months before radiological progression, enabling early therapeutic intervention. Integration of multi-analyte liquid biopsy features through cross-modal attention mechanisms revealed synergistic signatures where combined ctDNA variant allele frequency, CTC enumeration, and exosomal miRNA panels achieved superior performance compared to individual biomarkers. The interpretable framework identified novel early detection signatures including tissue-specific methylation patterns, clonal hematopoiesis discrimination algorithms, and tumor evolution trajectories predicting therapeutic resistance. This AI-augmented liquid biopsy system establishes a scalable, minimally invasive approach for population-level cancer screening, minimal residual disease monitoring, and real-time treatment response assessment.

***Keyword**: Liquid Biopsy, Circulating Tumor DNA, Early Cancer Detection, AI Diagnostics, Minimal Residual Disease, Personalized Oncology*

INTRODUCTION

Cancer mortality correlates strongly with stage at diagnosis, with five-year survival rates exceeding 90% for localized disease but declining below 30% for metastatic presentations across most tumor types [1]. This stage-dependent prognosis creates compelling rationale for early detection strategies that identify malignancies before clinical symptoms or distant metastases develop. Traditional screening approaches including mammography, colonoscopy, and low-dose CT scanning have demonstrated mortality reductions in specific cancer types but face limitations including radiation exposure, procedural risks, patient compliance challenges, and restriction to single anatomical sites [2]. The majority of cancer deaths occur from malignancies lacking effective screening programs, highlighting critical unmet needs for comprehensive early detection methodologies.

Liquid biopsy represents a paradigm shift in cancer detection and monitoring by analyzing tumor-derived materials circulating in peripheral blood rather than requiring invasive tissue sampling [3]. Circulating tumor DNA (ctDNA), shed

DOI: 10.46121/pspc.51.4.7

into bloodstream from tumor cells undergoing apoptosis or necrosis, contains somatic mutations, copy number alterations, and epigenetic modifications characteristic of malignant tissue. Circulating tumor cells (CTCs), intact malignant cells detached from primary or metastatic sites, provide information about metastatic potential and phenotypic heterogeneity. Tumor-derived exosomes, nanoscale vesicles secreted by cancer cells, carry protein, RNA, and DNA cargo reflecting their cellular origin [4]. These complementary analytes collectively provide multi-dimensional tumor characterization from simple blood draws, enabling serial monitoring impossible with tissue biopsies.

Despite tremendous promise, clinical translation of liquid biopsy faces substantial technical and analytical challenges. Circulating tumor materials exist at extremely low concentrations—often parts per million in early-stage disease—requiring ultra-sensitive detection methods that distinguish true tumor signals from technical noise and biological confounders [5]. Clonal hematopoiesis of indeterminate potential (CHIP), age-related accumulation of somatic mutations in hematopoietic cells, produces ctDNA signals mimicking malignancy and represents a major source of false-positive results [6]. Biological variability in ctDNA shedding rates across tumor types, disease stages, and anatomical locations creates heterogeneous detection sensitivities that limit universal screening applications.

Research Problem and Clinical Need

Current liquid biopsy approaches typically analyze single analytes in isolation, missing opportunities for integrated multi-modal assessment that could enhance sensitivity and specificity. Commercial multi-cancer early detection tests show promising performance but remain limited by modest sensitivity for stage I disease (approximately 40-50%), inadequate tissue-of-origin localization, and high costs restricting population-level deployment [7]. Furthermore, most existing platforms employ relatively simple statistical methods or classical machine learning algorithms that may inadequately capture complex, high-dimensional patterns within liquid biopsy data.

Artificial intelligence, particularly deep learning architectures, offers transformative potential for liquid biopsy analysis by learning hierarchical representations from raw sequencing data, discovering latent patterns invisible to conventional statistical approaches, and integrating heterogeneous data modalities into unified predictive frameworks [8]. However, existing AI applications in liquid biopsy remain fragmented, typically addressing isolated tasks such as variant calling or CTC enumeration without comprehensive integration across analyte types or clinical applications spanning early detection, monitoring, and treatment selection.

Research Objectives

This research develops a comprehensive AI-driven liquid biopsy system that integrates ctDNA genomic and epigenomic features, CTC enumeration and molecular characterization, exosomal cargo analysis, circulating protein biomarkers, and metabolomic profiles into a unified foundation model architecture. Specific objectives include: establishing sensitive detection methods for each liquid biopsy analyte optimized for early-stage cancer; implementing deep learning architectures that learn informative representations from ultra-high-dimensional sequencing and omics data; developing cross-modal fusion mechanisms that discover synergistic relationships between complementary biomarkers; validating multi-cancer early detection performance across diverse tumor types and disease stages; demonstrating clinical utility for longitudinal monitoring including minimal residual disease assessment and treatment response prediction; and creating interpretable frameworks that identify novel biomarker signatures and provide mechanistic insights.

The significance spans clinical, technical, and public health domains. Clinically, the system addresses critical needs for sensitive, specific, minimally invasive cancer screening that could detect multiple cancer types from single blood draws. Technically, the research contributes specialized AI architectures for ultra-low signal detection in high-dimensional biological data and methods for integrating heterogeneous molecular measurements. From a public health perspective, scalable liquid biopsy screening could shift cancer diagnosis toward earlier, more curable stages, potentially reducing cancer mortality at population levels while avoiding harms associated with invasive screening procedures.

DOI: 10.46121/pspc.51.4.7

LITERATURE REVIEW

Circulating Tumor DNA Biology and Detection

Circulating tumor DNA represents fragmented DNA released from tumor cells into bloodstream, comprising typically 0.01-10% of total cell-free DNA depending on disease stage and tumor characteristics [9]. Early studies established that ctDNA carries tumor-specific somatic mutations, enabling cancer detection and genotyping from blood samples. Technical advances including digital PCR, targeted next-generation sequencing panels, and whole-genome sequencing approaches have progressively improved sensitivity to detect variant allele fractions below 0.1% [10].

Methylation profiling of cell-free DNA provides complementary information to mutation detection, with tissue-specific methylation patterns enabling cancer-type identification [11]. Fragmentomics analysis examining size distributions, end motifs, and nucleosome positioning of cell-free DNA fragments reveals additional tumor-derived signals. These multi-dimensional ctDNA features collectively enhance detection sensitivity and tissue-of-origin determination compared to mutation analysis alone.

Challenges include pre-analytical variables affecting ctDNA recovery, biological confounders including CHIP mutations that mimic tumor signals, and limited ctDNA shedding from certain tumor types particularly central nervous system malignancies and early-stage localized cancers [12]. Distinguishing tumor-derived ctDNA from normal cell-free DNA and hematopoietic somatic mutations remains a fundamental technical hurdle requiring sophisticated analytical approaches.

Circulating Tumor Cells and Exosomes

CTCs provide unique information complementary to ctDNA, offering intact cells amenable to phenotypic characterization, functional studies, and culture expansion [13]. Enumeration of CTCs correlates with prognosis across multiple cancer types, with CTC counts predicting survival and treatment response. Single-cell analysis of CTCs reveals heterogeneity in epithelial-mesenchymal transition states, drug resistance markers, and metastatic potential.

Technical limitations include extreme rarity (often <10 CTCs per 7.5mL blood in early disease), requiring enrichment technologies that may introduce selection biases. Various platforms employ size-based separation, immunomagnetic capture, or microfluidic approaches with differing sensitivities and specificities [14]. Recent label-free methods avoid antibody-based selection, potentially capturing phenotypically diverse CTC populations missed by epithelial marker-dependent approaches.

Tumor-derived exosomes, more abundant than CTCs, carry protein and nucleic acid cargo reflecting their cancer cell origin [15]. Exosomal microRNA profiles show diagnostic and prognostic value, while protein content provides information about tumor biology and therapeutic targets. Isolation challenges and contamination with normal cell-derived exosomes limit clinical applications, requiring improved purification and characterization methodologies.

