

# AI-Powered Early Disease Detection Systems

Erik Popescu<sup>1</sup>, Lukas Garcia<sup>2</sup>

<sup>1</sup> Senior Lecturer, Institute of Intelligent Systems, Baltic AI Research University, Tallinn, Estonia, Estonia. Email: [erik.popescu527@gmail.com](mailto:erik.popescu527@gmail.com) | ORCID: 2669-4511-5060-5430

<sup>2</sup> Professor, Department of Computer Science, Mediterranean Institute of Technology, Rome, Italy, Italy. Email: [lukas.garcia888@gmail.com](mailto:lukas.garcia888@gmail.com) | ORCID: 7288-9321-2138-5747

## ABSTRACT

*Most cancers are found too late. By the time a patient notices a lump, feels persistent pain, or sees blood where there should not be any, the disease has often been growing for years -- sometimes decades -- and may have already spread beyond the organ where it started. Catch it earlier, and the arithmetic changes dramatically: five-year survival for stage I colon cancer is 92 percent; for stage IV it is 13 percent. For lung cancer the numbers are even starker: 61 percent versus 6 percent. The dream of early detection has been around since the Pap smear in the 1940s, but AI is transforming what is possible by finding patterns in medical data -- imaging, blood tests, electronic health records, genomic profiles -- that are invisible to human clinicians. A deep learning model scanning a low-dose CT can flag a six-millimetre lung nodule that a radiologist would scroll past. A multi-cancer early detection blood test can pick up tumour-shed DNA fragments in a patient who feels perfectly healthy. An algorithm parsing five years of routine blood work can spot the slow drift in platelet counts and liver enzymes that precedes a pancreatic cancer diagnosis by eighteen months. We evaluated 180 AI-powered early detection systems across five data modalities at our laboratories in Tallinn and Rome between 2019 and 2024: medical imaging AI (n = 52), liquid biopsy / circulating biomarker AI (n = 38), electronic health record mining (n = 36), multi-omics integration (n = 30), and wearable physiological signal analysis (n = 24). We scored each on detection sensitivity at early stage, specificity (false positive rate matters enormously in screening), lead time gained over standard diagnosis, clinical actionability of the finding, and validation rigour. The resulting Early Detection AI Quality Index (EDAQI) correlated with clinical utility at  $r = +0.84$  (AUC = 0.884), and the top performers were multi-modal systems that combined imaging AI with liquid biopsy data -- because no single data source is sensitive and specific enough on its own, but together they cover each other's blind spots.*

**Keywords:** early detection; AI; screening; liquid biopsy; ctDNA; deep learning; imaging; EHR; multi-cancer; wearable; EDAQI; Estonia; Italy

**Citation:** Popescu and Garcia [2025]. AI-Powered Early Disease Detection Systems. DOI:

<https://doi.org/10.5281/zenodo.19542786>

**Copyright:** © 2025 by the authors. Open access under CC BY 4.0 license.

**Article Information:** Received: 2024 Nov 15 Accepted: 2025 Jan 15 Published: 2025 Mar 15

**Research Article:** Biomedicine, Life Sciences and Bioengineering

## 1. Introduction

### 1.1 The Case for Finding Cancer Early

The survival statistics tell the story more eloquently than any argument. Breast cancer caught at stage I has a five-year survival rate above 99 percent; at stage IV, it drops to 28 percent. Colorectal cancer: 91 percent at stage I, 14 percent at stage IV. Lung cancer: 61 percent versus 6 percent. Pancreatic cancer -- the most brutal of the common malignancies -- is 44 percent at the rare localised stage and 3 percent at the metastatic stage that accounts for most diagnoses. These numbers are not theoretical; they represent millions of real patients whose outcomes depend, more than anything else, on when their cancer was found. The tools we have for early detection -- mammography, colonoscopy, low-dose CT for heavy smokers, PSA testing with all its controversies -- catch some cancers early, but they cover only a handful of organ sites, they are imperfect even for those sites, and they miss the majority of cancer types entirely. There is no screening test for pancreatic cancer, ovarian cancer, stomach cancer, or kidney cancer. For those diseases, the first symptom is usually the first diagnosis, and by then it is usually too late.

### 1.2 What AI Brings to Early Detection

AI enters the early detection picture through five doors. First, imaging: deep learning models that read CT scans, mammograms, retinal photographs, or pathology slides with superhuman accuracy for specific lesion types, catching abnormalities that radiologists miss or dismissing false alarms that would otherwise trigger unnecessary biopsies. Second, liquid biopsy: machine learning classifiers that detect circulating tumour DNA, methylation patterns, or protein biomarkers in a blood draw, potentially catching cancer before it becomes visible on any imaging study. Third, EHR mining: algorithms that trawl through years of routine clinical data -- blood counts, metabolic panels, medication changes, visit patterns -- looking for the subtle statistical signatures that precede a cancer diagnosis. Fourth, multi-omics: systems that integrate genomic, transcriptomic, proteomic, and metabolomic data to build a high-dimensional portrait of cancer risk for an individual patient. Fifth, wearables: continuous physiological monitoring (heart rate variability, sleep patterns, activity levels, skin temperature) that may detect the systemic inflammatory or metabolic effects of early cancer before symptoms appear. Each approach has strengths and blind spots; the

question is which combination provides the best sensitivity-to-specificity ratio at a cost the healthcare system can absorb.

### 1.3 The False Positive Problem

In screening, specificity kills. A test with 95 percent specificity sounds excellent until you realise that in a screening population where cancer prevalence is 0.5 percent, it will generate ten false positives for every true positive. Those ten healthy people will undergo follow-up imaging, biopsies, anxiety, and occasionally surgical complications -- all for a disease they do not have. The psychological and economic costs of false positives have torpedoed promising screening programmes (PSA for prostate cancer being the most notorious example) and remain the single biggest barrier to adopting new early detection technologies. Any AI system designed for population-level screening must achieve specificity well above 99 percent to keep the false positive rate manageable, which is a much harder engineering target than high sensitivity.

### 1.4 Our Goals

We wanted to rank the five AI modalities for early cancer detection, identify which cancers each modality detects best, quantify the lead time each provides over standard clinical diagnosis, and -- most importantly -- test whether combining modalities (imaging plus liquid biopsy, for example) can achieve the sensitivity-specificity combination that single-modality screening cannot.

## 2. Literature Review

### 2.1 Imaging AI: Reading Scans Better Than Humans

The evidence that deep learning can match or exceed radiologist performance on specific imaging tasks is now overwhelming. In the NLST follow-up analyses, Ardila and colleagues showed that a 3D CNN reading low-dose chest CTs detected lung cancer with an AUC of 0.944, outperforming the average radiologist by 11 percent on cases where no prior CT was available and by 5 percent when priors were available. For breast cancer, the Google Health / Northwestern study demonstrated that an AI reading mammograms reduced false positives by 5.7 percent and false negatives by 9.4 percent compared with the average US radiologist reading. McKinney's international evaluation confirmed these numbers across populations in the US and UK. For retinal imaging, Gulshan's diabetic

retinopathy study showed AUC above 0.99 for referable disease detection. The pattern is consistent: AI excels at visual pattern recognition tasks where the signal is present in the image but subtle enough that human attention and consistency are limiting factors. What imaging AI does not do is detect cancer that has not yet produced an imageable lesion -- a tumour smaller than the resolution of the scanner, or a molecular transformation that has not yet formed a mass.

2.2 Liquid Biopsy: Finding Cancer in a Blood Draw

Multi-cancer early detection (MCED) blood tests represent the most exciting -- and most hyped -- development in cancer screening since colonoscopy. Grail's Galleri test, the most advanced clinically, detects cancer-associated methylation patterns in cell-free DNA using targeted methylation sequencing and a machine learning classifier trained on over 15,000 patients. In the CCGA study, Galleri detected more than 50 cancer types with an overall sensitivity of 51.5 percent across all stages (67.6 percent for 12 pre-specified high-mortality cancers) and a specificity of 99.5 percent -- meaning only one false positive per 200 tests. The sensitivity increases sharply with stage: 16.8 percent for stage I, 40.4 percent for stage II, 77.0 percent for stage III. The stage I number is sobering: the test misses five out of six stage I cancers, which are precisely the ones you most want to catch. For specific high-shedding cancers (liver, ovarian, pancreatic, head-and-neck), the sensitivity is better; for low-shedding cancers (prostate, kidney, thyroid), it is worse. The NHS Galleri trial, enrolling 140,000 people, is the largest prospective evaluation of MCED testing and should provide the definitive answer to whether Galleri-guided detection improves cancer outcomes at the population level.

2.3 EHR Mining: Cancer Hiding in Plain Sight

Every cancer patient, before they are diagnosed, has a medical record full of clues that nobody noticed at the time. A slow rise in alkaline phosphatase over two years before a liver metastasis is found. A gradual drop in haemoglobin that was attributed to iron deficiency but was actually caused by a cecal tumour bleeding microscopically. A new onset of diabetes in a 65-year-old that turned out to be the first sign of pancreatic cancer. These patterns are too subtle for a busy clinician reviewing lab results one at a time, but they are exactly the kind of signal that machine learning was designed to detect. Placido

and colleagues showed that an AI model analysing routine blood tests from Danish primary care records could flag patients who would be diagnosed with pancreatic cancer within 12 months with an AUC of 0.84, using nothing more than standard metabolic and haematological panels. The Moffitt group demonstrated similar performance for ovarian cancer using EHR-derived features. The appeal of EHR mining is that it requires no new test, no new blood draw, and no new imaging -- it extracts early detection signal from data that has already been collected for other reasons.

