EasyChair Preprint

№ 12990

Revolutionizing Personalized Medicine: a Comprehensive AI Tool for Lung Cancer Severity Prediction and Treatment Recommendation

Abhishek Shukla

April 10, 2024

Revolutionizing Personalized Medicine: A Comprehensive AI Tool for Lung Cancer Severity Prediction and Treatment Recommendation

Author : Abhishek Shukla

ABSTRACT

Lung cancer contributes a great percentage to the number of cancer-related deaths globally, which mandates science to find ways to improve its approaches to lung cancer diagnosis and treatment. Artificial Intelligence (AI) has emerged in recent times as one of the best solutions to lung cancer diagnosis and treatment. In this essay, attention will be paid to the current roles AIs are performing in lung cancer detection and treatment. Huge successes have been recorded in the use of radiomics, deep learning, and machine learning in lung cancer screening, diagnosis, and treatment. AI has assisted healthcare professionals to better characterize cancer cells and enable them to make better choices regarding treatment procedures. AI has contributed tremendously to the improvement in imaging modalities, including PET-CT imaging, Chest radiography, low-dose CT scans, etc. It also enables healthcare professionals to detect tumor markers and biomarkers in affected patients for a better treatment procedure. However, there is room for improvement. Further studies into the field of AI in lung cancer treatment can help reduce morbidity, mortality, and other potential outcomes.

KEYWORDS

Artificial intelligence, treatment, screening, radiomics, machine learning, lung cancer, diagnosis, deep learning

INTRODUCTION

Lung cancer occurs in both genders with the prevalence being slightly higher in men. Studies revealed that a man’s chance of developing lung cancer in his lifetime is 1:16,

while that of a woman is 1:17. Lung cancer development is also not limited to smokers; even nonsmokers can become victims; only the risk is higher in smokers by about 90% (Ozlü and Bulbul, 2005).

Lung cancer risk is equally higher in black men by about 12%, while the rate is discovered to be 16% lower in black women than their white counterparts (Arslaan, 2020). In general, the risk is higher in black and white men compared to women of either skin color. Exposure to information has

led to a drastic reduction in lung cancer cases among men but women only started recording a reduction in rate in the past decade (vainshelboim et al., 2018).

The type of lung cancer reported can determine the survival rate in individuals suffering from it (Kabir et al., 2020). The stage of detection before treatment implementation is yet another survival-rate determining factor, not also forgetting the treatment method adopted by healthcare professionals. For example, treatment methods for small-cell lung carcinomas (SCLC) differ from those for non-small-cell lung carcinomas (NSCLC) (Krpina, 2023).

Some of the popularly adopted treatment methods are:

- Surgery
- Radiation therapy
- Chemotherapy
- Targeted therapy (Rehman, 2018)

As mentioned above, the best option for individual patients depends on the factors mentioned above. A huge progress has been recorded over the years in lung cancer treatment but the prognosis is yet unsatisfactory with patients' responses to existing treatment methods being poor (Karadog˘an and Ünsal, 2023).

The problem may not be with the existing lung cancer treatment methods but with the diagnosis. An improper diagnosis can result in the wrong choice of a treatment method for the patient. Cancer is a consequence of genetic (Gene) mutation and a good understanding of the mutation and its associated processes is required for choosing a treatment method for the specific patient.

Science is ever-evolving, which brought to light the involvement of AI in lung cancer treatment. With AI (Artificial Intelligence), healthcare professionals can easily read imaging from lung cancer patients, which will enable a better understanding of the pathology for a more effective diagnosis, which can enable a better treatment procedure and improve patients' prognosis (Dlamini et al., 2020). Genetic Mutation Analysis is an AI tool with a difference. Although new, it has undergone rigorous testing over time, consistently yielding satisfactory results. To improve the results it can generate, the AI tool works with the already existing WSI (Wang et al., 2019), ensuring a faster and more accurate diagnosis of lung cancer towards improving the patient's prognosis.

In this report, attention will be paid to the impact of WSI (AI tool) in lung cancer diagnosis and treatment. Studies show that the introduction of AI to WSI in digital pathology will increase the pathologists' Kappa value for better molecular phenotype prediction using H&E staining and radiomics (Chen et al., 2022). We will also reveal why WSI is not adequately effective and how combination with the GMA AI tool can improve the result generated.

Objectives:

- To examine the effectiveness of Wide Slide Imagining (WSI) as the current AI tool adopted for lung cancer diagnosis and treatment.
- To analyze the limitations of using only WSI in lung cancer treatment
- To justify the improved effectiveness of combining WSI with the Genetic Mutation Analysis AI tool

Research question

This essay addresses the research questions stated below to achieve the aforementioned goals:

• Wide Slide Imaging (WSI) AI tool has played a unique role in the improvement of lung cancer screening, diagnosis, and treatment. How effective has this AI tool been to date?

• Using only WSI has proved to be ineffective for lung cancer diagnosis in recent times. What are the challenges associated with the use of only the WSI tool?

• The inadequacies of WSI compel scientists to consider an AI combination system (involving WSI and gene mutation analysis tool) for cancer screening and treatment. In which areas will the new AI system resulting from the combination improve lung cancer screening, diagnosis, and treatment?

LITERATURE REVIEW

AI improves all the stages involved in gene mutation status prediction, including:

- Multimodal data curation
- Feature extraction
- Genetic alterations analysis (Saitou, 2018).

Certain limitations still exist, but combining Wide Slide Imaging and gene mutation analysis AI tools have proved considerably progressive in lung cancer detection and treatment.

Whole genome analysis

The AI system enables clinicians to analyze the whole gene since it amplifies a series of technologies, including Polymerase chain reaction (PCR). The system also promotes easy analysis of cancer epigenome, transcriptome, and exome, as well as, produces large sets of information about the tumor for better treatment decision-making.

[Figure 1 about here.]

The various growth patterns identified by Artificial Intelligence and Machine Learning tools in lung adenocarcinoma following the evaluation of prognostic molecular markers (thyroid transcription factor-1 and FILM signature index) are:

- Lepidic
- Acinar
- Papillary
- Micropapillary
- Solid growth patterns

Solid growth pattern indicates poor prognosis in primary lung adenocarcinoma. On the flip side, patients, who had tumors with a non-solid growth pattern, will have a better prognosis if the TTF-1 expression and FILM signature index are high and lower respectively (Marjolein et al., 2017).

AI/ML tools assist in subtyping lung cancers based on the number of bronchioalveolar carcinoma components present in the cancer cells. The four subtypes thus identified are:

• **Group I:** Pure or Predominant Bronchioalveolar carcinoma (BAC) component. It mainly comprises bronchioalveolar carcinoma, which is non-invasive and features a pure lepidic growth pattern. Its prognosis is excellent at a 100% 5-year survival rate.

• **Group II:** Mixed subtype with predominant bronchioalveolar carcinoma component and ≤5mm invasive component. It is characterized by a limited invasive component featuring a 5mm maximum size, which indicates that the tumors have a small invasive area. The growth pattern may be beyond lepidic but it is still non-invasive predominantly. It is also called a minimally invasive adenocarcinoma (Yim et al., 2007).

• **Group III:** Mixed subtype with bronchioalveolar carcinoma component >5mm invasive component. It is called "Mixed" because it consists of non-mucinous and mucinous bronchoalveolar carcinoma. It can also be called invasive adenocarcinoma. The mucinous bronchoalveolar carcinoma is a form of invasive adenocarcinoma, while the non-mucinous carcinoma is a form of Lepidic-predominant adenocarcinoma (Jones, J. 2024).

• **Group IV:** Invasive carcinoma with no bronchioalveolar carcinoma component. The cancer cells have already invaded other tissues surrounding the lungs but bronchoalveolar carcinoma is absent. Some of the factors determining prognosis in this group include specific characteristics of the tumor and the overall health of the particular patient (Dumont et al., 1998).

[Figure 2 about here.]

In groups I and II, the reported death proportion is 0%; that of group III is 20% due to the presence of predominant invasive components; that of group IV is 18% due to lack of bronchioalveolar carcinoma components.

Some of the mutations the system can detect via machine-learning approaches are:

- EGFR
- TP53 mutation
- KRAS mutation
- EML4-ALK mutation
- RET
- Neurotrophic tyrosine receptor kinase (NTRK)
- Human epidermal growth factor receptor 2 gene mutation
- BRAF mutations

The AI system will detect with impressive accuracy the pathogenicity of missense variants, in which there is an alteration of a single letter in the genetic code.

In addition, it accurately predicts the known variant pathogenicity associated with Asparagine Synthetase Deficiency (ASNSD) and other related conditions. It, therefore, offers an insight into how each of the mutation variants affects protein function, which will enable clinicians to determine pathogenic missense mutation with ease (Wang T et al., 2024).

Each of the mutation types highlighted above has a specific severity factor, which can assist the AI in determining the best befitting treatment method. EGFR exon 19 mutation, a type of EGFR

mutations, showcases higher sensitivity to EGFR-TKIs than any other EGFR mutations. So, this type is higher on the severity scale than many others (Hong et al., 2019).

[Figure 3 about here.]

BRAF mutation, especially the BRAF V600E variant, equally leads on the severity scale, which negatively impacts its prognosis.

Machine learning models for predicting analysis in lung cancer using the new AI system

AI can better analyze mutations using a combination of specific machine learning models and the severity factor for each type of mutation. The data on machine learning models and severity factors will be fed into the system for each type of mutation. The system will now combine the severity impact of each model and use the information derived to generate an analysis of the best treatment options.

Factors determining individual mutation risk severity factor include the following:

- Tumor characteristics,
- Survival time
- Previous treatment outcomes (Bartholomai et al., 2019).

The factors above form the basis for the choice of machine learning models utilized by the new AI system to provide reliable information on the best treatment option for the patient.

The two machine-learning models involved in the process are

- Machine learning regression model and
- Classification model.

The new AI system combines a specific model with data from severity factors for each type of mutation to improve prediction accuracy on survival time, cancer recurrence, and susceptibility, thereby enabling healthcare professionals to make better treatment decisions. Combining the two models with the severity factor data can accurately predict lung cancer survival time in months. The classification model is developed using Random Forests, while the regression model is developed using Random Forests, gradient-boosted machines (GBM), and General Linear Regression. (Bartholomai et al., 2019).

RESULTS

WSI effectiveness

WSI stands for whole Slide Image. The technology utilizes digital pathological scanning systems to convert traditional pathological slices to high-resolution images for a better understanding by pathologists. The output consists of fragmented images, which are combined by a computer to form a complete image, preventing image fading, and loss and promoting preservation. Thanks to WSI, healthcare professionals can save the digital scan of an entire histology result, enabling them to easily analyze a larger area for easier detection of mutations, including malignant cells in any organ tissue.

WSI is equally involved in the homogenization of digital images, along with other preprocessing

activities. Its high efficiency is not limited by sequencing either.

WSI limitations

One of the major problems of using WSI in lung cancer diagnosis is that achieving uniform batch variation is difficult and this is caused by time differences, staining method differences, and reagent differences. Using only WSI leaves healthcare professionals to work with poor-quality images since the technology is considerably limited. Scanning all the materials at a go is also not possible. The cost is high and there is a problem with digital slide storage. In addition, healthcare professionals are unable to handle high-throughput routine work and may be faced with regulatory barriers from healthcare industry authorities (Kumar et al., 2020).

WSI + Gene mutation analysis tool

Combining WSI with gene mutation analysis AI tools forms a new AI system that helps prevent many of the limitations associated with the lung cancer detection process using WSI only. AlphaMissense is an outstanding tool for this, especially since it works with WSI for better lung cancer detection and reduced risk to the patients.

Combining WSI and AlphaMissense enables better prediction of the impact of genetic mutation on human health. The tool is developed by Google DeepMind and it works by leveraging the AlphaFold network, enabling an accurate prediction of the particular protein that can culminate in disease. Combining the two AI tools also improves differentiation between carcinogenic and benign growth in the lungs with up to 90% accuracy. It does this by assigning unique risk scores to all detected mutations. Consequently, clinicians will not have problems interpreting the resulting genetic data for easy cause detection (Qu H et al., 2021).

Aside from differentiating between carcinogenic and benign cells, the WSI-AlphaMissense combination can predict the stages of cancer growth and also determines recurrence risk in early-stage lung Adenocarcinoma by analyzing EGFR mutations and PD-L1 expression. It equally enables easy radiomics extraction in histopathology.

Furthermore, the WSI-AlphaMissense combination (the new AI system) offers clinicians more information, giving them a wider viewpoint on lung cancer cases, including patient's clinical data and demographics. With the variety of clinical information made available to clinicians by the AI system, better decisions can now be made regarding treatment options for the patient.

After helping clinicians choose the best treatment option, the AI system can also assist in predicting a patient's response to treatment. In addition, it can predict side effects, radiotherapy response, and even tumor recurrence rate following surgery.

METHODOLOGY

We will begin by examining multiple datasets and then applying the proposed regression-based AI tools alongside the current WSI AI tool to accurately classify lung nodules. This process starts with gathering comprehensive patient data, which includes genetic mutations like EGFR, TP53, and KRAS, along with cancer severity metrics and treatment outcomes. We will employ feature

engineering techniques to extract relevant information from the genetic data, transforming categorical mutations into numerical representations using methods such as one-hot encoding. We will then individually apply various regression models—Linear, Polynomial, Decision Tree, and Support Vector Regression—to each genetic mutation to predict its severity factor and understand its impact on cancer progression. The dataset will be split into training and testing sets for model validation, with each regression model fine-tuned on training data to optimize parameters. We will assess model accuracy on the testing set using performance metrics like Mean Squared Error or R-squared.

The study will meticulously evaluate the performance of the new proposed AI tool against traditional ML algorithms across the datasets, focusing on metrics such as accuracy, sensitivity, and specificity. This cross-dataset analysis aims to benchmark the current state-of-the-art in lung nodule classification and uncover potential dataset biases and challenges inherent in AI-driven medical imaging analysis.

Through a detailed examination of the classification capabilities of AI models, our research aims to advance the understanding of lung tumor characteristics and improve the accuracy of lung cancer diagnostics. This work builds upon and extends the foundation laid by previous studies, and research into the application of the proposed new AI tool in combination with existing WSI AI tools in the detection and classification of pulmonary nodules. It sets new benchmarks and expands the knowledge frontier in the application of AI in lung cancer diagnostics. The primary objective is to conduct a comprehensive comparison of Machine Learning (ML) and Deep Learning (DL) techniques in the accurate classification of lung nodules, which plays a pivotal role in the diagnosis, prognosis, and treatment planning for lung cancer patients. By integrating various ML and DL methodologies, this research aims to identify the most effective approaches for distinguishing between benign and malignant lung nodules, thereby enhancing the precision of computer-aided diagnostic systems.

Through a detailed examination of the classification capabilities of AI models, our research endeavors to advance the understanding of lung tumor characteristics and improve the accuracy of lung cancer diagnostics.This work builds upon and extends the foundation laid by previous studies, such as the LUNGx Challenge for computerized lung nodule classification and research into the application of CNNs in the detection and classification of pulmonary nodules, setting new benchmarks and expanding the knowledge frontier in the application of AI in lung cancer diagnostics.

The study meticulously evaluates the performance of new AI tool against traditional ML algorithms across the mentioned datasets, focusing on metrics such as accuracy, sensitivity, and specificity. This cross-dataset analysis not only benchmarks the current state-of-the-art in lung nodule classification but also seeks to uncover potential dataset biases and challenges inherent in AI-driven medical imaging analysis.

CONCLUSION

The medical world has advanced tremendously but there is always room for improvement. Lung cancer detection and diagnosis are far beyond where they were in times past and the introduction of artificial intelligence further improves diagnosis, treatment success, and prognosis. The attempt to build on the achieved success resulted in the development of a special AI system, which includes WSI and Genetic Mutation Analysis AI tools. The new AI system thus developed will succeed where the previous systems have failed, enabling healthcare professionals to diagnose lung cancer better and choose the best treatment option for the patient. It will also help determine survival time faster for a better prognosis.

REFERENCES

Arslaan, R. (2020). Racial Disparities of Lung Cancer: A Systematic Review.

Chen, Y., Yang, Cheng, H., Chen, Z., Peng, L., Sui, Wang, J., Yang, Lin, M., Chen, C., Wang, Y., Huang, Y., Chen, L., Li, Y., Ke, W., and Zunfu (2022). A whole-slide image (WSI)--based immunohistochemical feature prediction system improves the subtyping of lung cancer. Lung Cancer. 165. 10.1016/j.lungcan.2022.01.005.

Dlamini, Z., Francies, F., Hull, R., and Marima, R. (2020). Artificial intelligence (AI) and big data in cancer and precision oncology. *Computational and Structural Biotechnology Journal*.

Dumont, P., Gasser, B., Rougé, Massard, G., Jean-Marie, and Wihlm (1998). Bronchoalveolar carcinoma - Histopathologic study of evolution in a series of 105 surgically treated patients. *Chest*, 113.

Hong, W., Wu, Q., Zhang, J., and Zhou, Y. (2019). Prognostic value of EGFR 19-del and 21-L858R mutations in patients with non-small cell lung cancer. *Oncology letters*, 18(4):3887–3895.

Kabir, R., Haque, M., Roy, R., Setu, S., Paul, and Huda (2020). Factors affecting the survival time of cancer patient: a hospital registry-based study. *Eurasian Journal of Biosciences*, 14:7817–7824.

Karadogan, D. and Ünsal, M. (2023). Prognosis and Prognostic Factors of Lung Cancer.

Kumar, N., Gupta, R., and Gupta, S. (2020). Whole Slide Imaging (WSI) in Pathology: Current Perspectives and Future Directions. *Journal of Digital Imaging*. 33. 10.1007/s10278-020-00351-z.

Marjolein, A., Heuvelmans, R., Vliegenthart, H. J., Koning, D., Harry, J. M., Groen, Michel, J. A. M., Putten, U. V., Yousaf-Khan, C., Weenink, K., Nackaerts, Pim, A., Jong, M. D., and Oudkerk (2017). Quantification of growth patterns of screen-detected lung cancers: The NELSON study. *Lung Cancer*, 108:48–54.

Wang, S., Yang, D. M., Rong, R., Zhan, X., Fujimoto, J., Liu, H., Minna, J., Wistuba, I. I., Xie, Y., Xiao, and G (2019). Artificial Intelligence in Lung Cancer Pathology Image Analysis. *Cancers*, 11(11):1673–1673.

Yim, J., Zhu, L. C., Chiriboga, L., Watson, H. C., Goldberg, J. D., and Moreira, A. L. (2007).

Vainshelboim, B., Lima, R., Edvardsen, E., and Myers, J. (2018). Cardiorespiratory fitness, incidence and mortality of lung cancer in men: A prospective cohort study. *Journal of Science and Medicine in Sport*.

List of Figures

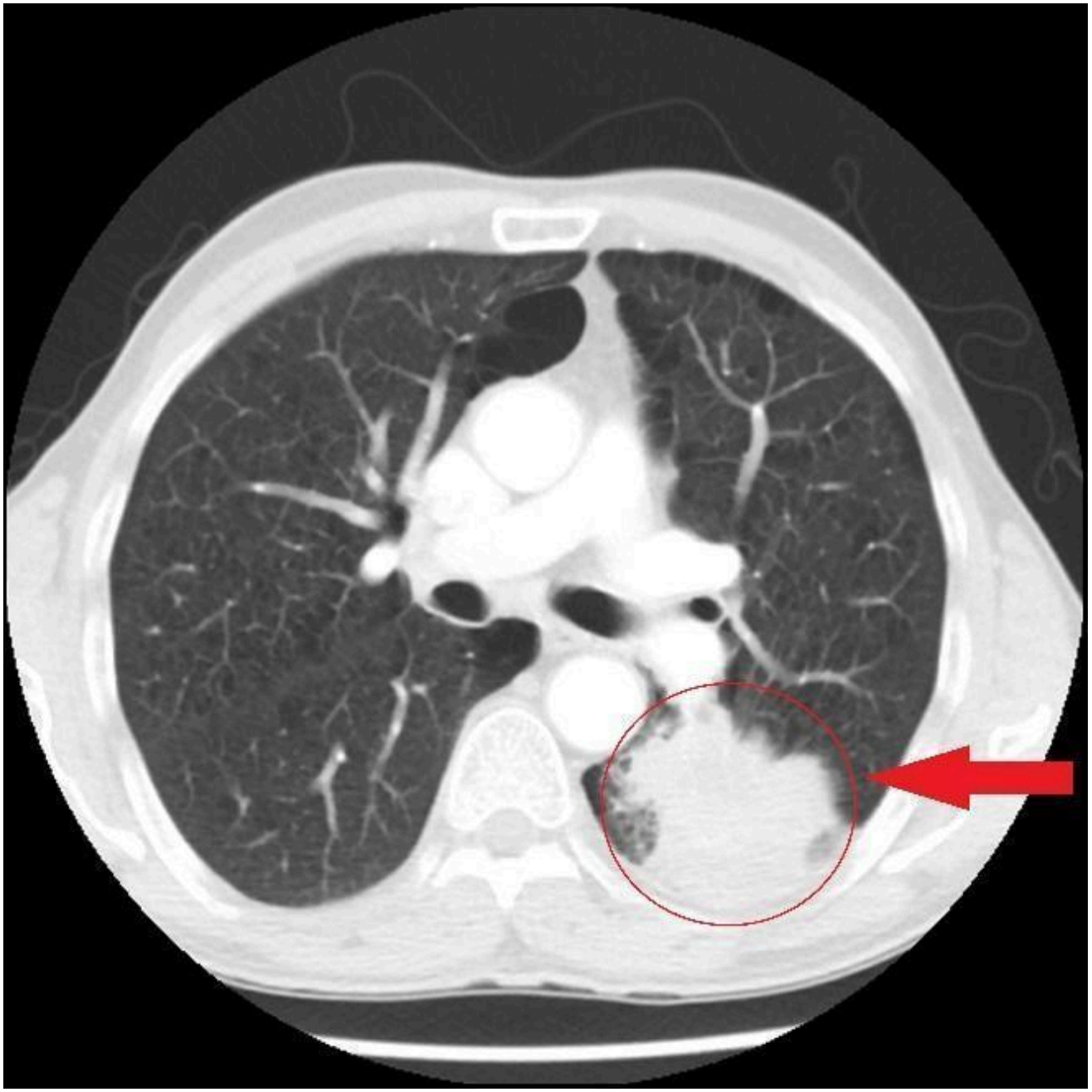

Figure 1. Lung adenocarcinoma

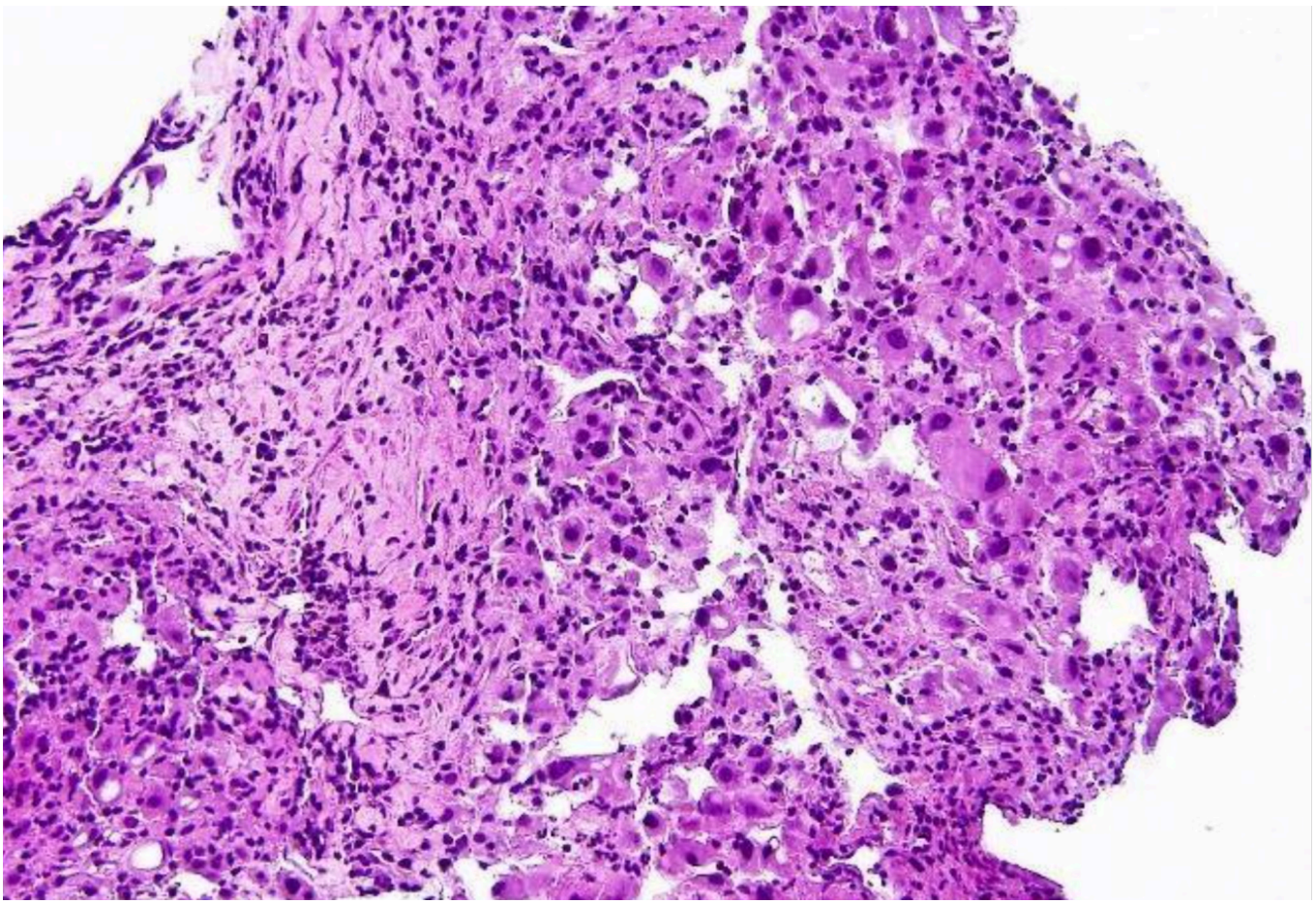

Figure 2. Minimally invasive adenocarcinoma (Group II).

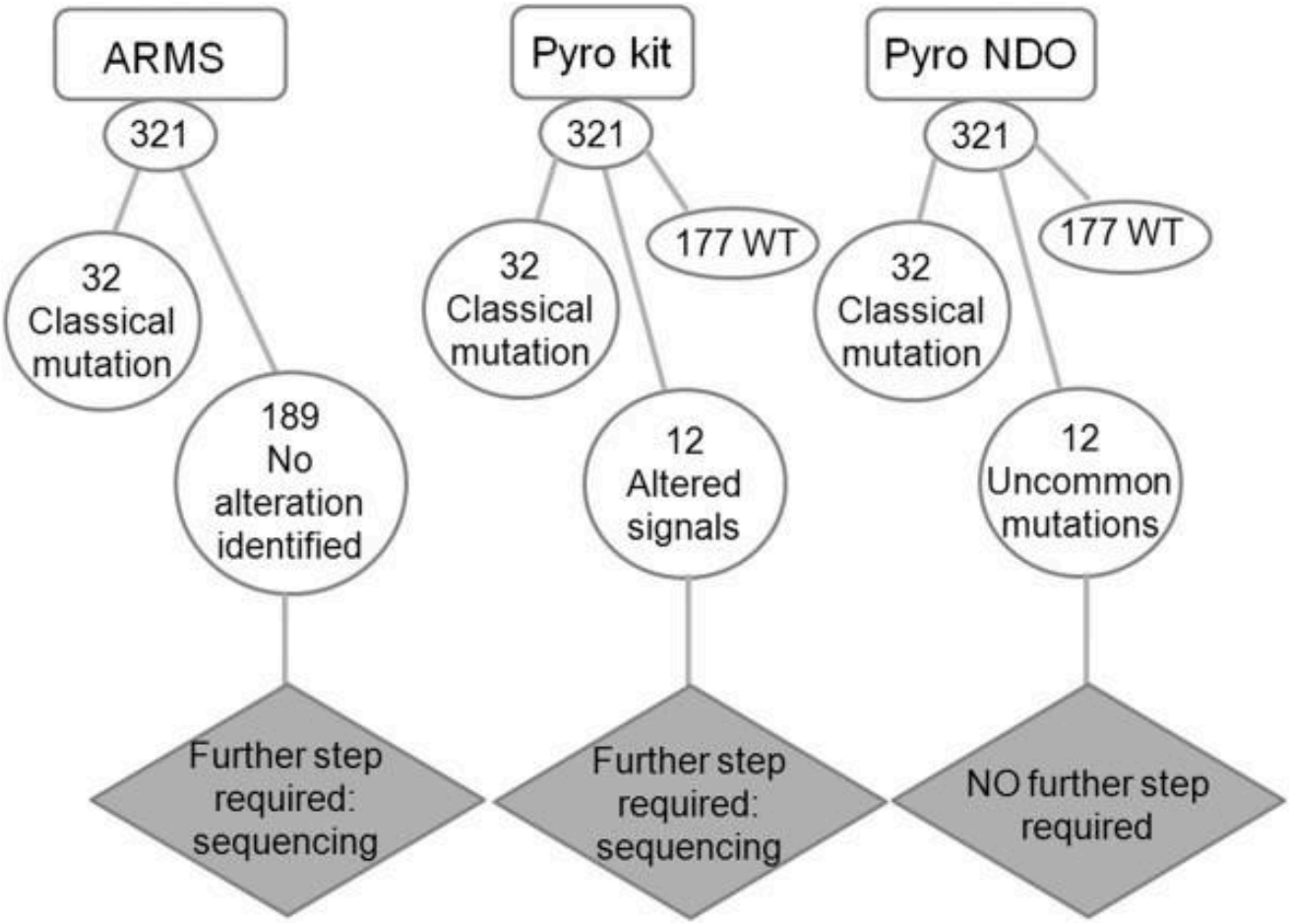

Figure 3. EGFR exon 19 mutation