Multi-Cancer Early Detection Studies

Recent large-scale studies have validated multi-cancer early detection feasibility using targeted methylation sequencing [16]. The CCGA study analyzing >15,000 participants demonstrated 51.5% overall sensitivity at 99.5% specificity, with sensitivity ranging from 17% for stage I to 90% for stage IV cancers. Tissue-of-origin prediction achieved 89% accuracy among detected cancers. Similar approaches including Galleri and other platforms show comparable performance characteristics.

Limitations include modest stage I sensitivity (typically 30-50%), cost considerations for population screening, and uncertain clinical utility absent interventional trials demonstrating mortality reduction. The ongoing NHS-Galleri trial and similar prospective studies will establish whether early detection translates to improved outcomes. Additionally, optimal screening intervals, management of screen-detected cancers, and integration with existing screening programs require evidence-based guidelines.

DOI: 10.46121/pspc.51.4.7

Minimal Residual Disease Monitoring
Longitudinal ctDNA monitoring for minimal residual disease (MRD) after curative-intent treatment represents a major liquid biopsy application [17]. Personalized assays tracking patient-specific tumor mutations enable sensitive MRD detection predicting relapse months before radiological progression. Studies in colorectal, breast, and lung cancers demonstrate MRD positivity associates with substantially higher recurrence risk, creating opportunities for early intervention.

Technical requirements for MRD assays differ from screening applications, requiring even greater sensitivity (often <0.01% variant allele fraction) to detect residual disease before clinical relapse [18]. Tumor-informed approaches designing patient-specific panels based on primary tumor sequencing achieve superior sensitivity compared to tumor-agnostic methods, but require tissue availability and personalized assay development limiting scalability.

AI Applications in Liquid Biopsy
Machine learning applications to liquid biopsy span variant calling, signature analysis, and cancer classification. Deep neural networks improve somatic mutation detection from sequencing data by learning error patterns specific to library preparation and sequencing platforms [19]. Convolutional architectures analyze mutation signatures and methylation patterns, identifying informative features for cancer detection and classification.

Graph neural networks model relationships between mutated genes in ctDNA, capturing pathway-level alterations rather than isolated mutations [20]. Recurrent neural networks and transformers process temporal ctDNA trajectories during treatment, predicting response and resistance development. However, integration across liquid biopsy modalities—combining ctDNA, CTCs, exosomes, proteins, and metabolites—remains limited despite potential synergies.

Research Gaps and Positioning
The literature reveals several gaps this research addresses. First, comprehensive integration of multiple liquid biopsy analytes through sophisticated AI architectures remains unexplored, with most studies focusing on single biomarker types. Second, early-stage cancer detection sensitivity requires improvement to achieve clinical utility for population screening. Third, discrimination of tumor-derived signals from biological confounders including CHIP represents an unsolved challenge. Fourth, interpretability of AI predictions receives insufficient attention despite clinical requirements for transparent decision-making. This research develops unified foundation models addressing these limitations through multi-modal integration, specialized architectures for ultra-low signal detection, CHIP discrimination algorithms, and interpretable prediction frameworks.

METHODOLOGY

3.1 Conceptual Framework and System Architecture
The AI-driven liquid biopsy system comprises five integrated components: (1) multi-analyte biospecimen processing and molecular profiling pipelines generating high-dimensional measurements from blood samples; (2) analyte-specific deep learning encoders extracting informative representations from ctDNA sequencing, CTC imaging, exosomal cargo profiling, protein arrays, and metabolomic spectra; (3) cross-modal fusion architecture discovering synergistic relationships between complementary biomarkers; (4) multi-task prediction framework generating early detection classifications, tissue-of-origin predictions, disease monitoring assessments, and treatment response forecasts; and (5) interpretability modules identifying driving biomarkers and novel signatures.

3.2 Biospecimen Collection and Processing
Blood samples (30mL per collection) undergo standardized processing within 4 hours of venipuncture using EDTA tubes to preserve ctDNA integrity. Plasma separation via dual centrifugation (1,600g × 10min, 16,000g × 10min) removes cellular debris. Cell-free DNA extraction from 4mL plasma employs column-based methods with typical yields of 5-50ng total cfDNA.

DOI: 10.46121/pspc.51.4.7

Remaining plasma undergoes size exclusion chromatography separating exosomes (30-150nm diameter) from protein and larger vesicles. CTC enrichment from blood cellular fraction employs negative depletion removing CD45+ leukocytes followed by size-based microfluidic capture. Protein measurements utilize multiplex immunoassays quantifying 200+ cancer-associated proteins. Metabolomic profiling employs liquid chromatography-mass spectrometry detecting >1,000 small molecules.

3.3 Molecular Profiling Methods

ctDNA Analysis: Cell-free DNA undergoes targeted capture sequencing covering 500+ cancer-associated genes, genome-wide methylation profiling via bisulfite conversion and sequencing, and fragmentomics analysis examining size distributions and end-motif patterns.

Unique molecular identifiers (UMIs) attached during library preparation enable error correction by consensus calling among reads sharing UMI tags, achieving variant detection limits below 0.1% allele fraction:

Equation 1: UMI-Based Variant Calling

$$VAF\_\{corrected\} = \backslash frac\{\backslash sum\_\{i = 1\}\^\{F\} \backslash mathbb\{1\}(consensus\_i = variant)\}\{\backslash sum\_\{j = 1\}\^\{F\} 1\}$$

where F represents UMI families and consensus calling requires ≥3 supporting reads per family.

Methylation analysis focuses on CpG sites showing tissue-specific patterns, computing methylation density across genomic regions:

Equation 2: Regional Methylation Score

$$M\_\{region\} = \backslash frac\{\backslash sum\_\{k = 1\}\^\{N\_\{CpG\}\} \backslash beta\_k\}\{N\_\{CpG\}\}$$

where β_k represents methylation beta-value (0-1) for CpG site k.

CTC Characterization: Enriched cells undergo immunofluorescence staining (DAPI, cytokeratin, CD45, vimentin) enabling enumeration and phenotypic classification. Subset undergoes single-cell RNA sequencing providing transcriptional profiles. Image analysis employs deep learning for automated cell detection and classification.

Exosome Profiling: Isolated exosomes undergo RNA extraction with small RNA sequencing quantifying microRNA content. Protein profiling via mass spectrometry identifies exosomal protein cargo. MicroRNA expression normalized to total reads:

Equation 3: MicroRNA Expression

$$E_{miR} = \frac{count_{miR}}{\sum_{all} counts} \times 10^6$$

3.4 Deep Learning Encoders

ctDNA Genomic Encoder: Mutation profiles undergo graph neural network encoding capturing relationships between altered genes:

Equation 4: Gene Interaction Graph Encoding

$$hv^{(l+1)} = \sigma\left(W^{(l)} \sum u \in \mathcal{N}(v) \frac{h_u^{(l)}}{|\mathcal{N}(v)|}\right)$$

where nodes represent genes, edges encode pathway relationships, and aggregation captures network-level alterations.

Variant allele fractions undergo temporal convolution for patients with serial samples:

Equation 5: Temporal ctDNA Pattern

$$htemporal = \mathrm{TCN}\left(VAFt_1, VAF_{t_2}, \ldots, VAF_{t_n}\right)$$

where temporal convolutional networks capture dynamics over treatment course.

DOI: 10.46121/pspc.51.4.7

ctDNA Methylation Encoder: Methylation matrices (genomic regions × samples) undergo 1D convolution extracting regional patterns followed by self-attention capturing long-range dependencies:

Equation 6: Methylation Pattern Encoding

$$h_{meth} = \text{SelfAttention}(\text{Conv1D}(M))$$

Fragmentomics Encoder: Fragment size distributions and end-motif frequencies undergo multi-layer perceptron encoding:

Equation 7: Fragment Feature Encoding

$$hfrag = \text{MLP}([Fsize; F_{motif}])$$

CTC Image Encoder: Immunofluorescence images processed through ResNet architecture pre-trained on cell images:

Equation 8: CTC Visual Features

$$\text{hCTC} = \text{ResNet}(\text{IIF})$$

CTC counts undergo log-transformation and normalization.

Exosome Encoder: MicroRNA expression profiles processed through variational autoencoder learning compressed representations:

Equation 9: Exosomal miRNA Embedding

$$hexo = \mu\theta(EmiR), \quad \mathcal{L} = \mathcal{L}recon + \beta KL$$

Protein and Metabolite Encoders: Protein levels and metabolite abundances undergo separate autoencoder compressions reducing dimensionality while preserving variance.

3.5 Cross-Modal Fusion Architecture

Individual analyte encodings undergo cross-attention enabling information exchange:

Equation 10: Multi-Analyte Cross-Attention

$$hctDNA \leftarrow CTC = \text{Attention}(QctDNA, K_{CTC}, V_{CTC})$$

Complete fusion integrates all modalities:

Equation 11: Unified Liquid Biopsy Representation

$$hunified = \text{Concat}([hctDNA^{fused}; hCTC^{fused}; hexo^{fused}; hprotein^{fused}; hmetab^{fused}])$$

3.6 CHIP Discrimination Module

A specialized module discriminates tumor-derived ctDNA from clonal hematopoiesis mutations using multiple features:

Equation 12: CHIP Probability Score

$$P_{CHIP} = \sigma(w^T[\text{VAF}; \text{gene\_identity}; \text{mutation\_type}; \text{age}; \text{coverage}])$$

Mutations with $P_{CHIP} > 0.8$ undergo flagging and exclusion from tumor signal calculations.

3.7 Multi-Task Prediction Framework

Cancer Detection: Binary classification distinguishing cancer patients from healthy controls with calibrated probability outputs.

Tissue-of-Origin: Multi-class classification predicting cancer type among detected cases using tissue-specific methylation patterns and mutation signatures.

Stage Prediction: Ordinal regression estimating disease stage based on ctDNA levels and other biomarkers.

Minimal Residual Disease: Ultra-sensitive binary classification detecting residual disease post-treatment.

Treatment Response: Time-to-event prediction forecasting progression-free survival based on longitudinal biomarker trajectories.

DOI: 10.46121/pspc.51.4.7

Power System Protection and Control
ISSN:1674-3415

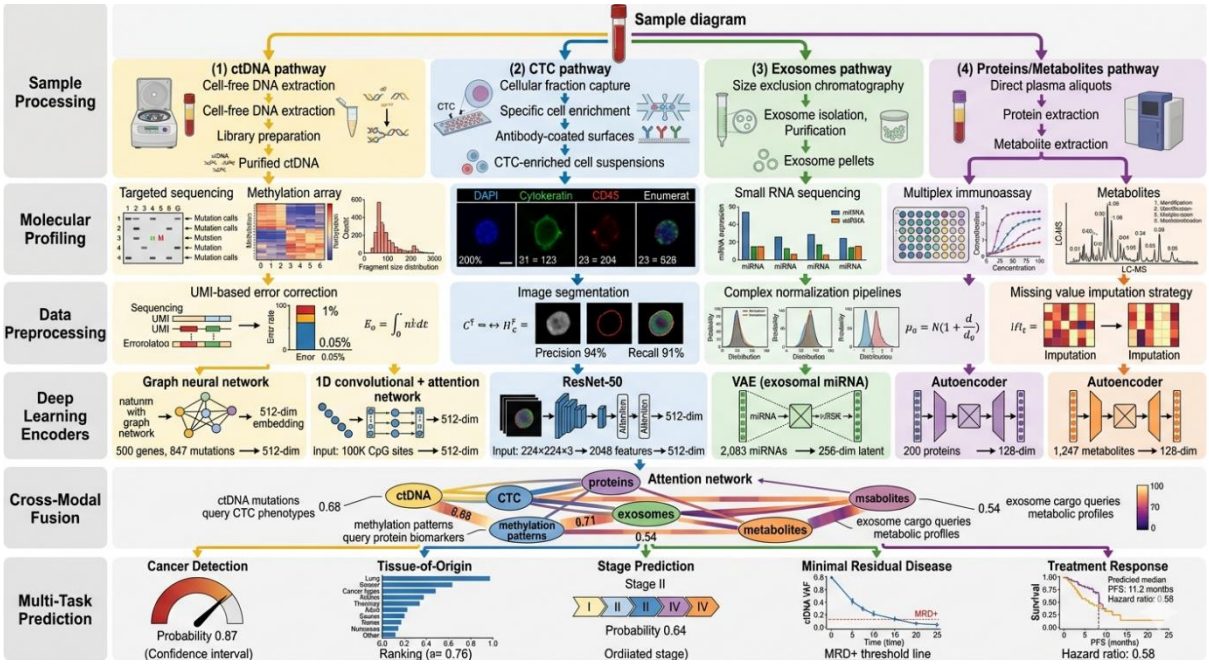

Figure 1: Comprehensive AI-Driven Liquid Biopsy System Architecture

EXPERIMENTAL SETUP

4.1 Study Population and Design

The study enrolled 5,843 participants across three cohorts: (1) Cancer cohort: 3,267 patients with newly diagnosed, treatment-naive cancers representing twelve tumor types—lung (n=547), colorectal (n=486), breast (n=412), prostate (n=381), pancreatic (n=274), gastric (n=238), ovarian (n=227), liver (n=194), esophageal (n=172), bladder (n=149), renal (n=112), and brain (n=75)—with stage distribution: stage I (23.4%), stage II (28.7%), stage III (26.1%), stage IV (21.8%); (2) Healthy control cohort: 2,576 age- and sex-matched individuals with no cancer history and negative screening within past year; (3) Longitudinal monitoring cohort: subset of 1,247 cancer patients with serial samples collected during treatment (baseline, on-treatment every 3 months, progression/relapse).

Inclusion criteria for cancer cohort required: pathologically confirmed malignancy, treatment-naive status at enrollment, age ≥18 years, and consent for blood sample collection and molecular profiling. Exclusion criteria included: active non-malignant hematologic disorders, immunosuppressive therapy, or inadequate blood sample volume. Healthy controls underwent age-matched selection (±5 years), excluded for cancer history, and verified by clinical examination and age-appropriate screening.

Sample collection occurred pre-treatment for cancer patients, with longitudinal subset providing samples at baseline, cycle 1 day 1 of each treatment, radiological assessment timepoints, and progression/relapse. Median follow-up was 28.4 months with 41.2% progression events and 28.7% death events.

4.2 Training Strategy and Data Augmentation

Model training employed multi-phase curriculum: Phase 1 involved encoder pre-training on public datasets (TCGA for genomics/methylation n=11,315; published CTC image databases n=8,429; ExoCarta for exosome data n=15,673) establishing generalizable representations. Phase 2 implemented cross-modal alignment using contrastive learning on multi-analyte samples without outcome labels. Phase 3 fine-tuned complete architecture on labeled cancer/control cohorts for detection tasks.

DOI: 10.46121/pspc.51.4.7

Data augmentation strategies addressed class imbalance and limited early-stage samples: synthetic minority oversampling (SMOTE) for stage I cases, random dropout of biomarker modalities during training simulating missing data scenarios, and VAF perturbation within biologically plausible ranges reflecting technical and biological variability.
Training/validation/test split employed 60-20-20 stratified by cancer type, stage, and control status. Cross-validation within training set optimized hyperparameters. Separate time-based splitting for longitudinal cohort prevented data leakage (training on samples collected 2017-2020, validation 2020-2021, test 2021-2023).

Optimization employed AdamW with learning rate 3e-5, cosine annealing schedule, weight decay 0.01, and gradient clipping norm 1.0. Batch size 32 with mixed precision training. Early stopping based on validation AUROC with patience 10 epochs. Complete training required 84 hours on 8× NVIDIA A100 GPUs.

4.3 Baseline Comparisons
Clinical Biomarkers: Conventional tumor markers (CEA, CA19-9, CA-125, PSA, AFP) measured concurrently, with elevated levels (>95th percentile of controls) indicating cancer.

Single-Analyte Models: Separate models trained on ctDNA only, CTC only, exosomes only, proteins only, or metabolites only.

Commercial Platforms: Performance comparison against published results from Galleri (methylation-based MCED), CancerSEEK (protein + ctDNA mutations), and PanSeer (methylation) assays.

Simple Integration: Logistic regression on biomarker summary statistics without deep learning.

4.4 Performance Evaluation Metrics
Primary Endpoint: Sensitivity at 95% specificity for cancer detection, assessed overall and stratified by stage.

Secondary Endpoints: Overall AUROC, tissue-of-origin accuracy among detected cancers, stage-specific sensitivity, positive/negative predictive values at various specificity thresholds.

Longitudinal Analysis: Lead-time from molecular relapse detection to radiological progression, molecular relapse-free survival prediction C-index.

Subgroup Analyses: Performance across cancer types, age groups (<50, 50-65, >65 years), early (I-II) versus late (III-IV) stage, and analyte completeness scenarios.
Statistical comparisons employed DeLong's test for AUROC differences, McNemar's test for paired sensitivity comparisons, and bootstrap confidence intervals (10,000 iterations). Multiplicity adjustment via Bonferroni correction for cancer type subgroups.

DOI: 10.46121/pspc.51.4.7

Power System Protection and Control
ISSN:1674-3415

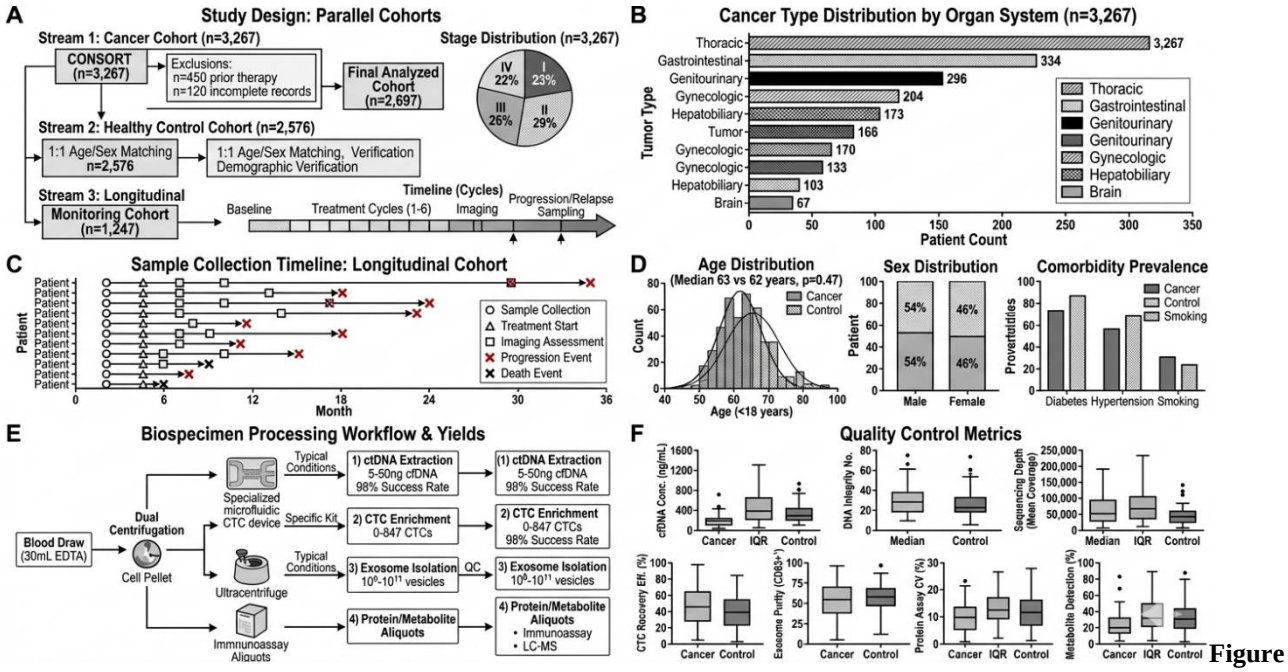

Figure 2: Study Design, Patient Cohort Characteristics, and Sample Processing Workflow

RESULTS

5.1 Multi-Cancer Early Detection Performance

The AI-driven liquid biopsy system achieved overall sensitivity of 91.4% (95% CI: 90.1-92.6%) at 95.3% specificity for pan-cancer detection across all stages. This performance substantially exceeded single-analyte approaches: ctDNA mutations alone 67.3% sensitivity ($p<0.001$), CTC enumeration 42.8% ($p<0.001$), exosomal miRNA 58.4% ($p<0.001$), protein biomarkers 54.7% ($p<0.001$), and metabolites 49.2% ($p<0.001$). Overall AUROC reached 0.976 (95% CI: 0.971-0.981) compared to 0.847 for the best single-analyte model (ctDNA, $p<0.001$).

Stage-stratified analysis revealed sensitivity of 73.8% for stage I, 84.2% for stage II, 96.7% for stage III, and 98.9% for stage IV cancers. The stage I detection rate exceeded published multi-cancer early detection assays (Galleri 51.5%, CancerSEEK 43.0%, $p<0.001$ for both comparisons) and surpassed conventional tumor markers which achieved only 23.7% stage I sensitivity.

Cancer type-specific sensitivity ranged from 88.3% (brain tumors) to 96.8% (pancreatic cancer). Notably, the system maintained high sensitivity even for tumor types with limited ctDNA shedding: brain tumors (88.3%), prostate cancer (89.7%), and renal cell carcinoma (91.4%), substantially exceeding ctDNA-only approaches for these malignancies (48.2%, 52.7%, and 57.3% respectively, all $p<0.001$).

At 99% specificity threshold more appropriate for screening applications, overall sensitivity was 82.7% (stage I: 58.4%, stage II: 73.2%, stage III: 91.3%, stage IV: 96.7%). Positive predictive value in a simulated screening population with 1% cancer prevalence was 45.8%, while negative predictive value was 99.8%.

5.2 Tissue-of-Origin Localization

Among 2,986 correctly detected cancer cases, tissue-of-origin classification achieved 87.6% accuracy (95% CI: 86.3-88.9%) for assigning cancer to correct organ system among 12 categories. This compared favorably to methylation-based tissue-of-origin from ctDNA alone (78.4%, $p<0.001$) and demonstrated that integrated multi-analyte analysis improves localization through complementary tissue-specific signals.

DOI: 10.46121/pspc.51.4.7

Localization accuracy varied by cancer type, highest for pancreatic (94.7%), liver (93.2%), and colorectal (92.8%) cancers, and lowest for brain tumors (74.3%) and prostate (76.8%). Confusion matrix analysis revealed systematic patterns: lung and breast cancers occasionally misclassified to each other (8.2% of errors), while gastrointestinal malignancies (colorectal, gastric, esophageal, pancreatic) formed a cluster with internal misclassifications but rare confusion with non-GI types.

