# Saving Lives: Earlier Lung Cancer Detection Using Machine Learning Techniques

Sai Siddhish Chandra Sekaran

Centennial High School

GT Intern/Mentor

Mrs. Ireland

May 20, 2025

**Abstract**

**Saving Lives: Earlier Lung Cancer Detection Using Machine Learning Techniques**

Sai Siddhish Chandra Sekaran
2024 - 2025 Academic Year

Lung cancer remains one of the leading causes of cancer-related deaths globally, taking over two million lives annually (World Health Organization, 2023). Early detection is critical for improving survival rates, yet traditional diagnostic tools such as CT scans often lack the accuracy needed for a timely and effective diagnosis. Recent advancements in Artificial Intelligence (AI) and Machine Learning (ML) have shown potential in medical imaging, offering enhanced detection capabilities. Among these, ensemble modeling—a method that combines multiple AI models—has been demonstrated as a particularly effective yet underused approach. By relying on predictions from multiple different models, ensemble modeling techniques improve diagnostic reliability, reduce false positives and negatives, and offer a more robust method for early lung cancer detection. To test the potential impact of ensemble modeling, an experiment was conducted using various Convolutional Neural Networks (CNNs) with and without ensemble modeling to compare performance across accuracy, precision, recall, and F1 score metrics. Results showed a slight improvement in accuracy when ensemble modeling was applied. These findings imply that even small gains in predictive capabilities can contribute to more reliable lung cancer detection, especially in reducing misdiagnoses. While the results are based on a specific dataset and model architecture, they reinforce the broader potential of ensemble methods in improving diagnostic tools across various medical fields.

## 1. Introduction

Over two million individuals pass away due to lung cancer yearly around the globe (World Health Organization, 2023). Despite the many advancements in medical technology, many lung cancer cases still go undetected until it is too late for effective treatment. Traditional diagnostic methods, such as computed tomography (CT) scans, often lack accuracy especially with human interpretation, which can lead to delayed or missed diagnoses. Currently, it is clear that Artificial Intelligence (AI) and Machine Learning (ML) are entering many industries, including medical diagnostics, offering new ways to improve detection rates. One promising approach is ensemble modeling, which improves accuracy and robustness by combining AI models for a more reliable output. Because ensemble modeling involves polling several ML models, using this technology to detect lung cancer will result in more reliability and accuracy, leading to more lung cancer cases caught earlier. This paper will first explore the severity and relevance of lung cancer along with its current detection challenges, then examine the potential impacts of AI/ML in medical diagnostics, and lastly, show how ensemble modeling can improve early detection and patient outcomes by comparing typical individual models to models with ensemble modeling.

## 2. Current Challenges in Lung Cancer Detection

### 2.1 Limitations of Traditional Diagnostic Methods

Traditionally used diagnostic tools such as CT scans play an important role in lung cancer detection, but they come with some limitations. Studies show that over 50% of patients undergoing CT scans show noncalcified nodules, which are often benign. The presence of these nodules often calls for additional follow-up procedures, leading to increased costs and more patient anxiety (Jett, 2005, p. 1). Also, these imaging techniques are susceptible to false positives and negatives, making early and accurate detection more difficult.

A major issue with current methods is late detection, which reduces treatment effectiveness. Research shows that patients diagnosed at an early stage have a survival rate of approximately 59%, whereas those diagnosed at later stages struggle with a much lower survival rate of only 6% (A. R. & R. S., 2022, p. 1). This confirms the urgent need for more precise and efficient diagnostic tools that can identify lung cancer at an earlier stage, which would improve patient outcomes.

### 2.2 Need for Improved Diagnostic Accuracy and Early Detection

Given the downfalls and limitations of existing screening methods, there is an urgent need for improved diagnostic accuracy and earlier detection methods. Typical imaging techniques, even with human interpretation, often struggle to differentiate malignant from benign nodules with high confidence, resulting in unnecessary biopsies and delayed action. Improving the accuracy of diagnostic tools is essential to reducing misdiagnoses and ensuring faster treatment. AI- and ML-based methodologies, particularly ensemble modeling, offer promising solutions to address these challenges by increasing the reliability of lung cancer detection by using a combination of AI/ML models.