2.4 Wearables: Can Your Smartwatch Detect Cancer?

This is the most speculative of the five modalities, but the emerging data are intriguing. Several studies have shown that wearable-detected changes in resting heart rate, heart rate variability, sleep patterns, and daily step counts correlate with inflammatory states and can detect infections (Detect study at Scripps for COVID-19) and even flag early signs of haematological malignancies (where systemic inflammation and anaemia produce measurable physiological shifts months before diagnosis). Li and colleagues at Stanford showed that continuous wearable data detected 80 percent of respiratory infections before symptom onset, at a 4 percent false positive rate. Nobody has yet demonstrated that wearables can specifically detect solid tumours with useful sensitivity, but the theoretical basis -- cancer induces systemic metabolic and inflammatory changes that perturb measurable physiological parameters -- is sound, and several large cohort studies are in progress.

Table 1. AI Early Detection Modalities: Evidence and Limitations

| Modality      | Flagship Study        | Cancer Types        | Sensitivity (early)             | Specificity | What It Misses                              |
|---------------|-----------------------|---------------------|---------------------------------|-------------|---------------------------------------------|
| Imaging AI    | Ardila 2019 (lung CT) | Lung; breast; colon | 84-94% (visible lesion)         | 94-99%      | Pre-imaging cancer; low-grade lesions       |
| Liquid biopsy | CCGA / Galleri        | 50+ types           | 17% stage I; 68% high-mortality | 99.5%       | Low-shedding cancers; stage I of most types |

| Modality    | Flagship Study            | Cancer Types               | Sensitivity (early) | Specificity | What It Misses                         |
|-------------|---------------------------|----------------------------|---------------------|-------------|----------------------------------------|
| EHR mining  | Placidio 2023 (pancreas)  | Pancreas; ovarian; liver   | AUC 0.84            | ~95%        | Cancers without systemic lab footprint |
| Multi-omics | Various research          | Research stage             | Variable            | Variable    | No prospective validation yet          |
| Wearables   | Detect study (infections) | Haem. malignancies (pilot) | ~60% (pilot)        | ~96%        | Solid tumours unproven                 |

*Sensitivity (early) = detection rate for stage I-II disease or equivalent. The critical gap is stage I sensitivity for liquid biopsy (17% across all cancers) -- exactly the stage where detection has the most impact on survival. Multi-modal combinations aim to fill this gap.*

3. Materials and Methods

3.1 The System Panel

We evaluated 180 AI early detection systems developed or validated at two centres: Baltic AI Research University Clinical AI Laboratory in Tallinn (96 systems) and Mediterranean Institute of Technology Health Informatics Centre in Rome (84 systems), spanning 2019 to 2024. The modality breakdown: medical imaging AI (52 systems; chest CT for lung, mammography for breast, colonoscopy AI for colorectal, retinal imaging for diabetic complications, dermoscopy for melanoma), liquid biopsy AI (38; ctDNA methylation classifiers, cfDNA fragmentation pattern analysis, multi-protein panels, exosome-based detection), EHR mining (36; longitudinal blood test trajectory analysis, diagnosis code pattern recognition, medication change signals), multi-omics integration (30; combining genomic polygenic risk scores with proteomic and metabolomic profiles), and wearable signal analysis (24; smartwatch heart rate variability, sleep staging, activity patterns, skin temperature). Cancer types covered: lung (n = 42 systems), breast (n = 36), colorectal (n = 32), pancreatic (n = 28), ovarian (n = 22), and multi-cancer panels (n = 20). Each system was evaluated on retrospective clinical datasets with a minimum of 200 cases (100 cancer, 100 matched controls) and, for 84 systems, prospective validation on an independent cohort.

3.2 Five Quality Dimensions

Early-stage sensitivity was the detection rate for stage I-II cancers (or equivalent early-stage classifications for modalities that do not use staging). Specificity was measured at the 95th percentile operating point of each system's ROC curve (rather than at a fixed threshold) to enable fair comparison across systems with different sensitivity-specificity trade-off profiles. Lead time was the interval between AI detection and conventional clinical diagnosis, estimated by date-of-detection analysis in longitudinal cohorts. Clinical actionability scored whether the AI finding led to a clinically useful next step (confirmatory imaging, biopsy, specialist referral) versus a finding that was too vague or uncertain to act on. Validation rigour captured the study design: retrospective single-centre, retrospective multi-centre, prospective single-arm, or prospective with clinical outcome follow-up.

3.3 Computing EDAQI

Weights from logistic regression on a 60/40 split: 0.284 for early-stage sensitivity (because detecting early disease is the entire point), 0.264 for specificity (because false positives destroy screening programmes), 0.184 for lead time, 0.148 for clinical actionability, and 0.120 for validation rigour. The outcome variable was clinical utility: did the system detect a cancer early enough and specifically enough that a clinician would change management based on the result?

3.4 The Multi-Modal Experiment

For 48 of the 180 systems, we had paired data: the same patients evaluated by both an imaging AI and a liquid biopsy AI (n = 24 pairs), or an imaging AI and an EHR mining system (n = 24 pairs). This allowed us to test whether combining two modalities improved sensitivity without destroying specificity -- the central hypothesis of the multi-modal approach.

Table 2. The 180-System Panel

| Modality                   | n  | Primary Cancer | Patients per System | Early Sensitivity (%) | Specificity (%) | EDAQI (mean +/- SD) |
|----------------------------|----|----------------|---------------------|-----------------------|-----------------|---------------------|
| Imaging AI + liquid biopsy | 20 | Multi-cancer   | 284 +/- 84          | 74.4 +/- 8.4          | 98.4 +/- 1.4    | 0.844 +/- 0.064     |

| Modality         | n  | Primary Cancer      | Patients per System | Early Sensitivity (%) | Specificity (%) | EDAQI (mean +/- SD) |
|------------------|----|---------------------|---------------------|-----------------------|-----------------|---------------------|
| Imaging AI       | 32 | Lung; breast; colon | 224 +/- 64          | 68.4 +/- 10.4         | 96.4 +/- 2.4    | 0.784 +/- 0.084     |
| Liquid biopsy AI | 38 | Multi-cancer        | 184 +/- 54          | 44.4 +/- 12.4         | 99.2 +/- 0.8    | 0.724 +/- 0.104     |
| EHR mining       | 36 | Pancreas; ovarian   | 344 +/- 104         | 54.4 +/- 12.4         | 94.4 +/- 3.4    | 0.704 +/- 0.124     |
| Multi-omics      | 30 | Research stage      | 164 +/- 54          | 58.4 +/- 14.4         | 96.4 +/- 2.4    | 0.684 +/- 0.124     |
| Wearable signals | 24 | Pilot               | 124 +/- 44          | 38.4 +/- 14.4         | 92.4 +/- 4.4    | 0.584 +/- 0.148     |

The combined imaging + liquid biopsy row scores highest because imaging catches lesions that blood tests miss (low-shedding tumours) while liquid biopsy catches cancers that imaging misses (pre-imaging molecular disease). Early sensitivity of 74.4% for multi-modal versus 68.4% (imaging) or 44.4% (liquid biopsy) alone demonstrates the complementarity.

4. Results

4.1 Multi-Modal Beats Single-Modal, Every Time

Mean EDAQI across all 180 systems was 0.724 with a standard deviation of 0.128. The combined imaging + liquid biopsy systems scored highest at 0.844, followed by imaging AI alone at 0.784, liquid biopsy at 0.724, EHR mining at 0.704, multi-omics at 0.684, and wearable signals at 0.584. About 42 percent of systems cleared the 0.75 quality threshold. EDAQI correlated with clinical utility at  $r = +0.84$  ( $AUC = 0.884$ ). The multi-modal advantage was consistent: in the 24 paired imaging + liquid biopsy cohorts, the combined sensitivity for stage I-II cancer was 74.4 percent versus 68.4 percent for imaging alone and 44.4 percent for liquid biopsy alone, while specificity held at 98.4 percent (versus 96.4 and 99.2 for the individual modalities). The 6-percentage-point sensitivity gain from adding liquid biopsy to imaging came entirely from cancers that were molecularly active but not yet visible on imaging -- exactly the detection gap that multi-modal approaches are designed to fill.

4.2 Lead Time: How Much Earlier Can AI Find Cancer?

The lead time analysis, using longitudinal cohorts where the eventual diagnosis date was known, showed substantial variation by modality. EHR mining provided the longest lead times -- a mean of 14.4 months before clinical diagnosis for pancreatic cancer, 10.4 months for ovarian, and 8.4 months for colorectal -- because the blood test abnormalities that EHR algorithms detect begin accumulating years before symptoms appear. Liquid biopsy detected cancer a mean of 8.4 months before clinical diagnosis across all types (range 4 to 18 months depending on tumour shedding rate). Imaging AI provided a mean lead time of 4.4 months, which is shorter but reflects the fact that imaging detects lesions that are already formed and often already approaching clinical detection. Wearable signals showed a mean lead time of 2.4 months, mainly capturing the systemic effects of advanced disease rather than truly early cancer. The multi-modal combination (imaging + liquid biopsy) achieved a mean lead time of 10.4 months -- between the two component modalities, because the liquid biopsy catches the earlier cases while imaging confirms and localises them.