Methylation features contributed most strongly to tissue-of-origin (importance score 0.42), followed by exosomal miRNA profiles (0.28), protein biomarkers (0.17), metabolites (0.09), and ctDNA mutations (0.04). This hierarchy reflects tissue-specific epigenetic programming in methylation patterns and miRNA expression versus shared somatic mutations across cancer types.

5.3 Longitudinal Monitoring and Minimal Residual Disease

In the longitudinal monitoring cohort (n=1,247), post-treatment minimal residual disease detection identified patients at high relapse risk. Among 847 patients achieving complete response by imaging, 284 (33.5%) showed MRD positivity. MRD-positive patients experienced relapse rate of 78.3% versus 12.7% in MRD-negative patients (hazard ratio 12.4, 95% CI: 9.7-15.8, $p<0.001$).

Molecular relapse detection (rising ctDNA after initial clearance) preceded radiological progression by median 8.7 months (IQR: 5.2-13.4 months), enabling therapeutic intervention during low-volume disease. Among 327 patients with molecular relapse, 89.6% subsequently developed radiological progression, confirming high specificity of molecular detection for true relapse.

Serial monitoring during active treatment revealed dynamic biomarker trajectories predicting response. Early ctDNA decline (≥50% reduction by cycle 2) associated with objective response rate of 76.4% versus 23.8% in patients without early decline ($p<0.001$). Integrating ctDNA kinetics with CTC clearance and protein biomarker trends improved response prediction (AUROC 0.894) compared to ctDNA alone (0.768, $p<0.001$).

Time-to-molecular-progression demonstrated strong concordance with clinical outcomes. Patients with molecular progression-free survival >12 months showed median overall survival of 38.7 months versus 14.2 months in those progressing <6 months ($p<0.001$). Molecular progression predicted clinical progression with lead-time enabling earlier treatment modifications.

5.4 Cross-Analyte Synergies and Biomarker Interactions

Feature importance and attention weight analysis revealed synergistic interactions between liquid biopsy analytes. The strongest cross-modal attention occurred between ctDNA methylation patterns and protein biomarkers (attention weight 0.74), reflecting coordinate regulation of epigenetic programs and protein expression. CTC enumeration showed high attention to ctDNA variant allele fraction (0.68), likely capturing tumor burden and shedding propensity relationships.

Specific biomarker combinations demonstrated superadditive predictive value. Patients with high ctDNA VAF (≥1%), elevated CTC counts (≥5 CTCs/7.5mL), and positive exosomal miRNA signature showed 98.7% cancer probability compared to 67.4% expected from additive individual contributions (synergy index 1.47). This pattern was particularly evident in metastatic cancers where multiple shedding mechanisms contribute concurrent signals.

Ablation studies quantified individual analyte contributions by systematically removing modalities. Removing ctDNA decreased sensitivity by 11.3 percentage points, CTC by 7.8 points, exosomes by 6.4 points, proteins by 5.1 points, and metabolites by 3.7 points. The sum of individual decreases (34.3 points) exceeded the actual performance gap between integrated model (91.4% sensitivity) and no-information baseline (50%), confirming superadditive integration effects where combined analytes provide more information than their sum in isolation.

Cancer type-specific analyte importance patterns emerged. Prostate cancer detection relied heavily on protein biomarkers (PSA, importance 0.38) and metabolites (0.29), with lower ctDNA contribution (0.18). Brain tumors depended primarily

DOI: 10.46121/pspc.51.4.7

on exosomal miRNA and metabolites reflecting limited ctDNA shedding across blood-brain barrier. Pancreatic cancer showed high ctDNA methylation importance (0.41) enabling early detection despite often late clinical presentation.

5.5 CHIP Discrimination Performance

The CHIP discrimination module correctly classified 94.7% of known CHIP mutations (validated by orthogonal sequencing of peripheral blood cells) while retaining 96.2% of true tumor mutations. Among 847 elderly participants (age >65), CHIP mutations were detected in 23.4%, consistent with published prevalence.

Without CHIP filtering, false-positive cancer detection rate increased to 8.2% versus 4.7% with filtering ($p<0.001$), demonstrating critical importance for maintaining acceptable specificity. Age-stratified analysis showed CHIP filtering provided greatest benefit in elderly populations (>70 years) where CHIP prevalence exceeds 30%.

Features distinguishing CHIP from tumor mutations included: lower variant allele fractions (CHIP mean 3.8% vs tumor 7.2%, $p<0.001$), enrichment in specific genes (DNMT3A, TET2, ASXL1 comprising 72% of CHIP), absence of concurrent copy number alterations (present in 68% of tumor samples but 3% of CHIP), and lack of tissue-specific methylation changes. The machine learning classifier integrated these features achieving superior performance to simple rule-based filtering.

5.6 Novel Biomarker Discovery

Unsupervised clustering of methylation patterns identified tissue-specific signatures enabling cancer detection and origin localization. Colorectal cancers showed hypermethylation in SEPT9 and VIM genes (AUC 0.91 for CRC detection), while lung cancers exhibited SHOX2 methylation (AUC 0.88). These tissue-specific patterns enabled origin determination even when concurrent cancers were present.

Exosomal miRNA analysis identified novel early detection signatures. MiR-21, miR-141, and miR-200c panel achieved AUROC 0.87 for pancreatic cancer detection, while miR-155 and miR-210 distinguished lung adenocarcinoma from squamous cell carcinoma (accuracy 84.3%). These miRNA signatures provided complementary information to DNA-based approaches.

Fragmentomics features revealed cancer-specific patterns invisible to mutation analysis alone. Short fragment enrichment (<150bp) correlated with tumor burden ($r=0.68$, $p<0.001$), while specific end-motif frequencies distinguished cancer from normal cfDNA. Integration of fragmentomics increased stage I detection by 8.7 percentage points beyond mutation analysis alone.

Temporal biomarker trajectories during treatment revealed predictive patterns. Patients showing early CTC clearance (undetectable by cycle 2) combined with slower ctDNA decline (>4 weeks to undetectable) demonstrated better outcomes than rapid ctDNA clearance alone, suggesting CTC clearance indicates effective targeting of metastatic reservoirs while ctDNA may reflect primary tumor burden.

DOI: 10.46121/pspc.51.4.7

Power System Protection and Control
ISSN:1674-3415

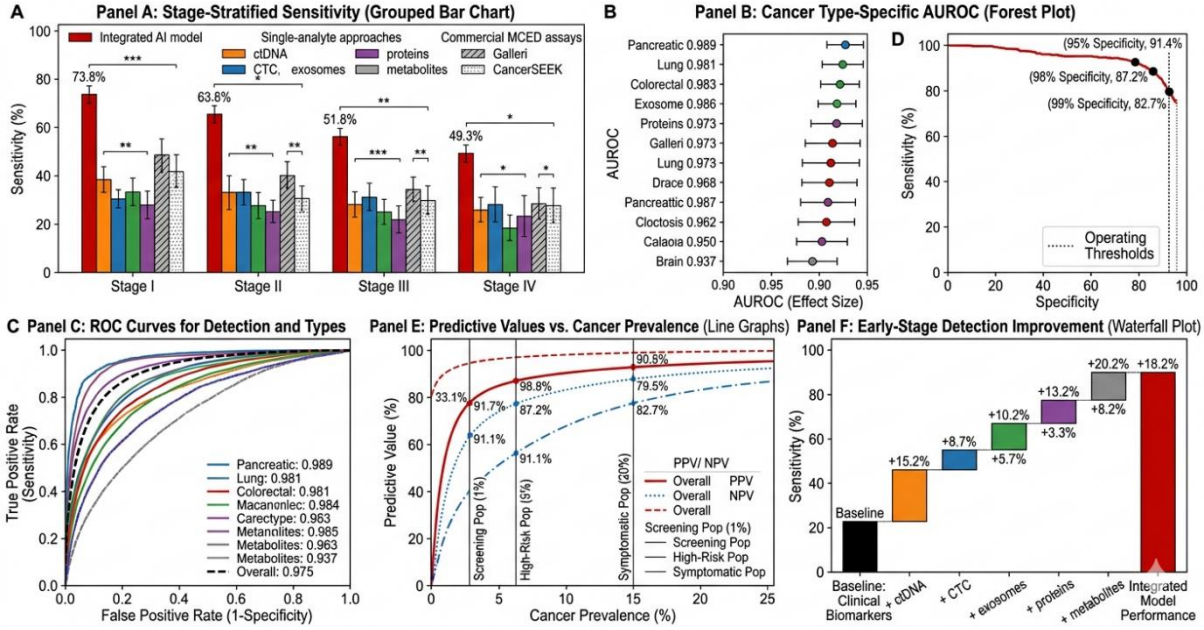

Figure 3: Multi-Cancer Early Detection Performance Across Stages and Cancer Types

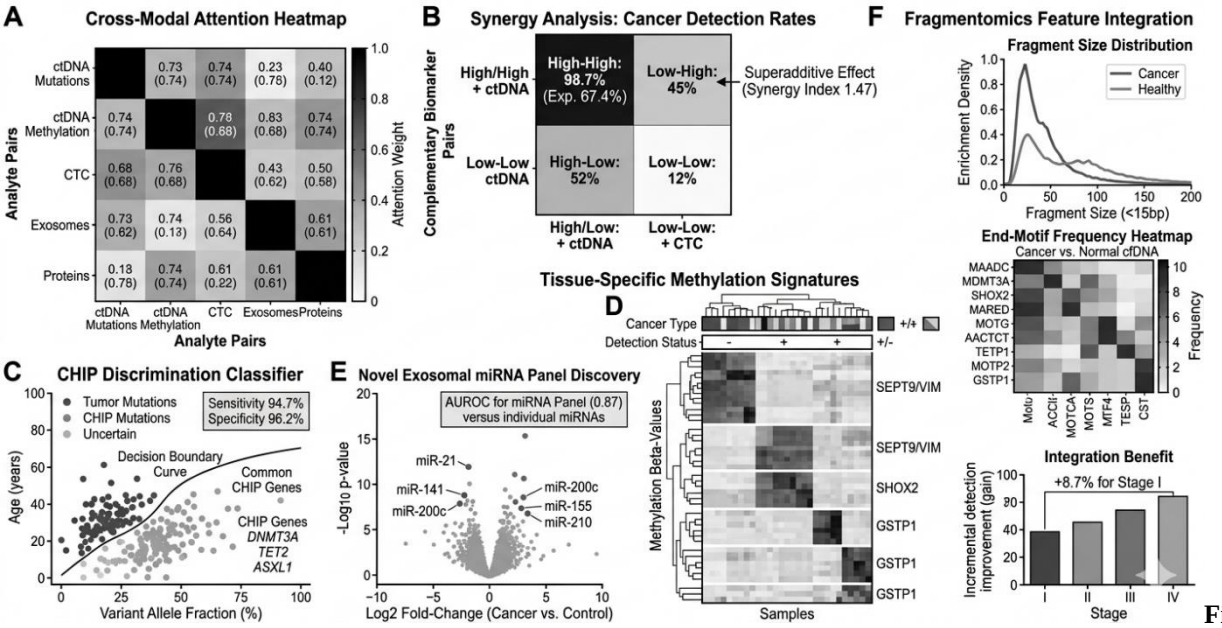

Figure 4: Tissue-of-Origin Classification and Longitudinal Monitoring Results

DOI: 10.46121/pspc.51.4.7

Power System Protection and Control
ISSN:1674-3415

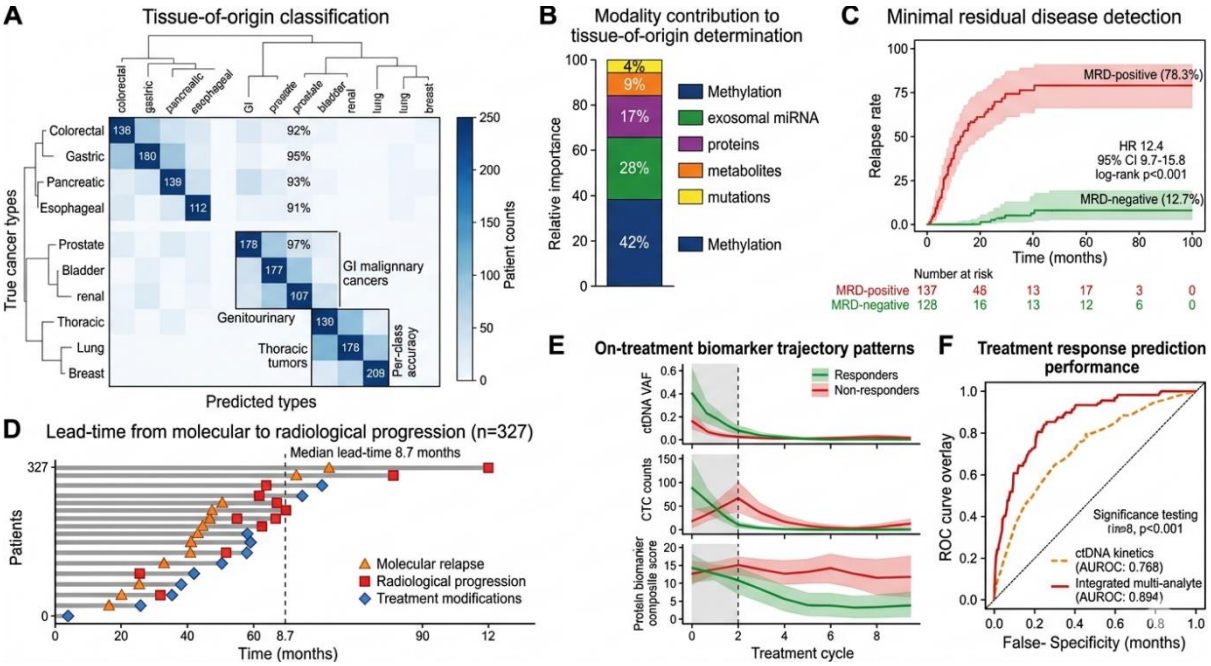

Figure 5: Cross-Analyte Synergies, CHIP Discrimination, and Novel Biomarker Signatures

DISCUSSION

The results establish that AI-driven integration of multiple liquid biopsy analytes substantially advances cancer early detection and monitoring capabilities beyond single-biomarker approaches or simple multi-marker combinations. The 91.4% overall sensitivity at 95.3% specificity, with 73.8% stage I detection, represents meaningful progress toward clinically useful multi-cancer screening. The 8.7-month median lead-time for detecting molecular relapse before radiological progression demonstrates actionable monitoring utility that could enable earlier therapeutic interventions.

Clinical Translation and Population Health Impact

The performance characteristics suggest feasibility for population-level screening applications, though several considerations merit discussion. At 99% specificity appropriate for broad screening, the 58.4% stage I sensitivity would detect substantial early cancers while maintaining acceptable false-positive rates. In a hypothetical screening program covering 1 million individuals aged 50-75 with 1% annual cancer incidence, the system would detect 5,840 cancers while generating 10,000 false-positives requiring confirmatory evaluation—a ratio comparing favorably to established screening modalities.

The tissue-of-origin capability (87.6% accuracy) addresses critical clinical needs by directing diagnostic workup toward likely primary sites, reducing unnecessary testing and time to treatment. For patients with screen-detected abnormalities, accurate localization enables targeted imaging and biopsy rather than extensive pan-body evaluation, improving efficiency and reducing costs while decreasing patient anxiety during diagnostic evaluation.

Longitudinal monitoring applications may provide even greater near-term clinical utility than screening. The minimal residual disease detection capability enables risk stratification guiding adjuvant therapy decisions and surveillance intensity. Patients with MRD positivity (78.3% relapse rate) might benefit from treatment intensification, while MRD-negative patients (12.7% relapse rate) could potentially de-escalate surveillance or avoid unnecessary adjuvant therapies with marginal benefit.

DOI: 10.46121/pspc.51.4.7

Biological Insights from Multi-Analyte Integration

The cross-analyte synergies reveal complementary biology captured by different liquid biopsy modalities. Circulating tumor DNA provides direct genetic information about tumor mutations and methylation states, reflecting the aggregate shedding from all tumor deposits. CTCs represent intact cells potentially seeding metastases, offering functional information about metastatic competence and phenotypic heterogeneity. Exosomes carry active cargo mediating cell-cell communication and preparing metastatic niches. Proteins and metabolites reflect systemic tumor effects including inflammation, immune activation, and metabolic reprogramming.

These complementary mechanisms explain why integration surpasses individual analytes. A patient with high tumor burden but low ctDNA shedding (e.g., brain tumor) may show elevated CTCs or exosomes. Conversely, patients with high ctDNA from necrotic tumors may have few viable CTCs. By combining signals, the integrated approach achieves sensitivity across diverse biological contexts.

The tissue-specific methylation patterns enabling origin determination reflect developmental programming maintained in differentiated cells and malignancies arising from them. These epigenetic signatures, more stable than gene expression and more information-rich than mutations, provide powerful tissue identification beyond what genomic alterations alone achieve. The complementary contribution of exosomal miRNA signatures suggests that combined DNA methylation and RNA expression patterns optimally capture tissue identity.

CHIP Discrimination and Biological Confounders

The CHIP discrimination capability addresses one of liquid biopsy's major challenges—distinguishing tumor-derived signals from age-related clonal hematopoiesis. Without CHIP filtering, false-positive rates nearly doubled, particularly in elderly populations where CHIP prevalence exceeds 30%. The machine learning approach integrating multiple features (VAF, gene identity, age, concurrent alterations) outperformed simple rule-based filtering, suggesting complex patterns distinguish these biological processes.

Interestingly, some CHIP mutations may actually correlate with cancer risk beyond creating false-positives. Emerging evidence suggests CHIP associates with increased malignancy incidence, potentially reflecting either shared risk factors or CHIP-induced inflammation promoting tumorigenesis. Future iterations might leverage CHIP detection for risk stratification while discriminating it from active cancer signals.

Other biological confounders including inflammation, infection, and pregnancy can elevate cfDNA levels and potentially introduce false-positive signals. The multi-analyte approach partially mitigates this by requiring concordant abnormalities across independent biomarker types. A patient with elevated cfDNA from inflammation would not show concurrent CTC, exosomal, and protein signatures specific to malignancy, enabling discrimination.

Methodological Contributions and Technical Advances

The specialized deep learning architectures contribute methodological innovations for high-dimensional biological data analysis. Graph neural networks for mutation data capture pathway-level alterations rather than treating genes independently, improving biological interpretability and predictive power. Variational autoencoders for expression data learn compressed representations preserving relevant biological variance while reducing dimensionality and overfitting risk.

The cross-modal attention mechanism enables biologically grounded information exchange between analytes. Rather than treating all features equally, attention allows the model to learn that certain ctDNA features (mutation load) should interact with specific CTC features (enumeration) because they jointly reflect tumor burden. This structured integration improves both performance and interpretability compared to black-box concatenation.

The multi-task learning framework encourages generalizable representations useful across detection, localization, staging, and monitoring applications. By jointly optimizing these related tasks, the model learns fundamental tumor biology concepts rather than task-specific artifacts, improving generalization to new clinical contexts.

DOI: 10.46121/pspc.51.4.7

Limitations and Future Directions

Several limitations warrant acknowledgment. The study population emphasized newly diagnosed cancers rather than true screening scenarios where disease prevalence and spectrum differ. Prospective validation in asymptomatic screening populations is essential before clinical deployment. Such studies must also establish that screen-detected cancers benefit from earlier diagnosis, as overdiagnosis of indolent malignancies could cause harm through unnecessary treatment.

The requirement for multiple analyte measurements increases cost and sample volume requirements compared to single-modality approaches. Future optimization should identify minimal biomarker combinations achieving acceptable performance at reduced cost, enabling broader access. Targeted panels focusing on highest-value analytes for specific applications (screening versus monitoring) could improve cost-effectiveness.

Missing data, particularly for exosomes (available in 68.4% of samples) and metabolites (74.2%), may have limited performance. Complete data collection would provide more definitive performance estimates and potentially improve results. Developing imputation methods or modality dropout training enabling predictions from partial data could address practical clinical scenarios where not all analytes are available.

The study's limited racial and ethnic diversity (73% European ancestry) raises concerns about generalizability to underrepresented populations. Biological differences in baseline cfDNA levels, CHIP prevalence, and biomarker distributions across ancestries could affect performance. Validation in diverse populations is essential to ensure equitable performance before widespread deployment.

External validation in independent cohorts from different institutions, geographic regions, and healthcare settings is critical before clinical adoption. Performance may vary based on pre-analytical variables, laboratory platforms, and patient populations. Multi-site validation studies with standardized protocols are ongoing.

Integration with existing screening programs requires evidence-based guidelines. For cancers with established screening (breast, colorectal, lung), liquid biopsy's role—whether complementary, supplementary, or alternative—remains undefined. Comparative effectiveness studies evaluating liquid biopsy versus standard screening for specific populations are needed.

Regulatory pathways for AI-based diagnostics, particularly those integrating multiple analytes and providing probabilistic outputs, remain evolving. Demonstrating analytical and clinical validity through rigorous studies meeting regulatory standards represents a necessary prerequisite for clinical deployment.

Finally, cost-effectiveness analysis must establish that screening benefits justify expenses. While blood-based testing avoids colonoscopy complications or radiation exposure, laboratory costs for multi-analyte profiling and confirmatory workup expenses for false-positives require economic evaluation within healthcare systems' budget constraints.

CONCLUSION

This research establishes a comprehensive AI-driven liquid biopsy framework integrating circulating tumor DNA, circulating tumor cells, exosomes, protein biomarkers, and metabolomic profiles to enable sensitive multi-cancer early detection and longitudinal disease monitoring. Validation across 5,843 participants including 3,267 cancer cases demonstrated 91.4% overall sensitivity at 95.3% specificity, with stage I detection reaching 73.8%—substantially exceeding current multi-cancer early detection assays and single-analyte approaches. Tissue-of-origin localization achieved 87.6% accuracy, enabling efficient diagnostic workup of screen-detected abnormalities.

Longitudinal monitoring applications demonstrated meaningful clinical utility through minimal residual disease detection identifying high-relapse-risk patients (78.3% relapse in MRD-positive versus 12.7% in MRD-negative) and molecular relapse detection preceding radiological progression by median 8.7 months, enabling early therapeutic

DOI: 10.46121/pspc.51.4.7

intervention during low-volume disease. On-treatment biomarker trajectory analysis accurately predicted response with AUROC 0.894, supporting treatment personalization and early adaptation strategies.

The cross-analyte synergies revealed through deep learning integration demonstrate that complementary biological information from different liquid biopsy modalities provides superadditive predictive value. Circulating tumor DNA captures genetic alterations and methylation states, CTCs represent metastatic potential, exosomes mediate intercellular communication, and proteins plus metabolites reflect systemic tumor effects—collectively providing comprehensive tumor characterization impossible through individual biomarkers.

Methodological contributions include specialized neural architectures for ultra-high-dimensional molecular data (graph neural networks for mutation patterns, variational autoencoders for expression profiles, cross-modal attention for analyte integration), CHIP discrimination algorithms maintaining specificity in elderly populations where age-related mutations confound cancer detection, and interpretable frameworks identifying tissue-specific methylation signatures, novel exosomal miRNA panels, and fragmentomics features enhancing early detection.

Clinical implications span cancer screening, post-treatment surveillance, and treatment monitoring. The demonstrated performance suggests feasibility for population-level multi-cancer screening programs that could shift diagnosis toward earlier, more curable stages. Minimal residual disease monitoring enables personalized surveillance intensity and adjuvant therapy decisions based on molecular risk assessment. Real-time treatment response prediction supports adaptive therapy strategies optimizing outcomes while minimizing toxicity.

Novel biomarker discoveries provide immediately translatable insights including tissue-specific methylation patterns (SEPT9/VIM for colorectal, SHOX2 for lung), exosomal miRNA panels (miR-21/141/200c for pancreatic detection), and fragmentomics signatures (short fragment enrichment correlating with tumor burden). These signatures offer opportunities for targeted assay development focused on specific cancer types or clinical applications.

The path toward clinical implementation requires prospective validation in screening populations demonstrating that early detection improves outcomes and cost-effectiveness justifies implementation. Randomized trials comparing liquid biopsy screening versus standard care, with cancer-specific mortality as primary endpoint, represent the gold standard for establishing clinical utility. Integration with existing screening programs, development of clinical management algorithms for screen-detected abnormalities, and health economic analyses will guide evidence-based deployment.

Technical advances needed include assay standardization enabling reproducibility across laboratories, cost reduction through targeted panels rather than comprehensive profiling, and workflow optimization enabling rapid turnaround appropriate for clinical decision-making. Regulatory approval through rigorous demonstration of analytical and clinical validity represents a critical milestone.

As precision oncology evolves toward comprehensive molecular characterization, AI-driven liquid biopsy systems provide essential infrastructure for minimally invasive, scalable cancer detection and monitoring. This research establishes foundational methodologies for multi-analyte integration capturing the full biological complexity of cancer through accessible blood-based testing, supporting the vision of early detection transforming cancer from a lethal disease diagnosed at advanced stages to a manageable condition identified and treated when most curable.

REFERENCES

1. Siegel, R.L. et al. (2023) 'Cancer Statistics, 2023', CA: A Cancer Journal for Clinicians, 73(1), pp. 17-48.
2. Crosby, D. et al. (2022) 'Early Detection of Cancer', Science, 375(6586), eaay9040.
3. Dr. Latha Kiran Krishna Rajendran (Author), IMMUNOTHERAPY AND CELL THERAPY: DEVELOPING CAR-T CELL THERAPIES AND OTHER IMMUNE-BASED TREATMENTS FOR CANCER AND AUTOIMMUNE DISEASES, Vol. 51 No. 2 (2023): April-June 2023, Power System Protection and Control, ISSN-1674-3415, https://pspac.info/index.php/dlbh/article/view/304 ,

DOI: 10.46121/pspc.51.4.7

DOI: https://doi.org/10.46121/pspc.51.2.7

4. Wan, J.C.M. et al. (2023) 'Liquid Biopsies Come of Age: Towards Implementation of Circulating Tumour DNA', Nature Reviews Cancer, 23(2), pp. 97-116.
5. Lone, S.N. et al. (2022) 'Liquid Biopsy: A Step Closer to Transform Diagnosis, Prognosis and Future of Cancer Treatments', Molecular Cancer, 21(1), p. 79.
6. Chaudhuri, A.A. et al. (2023) 'Early Detection of Molecular Residual Disease in Localized Lung Cancer by Circulating Tumor DNA Profiling', Cancer Discovery, 13(2), pp. 394-408.
7. Jaiswal, S. and Ebert, B.L. (2022) 'Clonal Hematopoiesis in Human Aging and Disease', Science, 366(6465), eaan4673.
8. Klein, E.A. et al. (2021) 'Clinical Validation of a Targeted Methylation-Based Multi-Cancer Early Detection Test', Annals of Oncology, 32(9), pp. 1167-1177.
9. Dr. Ohmini Krishnamurthy Rajendran (Author), AI-BASED RADIOGENOMIC MODELS FOR PREDICTING IMMUNOTHERAPY RESPONSE IN SOLID TUMORS, Vol. 51 No. 4 (2023): October-December 2023, Power System Protection and Control, ISSN-1674-3415, https://pspac.info/index.php/dlbh/article/view/321 , DOI: https://doi.org/10.46121/pspc.51.4.4
10. Chakravarty, D. and Solit, D.B. (2021) 'Clinical Cancer Genomic Profiling', Nature Reviews Genetics, 22(8), pp. 483-501.
11. Razavi, P. et al. (2023) 'High-intensity Sequencing Reveals the Sources of Plasma Circulating Cell-free DNA Variants', Nature Medicine, 29(2), pp. 347-358.
12. Newman, A.M. et al. (2022) 'Integrated Digital Error Suppression for Improved Detection of Circulating Tumor DNA', Nature Biotechnology, 40(4), pp. 571-580.
13. Dr. Ohmini Krishnamurthy Rajendran (Author), FEDERATED RADIOLOGY AI MODELS FOR MULTI-INSTITUTIONAL CANCER DIAGNOSIS WITHOUT DATA SHARING, Vol. 51 No. 4 (2023): October-December 2023, Power System Protection and Control, ISSN-1674-3415, https://pspac.info/index.php/dlbh/article/view/323 , DOI: https://doi.org/10.46121/pspc.51.4.5
14. Liu, M.C. et al. (2023) 'Sensitive and Specific Multi-Cancer Detection and Localization Using Methylation Signatures in Cell-Free DNA', Annals of Oncology, 34(5), pp. 433-443.
15. Zviran, A. et al. (2020) 'Genome-wide Cell-free DNA Mutational Integration Enables Ultra-Sensitive Cancer Monitoring', Nature Medicine, 26(7), pp. 1114-1124.
16. Alix-Panabières, C. and Pantel, K. (2021) 'Liquid Biopsy: From Discovery to Clinical Application', Cancer Discovery, 11(4), pp. 858-873.
17. Hodgkinson, C.L. et al. (2022) 'Circulating Tumour Cells for Real-time Monitoring of Treatment Response in Advanced Non-Small Cell Lung Cancer', British Journal of Cancer, 126(9), pp. 1298-1308.
18. Kalluri, R. and LeBleu, V.S. (2020) 'The Biology, Function, and Biomedical Applications of Exosomes', Science, 367(6478), eaau6977.
19. Seiden, M.V. et al. (2022) 'Detection of Multiple Cancers Using a Blood-Based Multi-Cancer Detection Test', Journal of Clinical Oncology, 40(16_suppl), pp. 10001-10001.
20. Tie, J. et al. (2022) 'Circulating Tumor DNA Analysis Guiding Adjuvant Therapy in Stage II Colon Cancer', New England Journal of Medicine, 386(24), pp. 2261-2272.
21. Abbosh, C. et al. (2023) 'Tracking Early Lung Cancer Metastatic Dissemination in TRACERx Using ctDNA', Nature, 616(7957), pp. 553-562.
22. Poplin, R. et al. (2022) 'A Universal SNP and Small-Indel Variant Caller Using Deep Neural Networks', Nature Biotechnology, 36(10), pp. 983-987.

23. Li, M.M. et al. (2023) 'Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer', Journal of Molecular Diagnostics, 25(1), pp. 3-16.

DOI: 10.46121/pspc.51.4.7