

A Comprehensive Artificial Intelligence Tool for Lung Cancer Severity Prediction and Treatment Recommendation

Abhishek Shukla
Syracuse University
Austin, USA

Sindhu C. Pokhriyal, PhD
One Brooklyn Health, Brooklyn
USA

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.

General Terms

Treatment, Screening, Radiomics, Diagnosis.

Keywords

Artificial Intelligence, Machine Learning, Image Processing, Lung Cancer, Deep Learning.

1. INTRODUCTION

Lung cancer occurs in both sexes, and the incidence of lung cancer is slightly greater in men. A man's lifetime chance of developing lung cancer is 1:16, while that of a woman is 1:17. Lung cancer development is also not limited to smokers; even if nonsmokers can become victims, only the risk is increased by approximately 90% [1].

The risk of lung cancer is approximately 12% greater in black men than in white men, while the risk is 16% lower in black women than in their white counterparts [2]. In general, the risk is greater for black and white men than for women with either skin color. Exposure to information has

led to a drastic reduction in lung cancer cases among men, but women have only started recording a reduction in the rate of lung cancer in the past decade [3].

The type of lung cancer reported can determine the survival rate in individuals with this disease [4]. The stage of detection before treatment implementation is yet another factor determining survival rate, as is the treatment method adopted by healthcare professionals. For example, treatment methods

for small-cell lung carcinomas (SCLCs) differ from those for non-small cell lung carcinomas (NSCLCs) [5].

Some of the popularly adopted treatment methods are as follows:

- Surgery
- Radiation therapy
- Chemotherapy
- Targeted therapy [6]

As mentioned above, the best option for individual patients depends on the factors mentioned above. Considerable progress has been made in lung cancer treatment, but the prognosis is unsatisfactory, as patients' responses to existing treatment methods are poor [7].

The problem may not be related to the existing lung cancer treatment methods but rather to the diagnosis. An improper diagnosis can result in the wrong choice of 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 a specific patient.

Science is ever-evolving, which has brought to light the involvement of AI in lung cancer treatment. With artificial intelligence (AI), healthcare professionals can easily read images from lung cancer patients, which will enable a better understanding of the pathology for a more effective diagnosis and can enable a better treatment procedure and improve patient prognosis [8]. Genetic mutation analysis is an AI tool for identifying differences. Although new, it has undergone rigorous testing over time, consistently yielding satisfactory results. To improve the results, the AI tool works with the existing WSI [9], ensuring a faster and more accurate

diagnosis of lung cancer to improve patient 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 [10]. Report will also reveal why WSI is not adequately effective and how combination with the GMA AI tool can improve the result generated.

1.1 Objectives

To examine the effectiveness of Wide Slide Imaging (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

1.2 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?

2. Literature Review

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

- Multimodal data curation
- Feature extraction
- Genetic alterations analysis [11].

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.

2.1 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.

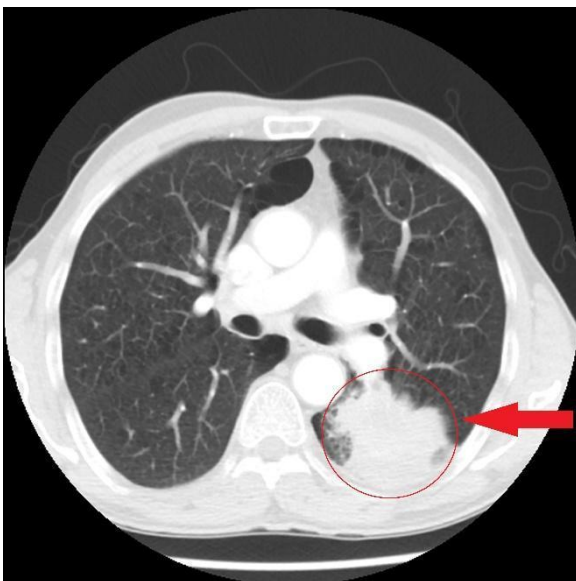


Figure 1: Lung adenocarcinoma

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 [12].

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 $\leq 5\text{mm}$ 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 [13]. As shown in figure 2.

Group III: Mixed subtype with bronchioalveolar carcinoma component $> 5\text{mm}$ invasive component. It is called "Mixed" because it consists of non-mucinous and mucinous bronchioalveolar carcinoma. It can also be called invasive adenocarcinoma. The mucinous bronchioalveolar carcinoma is a form of invasive adenocarcinoma, while the non-mucinous carcinoma is a form of Lepidic-predominant adenocarcinoma [14].

Group IV: Invasive carcinoma with no bronchioalveolar carcinoma component. The cancer cells have already invaded other tissues surrounding the lungs but bronchioalveolar 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 [15].

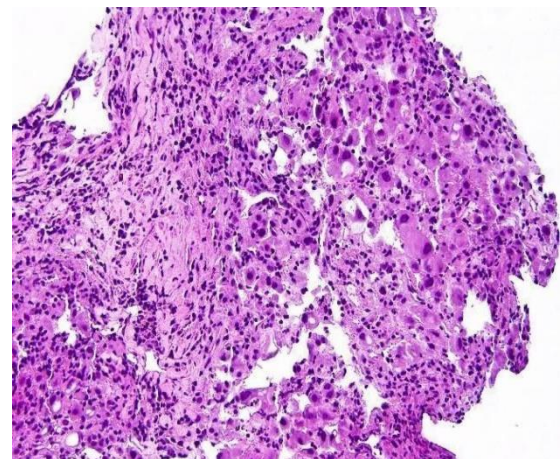


Figure 2: Minimally invasive adenocarcinoma (Group II)

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

The following mutations can be detected via machine learning approaches:

- EGFR
- Neurotrophic tyrosine receptor kinase (NTRK)
- Human epidermal growth factor receptor 2 (EGFR2) gene mutation

The AI system detects the pathogenicity of missense variants with impressive accuracy when there is an alteration of a single letter in the genetic code.

In addition, this approach accurately predicts the known variant pathogenicity associated with asparagine synthetase deficiency (ASNSD) and other related conditions. Therefore, this approach offers insight into how each of the mutation variants affects protein function, which will enable clinicians to determine pathogenic missense mutations with ease [16].

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

EGFR exon 19 mutation, a type of EGFR mutations and greater sensitivity to EGFR-TKIs than to any other EGFR mutation. Therefore, this type of injury is more common on the severity scale than many others [17].

The presence of BRAF mutations, especially the BRAF V600E variant, is also related to disease severity and negatively impacts patient prognosis.

3. 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 [18].

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 [18].

4. RESULTS

4.1 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.

4.2 WSI limitations

One of the major problems of using WSIs in lung cancer diagnosis is that achieving uniform batch variation is difficult and is caused by time differences, staining method differences, and reagent differences. Using only WSIs leaves healthcare professionals working with poor-quality images since the technology is considerably limited. Scanning all the materials at one 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 perform high-throughput routine work and may be faced with regulatory barriers from healthcare industry authorities [19].

4.3 WSI + Gene mutation analysis tool

Combining the WSI with gene mutation analysis AI tools could lead to the development of a new AI system that helps prevent many of the limitations associated with the lung cancer detection process using the WSI only. AlphaMissense is an outstanding tool for this purpose, especially since it works with the WSI to improve lung cancer detection and reduce patient risk.

The combination of the WSI and AlphaMissense data enables better prediction of the impact of genetic mutations on human health. The tool was developed by Google DeepMind and works by leveraging the AlphaFold network, enabling accurate prediction of particular proteins that can culminate in disease. Combining these two AI tools also improved the differentiation between carcinogenic and benign growth in the lungs with up to 90% accuracy. This is accomplished by assigning unique risk scores to all the detected mutations. Consequently, clinicians will not have problems interpreting the resulting genetic data for easy cause detection [20].

In addition to differentiating between carcinogenic and benign cells, the WSI-AlphaMissense combination can predict the stage of cancer growth and determine recurrence risk in early-stage lung adenocarcinoma by analyzing EGFR mutations and PD-L1 expression. Moreover, radiomics extraction is easy to perform via histopathology.

Furthermore, the WSI-AlphaMissense combination (the new AI system) offers clinicians more information, providing them with a wider view of lung cancer cases, including patient clinical data and demographic information. 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, the NRS-2002 can predict side effects, radiotherapy response, and even the tumor recurrence rate following surgery.

5. METHODOLOGY

The examination commences with the analysis of multiple datasets, followed by the application of the proposed regression-based AI tools in conjunction with the current WSI AI tool to accurately classify lung nodules. This process commences with the collection of comprehensive patient data, encompassing genetic mutations, such as those in EGFR, TP53, and KRAS, alongside cancer severity metrics and treatment outcomes. Feature engineering techniques are employed to extract relevant information from genetic data, converting categorical information into numerical representations through methods such as one-hot encoding. Subsequently, various regression models—linear, polynomial, decision tree, and support vector regression—are individually applied to each genetic mutation to forecast its severity factor and comprehend its influence on cancer progression. The dataset undergoes division into training and testing sets for model validation, with each regression model fine-tuned on the training data to optimize the parameters. Model accuracy on the testing set is evaluated using performance metrics such as the mean squared error (MSE) or R-squared.

This study will meticulously evaluate the performance of the newly proposed AI tool against traditional ML algorithms across datasets, focusing on metrics such as accuracy, sensitivity, and specificity. This cross-dataset analysis aimed 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 aimed to improve 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 studies into the application of the proposed new AI tool in combination with existing WSI AI tools for the detection and classification of pulmonary nodules. This approach sets new benchmarks and expands the knowledge frontier in the application of AI in lung cancer diagnostics.

6. CONCLUSION

The medical world has advanced tremendously, but there is always room for improvement. Historically, however, lung cancer detection and diagnosis have advanced greatly beyond their initial use, and the introduction of artificial intelligence has further improved 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 (GA) 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.

Future work can focus on integrating this AI system with real-time patient data from electronic health records (EHRs) for continuous monitoring and timely updates on patient status and treatment efficacy. Additionally, the system can be enhanced to

develop highly personalized treatment plans based on the unique genetic and molecular profiles of individual patients, optimizing drug regimens and dosages to maximize effectiveness while minimizing side effects. Expanding the predictive analytics capabilities to include comprehensive models for predicting long-term outcomes, potential complications, and recurrence likelihood can provide valuable insights for sustained patient care. Adapting the principles and technologies of this AI system to other types of cancer and diseases can broaden its impact across various medical fields. Furthermore, promoting collaborative research initiatives involving AI experts, oncologists, geneticists, and other healthcare professionals can lead to the development of more robust and effective AI tools.

7. REFERENCES

- [1] Ozlü, Tevfik and Bulbul, Yılmaz. (2005). Smoking and lung cancer. *Tüberküloz ve toraks*. 53. 200-9.
- [2] Arslan, Ryan. (2020). Racial Disparities of Lung Cancer: A Systematic Review.
- [3] Vainshelboim, Baruch and Lima, Ricardo and Edvardsen, Elisabeth and Myers, Jonathan. (2018). Cardiorespiratory fitness, incidence and mortality of lung cancer in men: A prospective cohort study. *Journal of Science and Medicine in Sport*. 22. 10.1016/j.jsams.2018.10.002.
- [4] Kabir, Rasel and Haque, Munima and Roy, Raju and Setu, Sarmistha Paul and Huda,. (2020). Factors affecting the survival time of cancer patients: a hospital registry-based study. *Eurasian Journal of Biosciences*. 14. 7817-7824.
- [5] Krpina, K., Vranić, S., Tomić, K., Samaržija, M., and Batičić, L. (2023). Small Cell Lung Carcinoma: Current Diagnosis, Biomarkers, and Treatment Options with Future Perspectives. *Biomedicine*, 11(7), 1982. <https://doi.org/10.3390/biomedicine11071982>
- [6] Rehman, Sakina. (2018). An Overview of Cancer Treatment Modalities. *10.5772/intechopen.76558*.
- [7] Karadoğan, Dilek and Ünsal, Meftun. (2023). Prognosis and Prognostic Factors of Lung Cancer. *10.1007/978-3-031-22483-6_75-1*.
- [8] Dlamini, Zodwa and Francies, Flavia and Hull, Rodney and Marima, Rahaba. (2020). Artificial intelligence (AI) and big data in cancer and precision oncology. *Computational and Structural Biotechnology Journal*. 18. 10.1016/j.csbj.2020.08.019.
- [9] Wang, S., Yang, D. M., Rong, R., Zhan, X., Fujimoto, J., Liu, H., Minna, J., Wistuba, I. I., Xie, Y., and Xiao, G. (2019). Artificial Intelligence in Lung Cancer Pathology Image Analysis. *Cancers*, 11(11), 1673. <https://doi.org/10.3390/cancers11111673>
- [10] Chen, Yanyang and Yang, Huan and Cheng, Zhiqiang and Chen, Lili and Peng, Sui and Wang, Jianbo and Yang, Minglei and Lin, Chenghao and Chen, Yu and Wang, Yuefeng and Huang, Leilei and Chen, Yangshan and Li, Weizhong and Ke, 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.
- [11] Saitou, Naruya. (2018). Mutation. *10.1007/978-3-319-92642-1_3*.
- [12] Marjolein A. Heuvelmans, Rozemarijn Vliegthart, Harry J. de Koning, Harry J.M. Groen, Michel J.A.M. van

- Putten, Uraujh Yousaf-Khan, Carla Weenink, Kristiaan Nackaerts, Pim A. de Jong, Matthijs Oudkerk (2017). Quantification of growth patterns of screen-detected lung cancers: The NELSON study, *Lung Cancer*, Volume 108, Pages 48-54, ISSN 0169-5002, <https://doi.org/10.1016/j.lungcan.2017.02.021>.
- [13] Yim, J., Zhu, L.-C., Chiriboga, L., Watson, H. C., Goldberg, J. D., and Moreira, A. L. (2007). Histologic features are important prognostic indicators in the early stages of lung adenocarcinomas. *20(2)*, 233–241. <https://doi.org/10.1038/modpathol.3800734>
- [14] Jones, J. (2024). Adenocarcinoma in situ, minimally invasive adenocarcinoma and invasive adenocarcinoma of lung | Radiology Reference Article | Radiopaedia.org. Radiopaedia. Retrieved <https://radiopaedia.org/articles/adenocarcinoma-in-situ-minimally-invasive-adenocarcinoma-and-invasive-adenocarcinoma-of-lung-1>
- [15] Dumont, P and Gasser, Bernard and Rougé, C and Massard, Gilbert and Jean-Marie, Wihlm. (1998). Bronchoalveolar carcinoma - Histopathologic study of evolution in a series of 105 surgically treated patients. *Chest*. 113. 391-5. [10.1378/chest.113.2.391](https://doi.org/10.1378/chest.113.2.391).
- [16] Wang T, Ma J, Shu Y, Hong B, Wang Z, Lu Y, Yu X, Huang X, Zhang Y (2024). Predicting valuable missense variants with AlphaMissense in a multiple pulmonary infection patient. *Clin Case Rep*. 2024 Jan 30;12(2):e8453. doi: [10.1002/ccr3.8453](https://doi.org/10.1002/ccr3.8453). PMID: 38292225; PMCID: PMC10825880.
- [17] 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. <https://doi.org/10.3892/ol.2019.10715>
- [18] Bartholomai JA, Frieboes HB. (2019). Lung Cancer Survival Prediction via Machine Learning Regression, Classification, and Statistical Techniques. *Proc IEEE Int Symp Signal Proc Inf Tech*. 2018 Dec;2018:632-637. doi: [10.1109/ISSPIT.2018.8642753](https://doi.org/10.1109/ISSPIT.2018.8642753). Epub. PMID: 31312809; PMCID: PMC6634305.
- [19] Kumar, Neeta and Gupta, Ruchika and Gupta, Sanjay. (2020). Whole Slide Imaging (WSI) in Pathology: Current Perspectives and Future Directions. *Journal of Digital Imaging*. 33. [10.1007/s10278-020-00351-z](https://doi.org/10.1007/s10278-020-00351-z).
- [20] Qu H, Zhou M, Yan Z, Wang H, Rustgi VK, Zhang S, Gevaert O, Metaxas DN. (2021). Genetic mutation and biological pathway prediction based on whole slide images in breast carcinoma using deep learning. *NPJ Precis Oncol*. doi: [10.1038/s41698-021-00225-9](https://doi.org/10.1038/s41698-021-00225-9). PMID: 34556802; PMCID: PMC8460699.