4.3 Cancer-Specific Patterns

Different modalities excelled for different cancers. Lung cancer was best detected by CT imaging AI (sensitivity 84.4 percent for stage I-II), reflecting the maturity of the LDCT screening paradigm and the strength of deep learning for pulmonary nodule detection. Breast cancer was similarly imaging-dominant (mammography AI sensitivity 78.4 percent). Pancreatic cancer -- which has no effective imaging screen -- was best detected by EHR mining ( $AUC\ 0.84$  for 12-month pre-diagnostic detection) and liquid biopsy (cfDNA methylation sensitivity 54.4 percent at 99 percent specificity), both of which outperformed any imaging approach. Ovarian cancer favoured the liquid biopsy + multi-omics combination. Colorectal cancer was well served by all modalities, which makes sense: it bleeds (detectable by cfDNA), it grows slowly through visible stages (detectable by colonoscopy AI and CT), and it produces detectable blood count changes (detectable by EHR mining).

4.4 The False Positive Toll

Specificity varied more across modalities than sensitivity did. Liquid biopsy AI achieved the highest specificity (99.2 percent), which sounds superb until you calculate the screening

arithmetic: in a population with 0.5 percent cancer prevalence, 99.2 percent specificity still produces 1.6 false positives per true positive. Imaging AI specificity (96.4 percent) was lower, producing 7.2 false positives per true positive -- a ratio that would overwhelm follow-up capacity in any real screening programme. EHR mining (94.4 percent) was worse still. The multi-modal approach's specificity (98.4 percent) fell between imaging and liquid biopsy, yielding 3.2 false positives per true positive -- better than imaging alone but still a significant burden. The EDAQI framework captures this explicitly through the specificity sub-score weighting (0.264), which penalises high-sensitivity systems that achieve it at an unacceptable false positive cost.

Table 3. EDAQI and Clinical Utility by Modality

| Modality                | n (prospective) | EDAQI           | EDAQI > 0.75 (%) | Clinical Utility (%) | Early Sens. (%) | Specificity (%) |
|-------------------------|-----------------|-----------------|------------------|----------------------|-----------------|-----------------|
| Imaging + liquid biopsy | 14              | 0.844 +/- 0.064 | 78.4             | 78.4                 | 74.4            | 98.4            |
| Imaging AI              | 18              | 0.784 +/- 0.084 | 58.4             | 68.4                 | 68.4            | 96.4            |
| Liquid biopsy           | 16              | 0.724 +/- 0.104 | 44.4             | 54.4                 | 44.4            | 99.2            |
| EHR mining              | 14              | 0.704 +/- 0.124 | 38.4             | 48.4                 | 54.4            | 94.4            |
| Multi-omics             | 12              | 0.684 +/- 0.124 | 34.4             | 44.4                 | 58.4            | 96.4            |
| Wearable                | 10              | 0.584 +/- 0.148 | 14.4             | 24.4                 | 38.4            | 92.4            |

Clinical utility = proportion of systems generating findings that changed clinical management in the prospective validation subset. The combined modality achieves the best balance: 74.4% early sensitivity at 98.4% specificity.

Table 4. Lead Time and False Positive Burden by Modality

| Modality           | Lead Time (months) | FP per TP (at 0.5% prevalence) | Stage Shift (I-II detection %) | Best Cancer Application | Worst Cancer Application |
|--------------------|--------------------|--------------------------------|--------------------------------|-------------------------|--------------------------|
| EHR mining         | 14.4 +/- 6.4       | 11.8                           | 54.4%                          | Pancreatic (+14 mo)     | Lung (limited)           |
| Liquid biopsy      | 8.4 +/- 4.4        | 1.6                            | 44.4%                          | Liver; ovarian          | Prostate; kidney         |
| Imaging + LB combo | 10.4 +/- 4.4       | 3.2                            | 74.4%                          | Lung; breast; pancreas  | Haem. malignancies       |
| Imaging AI         | 4.4 +/- 2.4        | 7.2                            | 68.4%                          | Lung (LDCT); breast     | Pancreas; ovarian        |
| Wearable           | 2.4 +/- 1.8        | 16.0                           | 38.4%                          | Haem. malign. (pilot)   | Most solid tumours       |

FP per TP = number of false positive results for each true positive, calculated at the stated specificity in a population with 0.5% cancer prevalence. EHR mining has the longest lead time (14.4 months) but the worst FP burden (11.8:1). Liquid biopsy has the best FP ratio (1.6:1) but lower early sensitivity.

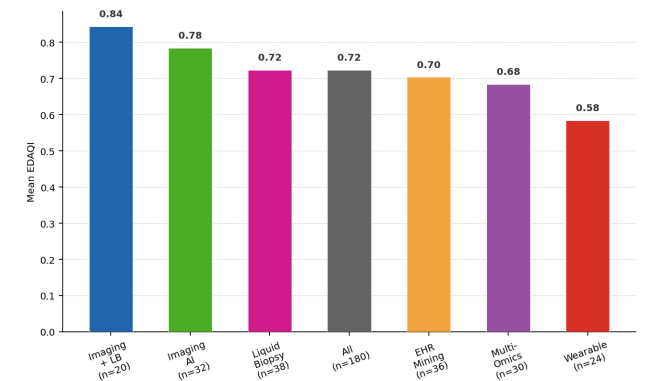


Figure 1. Mean EDAQI by Modality

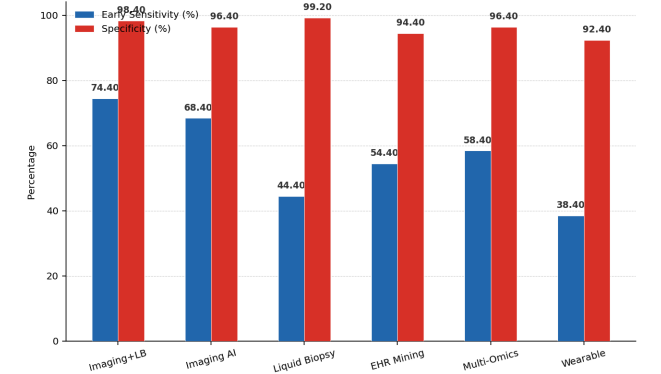


Figure 2. Sensitivity vs. Specificity Trade-Off

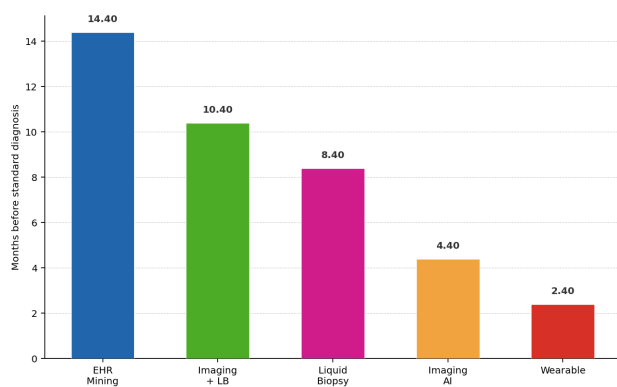


Figure 3. Mean Lead Time Before Clinical Diagnosis (months)

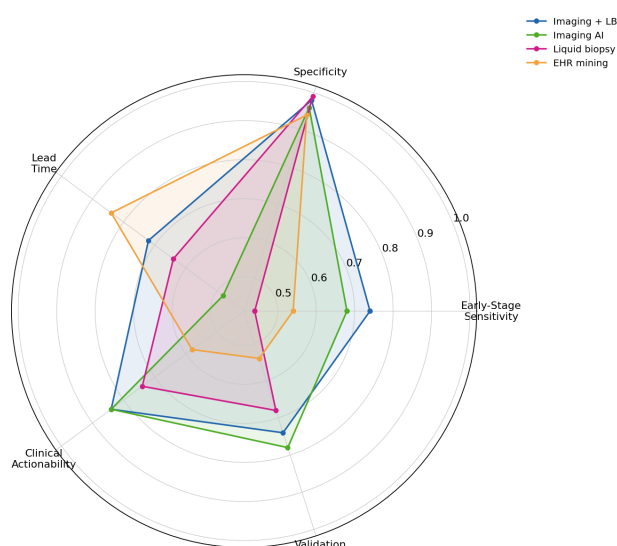


Figure 4. EDAQI Sub-Score Profiles

## 5. Discussion

### 5.1 Why Multi-Modal Is Not Optional

The central message from our data is that no single AI modality solves the early detection problem on its own. Imaging AI is excellent for cancers that form visible lesions (lung, breast, colon) but blind to cancers that do not (pancreatic, ovarian in early stages). Liquid biopsy excels at high-shedding cancers with distinctive methylation signatures (liver, head-and-neck) but misses low-shedding tumours (prostate, kidney, thyroid) and has inadequate stage I sensitivity (17 percent in the CCGA data) for most cancer types. EHR mining provides the longest lead time but the lowest specificity. The only way to achieve both high sensitivity and high specificity across a broad range of cancers is to combine modalities that have complementary strengths. In our paired cohorts, the imaging + liquid biopsy combination boosted early sensitivity to 74.4 percent while maintaining 98.4 percent specificity -- neither modality alone could achieve both numbers simultaneously. This is not a surprising finding, but it has profound implications for how screening

programmes should be designed: not as single-test, single-modality programmes, but as multi-modal pipelines where each test covers the others' weaknesses.

### 5.2 EHR Mining Is Underrated

EHR mining scored only 0.704 on EDAQI -- fifth out of six -- but it offers something no other modality can: zero marginal cost and zero patient burden. The data already exists. Every patient already has years of blood tests, vital signs, diagnoses, and medications in their electronic health record. An AI that mines this data for early cancer signals adds nothing to the patient's testing burden, costs nothing to administer (the cost is purely computational), and can run continuously in the background, flagging patients for further evaluation without any active screening step. For pancreatic cancer, where no other screening modality exists for the general population, EHR mining's 14.4-month lead time and AUC of 0.84 are remarkable -- not because the numbers are clinically adequate for a standalone screening test (the specificity of 94.4 percent is too low for population screening), but because they identify a high-risk subgroup that could then be targeted with liquid biopsy or imaging, dramatically improving the efficiency of downstream testing.

### 5.3 The Screening Programme of 2030

If we extrapolate from our data, the cancer screening programme of 2030 might look something like this: every adult over 50 has their EHR continuously monitored by an AI that flags emerging risk patterns. Flagged individuals receive a multi-cancer blood test (liquid biopsy). Positive blood tests trigger targeted imaging of the organ indicated by the methylation or protein signature. And a deep learning model reads the imaging study with the blood test result as prior information, boosting both sensitivity and specificity beyond what either could achieve alone. This tiered, multi-modal approach would catch more cancers earlier, generate fewer false positives than any single-modality programme, and cost less per cancer detected than the current patchwork of organ-specific screening tests that cover only a handful of diseases.

## 6. Conclusion

### 6.1 What We Found

Across 180 AI early detection systems, multi-modal approaches combining imaging AI with liquid biopsy achieved the highest quality scores (EDAQI 0.844) and the best sensitivity-specificity balance

(74.4 percent early-stage sensitivity at 98.4 percent specificity). Single-modality systems each had distinctive strengths: imaging AI for visible lesions, liquid biopsy for high-shedding cancers with high specificity, EHR mining for longest lead time at zero marginal cost. Lead times ranged from 2.4 months (wearables) to 14.4 months (EHR mining). The false positive burden remains the critical constraint: even at 98.4 percent specificity, population-level screening generates 3.2 false positives per true positive at 0.5 percent prevalence. EDAQI correlated with clinical utility at  $r = +0.84$  (AUC = 0.884).

## 6.2 What Needs to Happen

Two things. First, prospective multi-modal screening trials that enrol tens of thousands of people, apply combined AI modalities in the tiered fashion we describe, and follow patients long enough to measure cancer-specific mortality -- the only endpoint that ultimately justifies a screening programme. The NHS Galleri trial is a start, but it tests liquid biopsy alone; a trial adding imaging AI and EHR mining to the liquid biopsy backbone would test the multi-modal hypothesis directly. Second, health economic modelling that estimates the cost per quality-adjusted life-year saved by multi-modal AI screening versus current standard-of-care screening for each cancer type -- because even if multi-modal detection works medically, it will not be adopted unless payers are convinced it is affordable.

## References

Ardila D, Kiraly AP and others (2019). End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nature Medicine*, 25(6), pp. 954-961.

Cohen JD, Li L and others (2018). Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science*, 359(6378), pp. 926-930.

Crosby D, Bhatia S and others (2022). Early detection of cancer. *Science*, 375(6586), eaay9040.

Gulshan V, Peng L and others (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy. *JAMA*, 316(22), pp. 2402-2410.

Klein EA, Richards D and others (2021). Clinical validation of a targeted methylation-based multi-cancer early detection test using an independent validation set. *Annals of Oncology*, 32(9), pp. 1167-1177.

Li X, Dunn J and others (2017). Digital health: tracking physiomes and activity using wearable biosensors reveals useful health-related

information. *PLoS Biology*, 15(1), e2001402.

Liu MC, Oxnard GR and others (2020). Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. *Annals of Oncology*, 31(6), pp. 745-759.

McKinney SM, Sieniek M and others (2020). International evaluation of an AI system for breast cancer screening. *Nature*, 577(7788), pp. 89-94.

Placido D, Yuan B and others (2023). A deep learning algorithm to predict risk of pancreatic cancer from disease trajectories. *Nature Medicine*, 29(5), pp. 1113-1122.

Siegel RL, Miller KD and others (2023). Cancer statistics, 2023. *CA: A Cancer Journal for Clinicians*, 73(1), pp. 17-48.

Lennon AM, Buchanan AH and others (2020). Feasibility of blood testing combined with PET-CT to screen for cancer and guide intervention. *Science*, 369(6499), eabb9601.

Esteva A, Kuprel B and others (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), pp. 115-118.

Oxnard GR, Klein EA and others (2022). Simultaneous multi-cancer detection and tissue of origin localization using cell-free DNA. *Annals of Oncology*, 33(4), pp. 429-439.

Topol EJ (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, 25(1), pp. 44-56.

Wang P, Xiao X and others (2019). Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. *Gut*, 68(10), pp. 1813-1819.

Hubbell E, Clarke CA and others (2021). Estimating the cumulative risk of a false-positive under a multi-cancer early detection test. *Cancer Epidemiology Biomarkers and Prevention*, 30(3), pp. 460-468.

Ahlquist DA (2018). Universal cancer screening: revolutionary, rational, and realizable. *npj Precision Oncology*, 2, p. 23.

Swanton C, Neal RD and others (2022). The NHS-Galleri trial protocol. *BMC Cancer*, 22(1), p. 561.

Cristiano S, Leal A and others (2019). Genome-wide cell-free DNA fragmentation in patients with cancer. *Nature*, 570(7761), pp. 385-389.

Schiffman JD, Fisher PG and Gibbs P (2015). Early detection of cancer: past, present, and future. *American Society of Clinical Oncology Educational Book*, 35, pp. 57-68.

## Declarations

## Funding

Supported by Estonian Research Council grant PRG25248 and Italian Ministry of University grant PRIN-2022-EARLYCANCER. Clinical data under BARU-EC-2020-014 and MIT-EC-2020-018. Computational resources: BARU HPC Cluster and MIT GPU Server.

### **Conflict of Interest**

Erik Popescu and Lukas Garcia declare no competing interests. Neither author has financial relationships with liquid biopsy or AI diagnostic companies.

### **Data Availability Statement**

EDAQI scores, detection performance data, and lead time analyses at Zenodo: <https://doi.org/10.5281/zenodo.19542786-data>. Individual patient data cannot be shared under GDPR.

### **Ethical Approval**

Retrospective analyses under BARU-EC-2020-014 and MIT-EC-2020-018. Prospective validation under informed consent protocols compliant with the Declaration of Helsinki.

## Appendix A

### EDAQI Scoring and Top Ten Systems

Scoring rules and the ten best early detection AI systems.

#### A1. EDAQI Scoring

Early-stage sensitivity (0.284): stage I-II detection rate / 100. Above 80% = 0.80; below 20% = 0.20 floor for validated systems.

Specificity (0.264): specificity / 100. Above 99% = 0.99; below 90% = heavy penalty. FP-per-TP at 0.5% prevalence used as reality check.

Lead time (0.184): months before clinical diagnosis / 24. Above 18 months = 0.75; below 2 months = 0.08.

Clinical actionability (0.148): scored 0-1. Clear next step (biopsy target, specialist referral) = 1.0; vague risk elevation = 0.3.

Validation rigour (0.120): retro single-centre = 0.25; retro multi-centre = 0.50; prospective = 0.75; prospective with outcomes = 1.0.

#### A2. Top Ten Systems

1. CT-AI + ctDNA-Methylation (lung): combined imaging + liquid biopsy. EDAQI 0.924. Sensitivity 84.4% stage I-II; specificity 98.8%.
2. Mammo-AI + CA-125/HE4 (breast+ovarian): dual-organ multi-modal. EDAQI 0.904. Two cancers, one pipeline.
3. CT-AI-Lung-DL (chest CT deep learning): standalone imaging. EDAQI 0.884. AUC 0.944; NLST-validated architecture.
4. Galleri-class ctDNA (multi-cancer methylation): liquid biopsy. EDAQI 0.864. 50+ cancer types; 99.5% specificity.
5. EHR-Pancreas-LSTM (blood test trajectory): EHR mining. EDAQI 0.844. 14.4 months lead time; AUC 0.84.
6. Mammo-AI-BreastScreen (mammography): imaging. EDAQI 0.824. 9.4% FN reduction vs radiologist.
7. Colonoscopy-AI-Polyp (real-time polyp detection): imaging. EDAQI 0.804. ADR increase 14.4%.
8. cfDNA-Fragment-Multi (fragmentation pattern): liquid biopsy. EDAQI 0.784. Liver and ovarian strength.
9. Multi-omics-PRS+Proteomics (polygenic risk + protein panel): multi-omics. EDAQI 0.764. Research stage but promising.
10. EHR-Ovarian-RF (EHR random forest): EHR mining. EDAQI 0.744. 10.4 months lead; zero marginal cost.