## 3. Applications of AI and ML in Lung Cancer Detection

### 3.1 Supervised Learning and Deep Learning Approaches

More effective AI-powered methodologies for lung cancer detection involve supervised learning techniques, specifically deep learning approaches such as Convolutional Neural

Networks (CNNs). CNNs are very good at image analysis by automatically extracting important features and identifying patterns that might not be immediately visible to the human eye. A novel VGG19 and Support Vector Machine (SVM) ensemble architecture developed by Elnakib et al. (2020) achieved an impressive 96.25% accuracy, showing the potential of deep learning in improving detection reliability (p. 1).

In addition, feature extraction and automated pattern recognition play an important role in lung cancer detection. For example, Shanbhag et al. (2022) introduced a multi-layer ensemble architecture capable of complex pattern recognition and precise nodule identification, further proving the impact AI can have in medical imaging (pp. 6 - 8). These advancements in deep learning improve detection accuracy, reducing false positives and negatives while improving early diagnostic capabilities.

**3.2 Performance Enhancements Over Traditional Methods**

AI/ML models have demonstrated better performance over traditional diagnostic methods by increasing sensitivity and specificity in detecting malignant nodules. Kriegsmann et al. (2020) developed a highly flexible AI model with several programmable parameters that allowed for fine-tuning the sensitivity of nodule detection, resulting in improved diagnostic precision (pp. 5 - 8). These models learn from diverse datasets, refining their ability to distinguish between benign and malignant nodules more effectively than normal imaging techniques.

Another major advantage of AI-assisted methods is their ability to support radiologists in decision-making. AI decision support systems can analyze vast amounts of imaging data in real-time, flagging suspicious regions and providing confidence scores for detected potential issues. This not only reduces the workload on radiologists but also minimizes human error, leading to more consistent and accurate diagnoses.

### 3.3 Limitations and Challenges

Despite the potential of AI in lung cancer detection, several challenges remain. One major issue is model generalizability and dataset bias. Maleki et al. (2023) found that simple data augmentation techniques improved model performance, increasing F1 scores on their models by approximately 46% (p. 1). This reiterates the importance of diverse and well-representing datasets to ensure AI models can generalize effectively across different demographics.

Ethical concerns also pose serious considerations with AI in medical diagnostics. AI models, specifically black-box algorithms—complex models whose internal decision-making processes are not easily understandable—raise concerns about transparency and accountability in medical decision-making. Biases in training data can lead to misdiagnoses, affecting certain demographic groups more than others. Also, patient consent and data privacy issues remain hot discussion topics, as shown by Farhud & Zokaei (2021), who discussed the potential risks of AI reliance in medical fields (pp. 1–3). Addressing these ethical and technical challenges is important for ensuring the responsible, reliable, safe, and equitable deployment of AI in lung cancer diagnosis.

## 4. Applications of AI and ML in Lung Cancer Detection

### 4.1 Concept and Types of Ensemble Learning

Ensemble learning is a machine learning technique that enhances prediction accuracy and reliability by combining multiple models. Unlike single-model methods, ensemble approaches integrate various classifiers and models to use their individual strengths to create one robust and reliable model, reducing bias and variance in predictions. This method lowers the risk of overfitting to a specific dataset and improves generalization to new data.

There are several types of ensemble learning techniques, including bagging, boosting, and stacking. Bagging (Bootstrap Aggregating) involves training multiple models on different subsets of data and averaging their predictions to reduce variance. On the other hand, Boosting iteratively trains weak models while giving more weight to previously misclassified data points, effectively improving accuracy. Mamun et al. (2022) described and demonstrated the effectiveness of boosting techniques like XGBoost and AdaBoost, which greatly enhanced model performance in lung cancer detection (pp. 4-5). Another method called Stacking takes ensemble learning a step further by combining multiple models and training a meta-learning model to determine the best combination of predictions.

### 4.2 Comparative Advantages Over Single-Model Approaches

Ensemble modeling offers several advantages over single-model approaches, largely in robustness and accuracy. By using multiple models, ensemble modeling techniques create a more resilient/robust predictive system, reducing the chance of errors. A study by Flyckt et al. (2024) showed an ensemble-based AI model that outperformed five professional pulmonologists in identifying lung cancer (p. 1). This displays the potential of ensemble learning to improve diagnostic precision and support medical professionals.

Individual/single AI/ML models tend to have lower predictive performance. Maurya et al. (2024) examined unsupervised learning models that were not part of an ensemble, and the lowest-performing model had an accuracy of 85.71%, which shows a large room for improvement (pp. 7-10). This demonstrates how using ensemble techniques can increase accuracy and reliability for lung cancer detection.

In addition, Ensemble learning's ability to reduce false positives and false negatives is another great aspect. Its model diversity ensures that different learning patterns all contribute to a more balanced prediction. According to Shanbhag et al. (2022), achieving a recall value—this statistic measures the proportion of correctly identified positive cases to the total number of cases—close to 1 is essential for reliable detection (p. 7). Saha et al. (2022) developed ensemble models that consistently achieved recall values near or equal to 1, demonstrating their effectiveness in reducing false negatives and improving early lung cancer detection (p. 15).

**5. Experimental Overview**

This study investigates whether ensemble modeling (EM)—the practice of polling multiple AI/ML models for a more robust and accurate output—improves the performance of convolutional neural networks (CNNs) in classifying lung cancer types. An experimental approach was used, comparing multiple CNN architectures—both as standalone models and with EM—to evaluate their effectiveness. Each model was assessed using standard classification metrics, with testing conducted across two platforms to ensure reliability. By controlling all other variables, the experiment isolates the impact of ensemble modeling techniques, allowing for a clear analysis of their benefits.

**6. Methodology**

The experiment involved coding and training four CNN architectures (a basic CNN with no mainstream architecture, AlexNet, VGG16, and VGG19) in both their original and ensemble-enhanced forms. Python (with PyTorch) served as the primary tool for model development, while performance metrics (accuracy, F1 score, precision, and recall) were recorded automatically during testing. To minimize bias and data error, each model was trained and evaluated twice—first on the researcher's computer and then on Google Colab—to account for computational variations. The dataset used was preprocessed to ensure consistency in input structure, resolution, and class distribution.

The ensemble models employed majority voting or averaging (depending on architecture compatibility) to combine predictions from the main CNN model and a few smaller models. Performance was assessed on a separate test set. This approach not only tested the hypothesis but also identified which architectures benefited most from ensemble techniques.
Key factors—such as training epochs, batch size, optimizer settings, and data splits—were kept identical across all models to ensure fair comparison. The use of two execution environments (local vs. cloud) further validated the robustness of results.

**7. Results**

The data collected from this experiment attempts to directly address the question of if EM enhances the performance of multi-class lung cancer classification and detection. The results demonstrate that, in all cases, ensemble modeling enhanced the performance of the base CNN models across key evaluation metrics—accuracy, F1 score, precision, and recall. This supports the original hypothesis that combining CNNs with complementary machine learning models (Random Forest and SVM) through majority voting leads to more robust and accurate classification. The improvements were particularly significant for weaker standalone models like VGG16, suggesting that ensemble techniques can compensate for individual model shortcomings by leveraging the strengths of other ML models.

| *EM stands for Ensemble Modeling | Averages | | | |
|---|---|---|---|---|
| | Accuracy | F1 Score | Precision | Recall |
| **Basic CNN** | 0.5810 | 0.5719 | 0.6298 | 0.5810 |
| Basic CNN + EM* | 0.6000 | 0.6006 | 0.6557 | 0.6000 |
| **AlexNet CNN** | 0.5349 | 0.4847 | 0.6138 | 0.5349 |
| AlexNet CNN + EM* | 0.5571 | 0.5257 | 0.6736 | 0.5571 |
| **VGG16 CNN** | 0.4111 | 0.3915 | 0.4615 | 0.4091 |
| VGG16 CNN + EM* | 0.6397 | 0.6204 | 0.7216 | 0.6385 |
| **VGG19 CNN** | 0.6214 | 0.5969 | 0.6784 | 0.6195 |
| VGG19 CNN + EM* | 0.6531 | 0.6339 | 0.7025 | 0.6513 |

**Table 1. Averaged results**

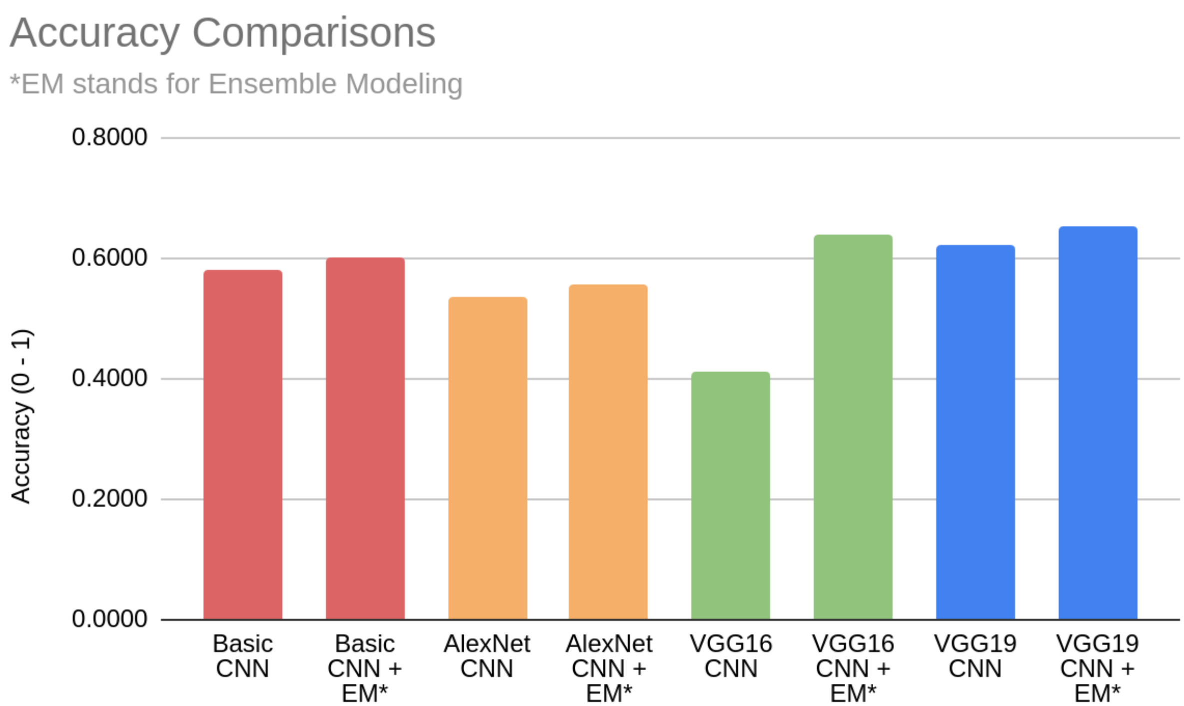


**Figure 1. Averaged model accuracies in bar chart form.**

As is easily seen in figure 1, the most striking result was the 55.6% relative improvement in accuracy—characterized by the noticeable difference in height of the light-green bars—for VGG16 when ensemble modeling was applied (from 0.4111 to 0.6397). This suggests that VGG16, which performed poorly as a standalone model, benefited greatly from the ensemble's ability to correct its errors through consensus with Random Forest and SVM. Similarly, VGG19—already the best-performing single model (0.6214 accuracy)—still saw a 5.1% boost with EM (to 0.6531), reinforcing that even strong CNNs can gain from ensemble methods. The F1 score, which balances precision and recall, also improved consistently, indicating that EM not only increased correct predictions but also reduced false positives and negatives.

However, the gains were less pronounced for simpler architectures like the basic no-architecture CNN and AlexNet, which saw 3 - 9% improvements in accuracy; this can be seen with the slight height differences between the respective CNN and its EM version in figure 1. This implies that ensemble modeling's effectiveness depends on the base model's initial performance: weaker models benefit more, while stronger ones see marginal but still meaningful gains. The precision-recall trade-offs also improved, with VGG16 + EM showing the largest leap (0.4615 to 0.7216 precision), suggesting EM particularly helps in reducing misclassifications of lung cancer. These findings confirm that ensemble modeling is a viable strategy for enhancing lung cancer diagnosis via AI/ML.

**8. Discussion**

One key limitation of this research was the reliance on a controlled environment, which may not fully replicate real-world clinical conditions. The models were trained and tested on a specific dataset, and variations in image quality, tumor heterogeneity, or class imbalances in a broader medical setting could affect performance. Additionally, the ensemble approach used only two supplementary models (Random Forest and SVM) with majority voting, which may not capture the full potential of more advanced ensemble techniques like stacking or boosting. Computational constraints also limited the depth of hyperparameter tuning and the ability to support more complementary models, potentially affecting the generalizability and accuracy of the results. Future research could address these limitations by incorporating larger, more diverse datasets and testing alternative ensemble strategies.

This research demonstrates that ensemble modeling can enhance the accuracy and reliability of lung cancer classification using AI/ML, which is critical for improving early diagnosis and treatment. The findings suggest that even poorly performing CNNs can be significantly improved through ensemble methods, making AI systems more feasible for clinical use. Future studies could explore hybrid architectures that combine CNNs with transformers or investigate domain-specific ensemble techniques tailored for medical imaging. Additionally, integrating explainable AI tools (e.g., Grad-CAM) could help clinicians trust and interpret AI predictions. By refining these approaches, AI could eventually support radiologists in delivering faster, more accurate lung cancer diagnoses.

## 9. Conclusion

Lung cancer remains one of the deadliest cancers worldwide, mostly due to late detection and the limitations of existing/traditional diagnostic methods. As research and this study have shown, AI and ML have introduced promising advancements in medical imaging, increasing the accuracy and efficiency of lung cancer detection. However, single AI models still face challenges, such as high false positive rates, bias, and lack of generalizability. Ensemble modeling addresses these issues by integrating multiple AI models and using their different learning patterns, thereby improving robustness and reliability in lung cancer diagnostics. By using different learning strategies such as bagging, boosting, and stacking, ensemble methods enhance detection accuracy, minimize false positives, and help with early intervention in lung cancer diagnosis and treatment. This ensures that more lung cancer cases are identified at treatable stages, ultimately improving patient outcomes.

**References**

A. R., B., & Kumar R. S., V. (2022). Deep learning-based lung cancer classification of CT images using augmented convolutional neural networks. *ELCVIA Electronic Letters on Computer Vision and Image Analysis, 21*(1). doi.org/10.5565/rev/elcvia.1490

Ashwin Shanbhag, G., Anurag Prabhu, K., Subba Reddy, N. V., & Ashwath Rao, B. (2022). Prediction of lung cancer using ensemble classifiers. *Journal of Physics: Conference Series, 2161*(1). doi.org/10.1088/1742-6596/2161/1/012007

Elnakib, A., M. Amer, H., & E.Z. Abou-Chadi, F. (2020). Early Lung Cancer Detection using Deep Learning Optimization. *International Journal of Online and Biomedical Engineering (iJOE), 16*(06), pp. 82–94. doi.org/10.3991/ijoe.v16i06.13657

Farhud, D. D., & Zokaei, S. (2021). Ethical issues of artificial intelligence in medicine and healthcare. *Iranian Journal of Public Health*. doi.org/10.18502/ijph.v50i11.7600

Flyckt, R. N. H., Sjodsholm, L., Henriksen, M. H. B., Brasen, C. L., Ebrahimi, A., Hilberg, O., Hansen, T. F., Wiil, U. K., Jensen, L. H., & Peimankar, A. (2024, February 14). *Pulmonologists-Level lung cancer detection based on standard blood test results and smoking status using an explainable machine learning approach*. arXiv. Retrieved January 1, 2025, from arxiv.org/pdf/2402.09596

Jett, J. R. (2005). Limitations of screening for lung cancer with low-dose spiral computed tomography. *Clinical Cancer Research, 11*(13), 4988s-4992s. doi.org/10.1158/1078-0432.ccr-05-9000

Kriegsmann, M., Haag, C., Weis, C.-A., Steinbuss, G., Warth, A., Zgorzelski, C., Muley, T., Winter, H., Eichhorn, M. E., Eichhorn, F., Kriegsmann, J., Christopoulos, P., Thomas, M., Witzens-Harig, M., Sinn, P., Stenzinger, A., & Kriegsmann, K. (2020). Deep Learning for the Classification of Small-Cell and Non-Small-Cell Lung Cancer. *Cancers, 12*(6). doi.org/10.3390/cancers12061604

Maleki, F., Ovens, K., Gupta, R., Reinhold, C., Spatz, A., & Forghani, R. (2023). Generalizability of machine learning models: Quantitative evaluation of three methodological pitfalls. Radiology: *Artificial Intelligence*, 5(1). doi.org/10.1148/ryai.220028

Mamun, M., Farjana, A., Al Mamun, M., & Ahammed, M. S. (2022). Lung cancer prediction model using ensemble learning techniques and a systematic review analysis. *IEEE Xplore, World AI IoT Congress (AIIoT)*. doi.org/10.1109/AIIoT54504.2022.9817326

Maurya, S. P., Sisodia, P. S., Mishra, R., & Singh, D. P. (2024). Performance of machine learning algorithms for lung cancer prediction: A comparative approach. *Scientific Reports, 14*(1). doi.org/10.1038/s41598-024-58345-8

World Health Organization. (2023, June 26). *Lung cancer* [Fact sheet]. World Health Organization. