
Unlocking Early Lung Cancer Detection through AI in General Practice



Lung cancer remains one of the deadliest forms of cancer, often detected too late for effective treatment. While general practitioners are typically the first point of contact for patients, around 80% of lung cancer cases are still identified at an advanced stage. Early diagnosis is vital, yet existing methods struggle to catch early indicators. A recent study explored how artificial intelligence, particularly natural language processing (NLP), can harness the untapped potential of general practice records to identify signs of lung cancer months before referral.

Leveraging Routine Data with AI Tools

The study investigated whether routinely recorded data in general practice could help flag potential lung cancer cases earlier. Researchers analysed data from over 525,000 patients across four Dutch academic primary care networks. The cohort included 2,386 confirmed cases of lung cancer. Two AI models were developed: one based solely on free-text consultation notes, and another incorporating both text and structured data, such as symptom codes and patient demographics. These models aimed to identify patients at risk up to five months before official diagnosis—approximately four months before referral.

Must Read: [Improving Patient Understanding of Clinical Notes with AI](#)

Free-text entries, which are typically underutilised in predictive tools, were processed using a phrase skip-gram algorithm to convert language into meaningful variables. Logistic regression was then applied to predict the likelihood of a lung cancer diagnosis. By avoiding reliance on manually coded data—often incomplete or selective—the models captured subtler, potentially overlooked patterns. This approach provided a richer, more comprehensive foundation for risk prediction.

Evaluating Model Performance and Predictive Value

Internal validation showed that both models performed strongly, with the text-only model achieving an area under the curve (AUROC) of 0.88. When tested externally using a leave-one-centre-out approach, performance slightly declined but remained robust with an AUROC around 0.79. These figures highlight the models' ability to distinguish between patients with and without lung cancer several months before typical diagnosis pathways would detect it.

However, accuracy is only part of the equation. Despite high discrimination, the models' positive predictive values (PPVs) remained modest due to the low prevalence of lung cancer. At a 3% PPV, one patient in 33 flagged would truly have cancer. This sensitivity translates into identifying 62% of future lung cancer patients approximately four months earlier, although 40% would still be missed and a significant number flagged unnecessarily. Such trade-offs underline the importance of adjusting thresholds to match clinical context and resource availability.

Calibration plots demonstrated that predicted probabilities closely aligned with observed outcomes, suggesting that the model outputs are trustworthy for clinical interpretation. Although incorporating structured data alongside text did not significantly boost predictive performance, it offers flexibility for practices preferring a broader data spectrum. Nonetheless, the simplicity and performance of the text-only model make it an appealing option for integration into routine care.

Implications, Limitations and Future Perspectives

This proof-of-concept study offers valuable insights into how AI could reshape early lung cancer detection in primary care. With general practitioners acting as gatekeepers to specialist services, their electronic records are a critical yet underused resource. By automatically

© For personal and private use only. Reproduction must be permitted by the copyright holder. Email to copyright@mindbyte.eu.

analysing decades of patient data, AI systems can alert clinicians to subtle, cumulative risk indicators that might otherwise go unnoticed.

Nevertheless, implementation poses challenges. False positives could lead to unnecessary tests, patient anxiety and increased healthcare costs. Moreover, neural network models function as 'black boxes'—they do not explain which specific features contributed to a given prediction, which may hinder clinician trust. Additionally, international differences in healthcare systems, coding practices and data quality could limit the portability of the models without region-specific retraining and validation.

Privacy and data governance remain vital. Although the study complied with GDPR and anonymised data rigorously, real-world deployment would require stringent safeguards. There's also a need for ethical oversight to prevent misuse or overreliance on algorithmic predictions. Further research should explore the integration of this AI approach with clinician workflows and assess its impact on patient outcomes through prospective trials.

Despite these limitations, the method shows promise for other hard-to-detect cancers such as pancreatic, ovarian and oesophageal, which often present with vague symptoms. Expanding this work across larger datasets and international collaborations could help validate the approach for broader use. Ultimately, this AI tool is not a screening method but a support mechanism—an additional layer to help clinicians detect serious conditions earlier and improve prognosis through timely intervention.

The application of artificial intelligence to routine general practice records marks a significant step forward in the early detection of lung cancer. By using natural language processing to mine unstructured text, the models developed in this study demonstrated the potential to identify high-risk patients up to five months before formal diagnosis. While challenges around false positives, interpretability and implementation remain, the findings pave the way for smarter, more proactive primary care. Integrating AI into clinical workflows could enhance the diagnostic reach of general practitioners, ultimately improving survival outcomes for patients facing one of the most lethal cancers.

Source: [British Journal of General Practice](#)

Image Credit: [iStock](#)

Published on : Tue, 13 May 2025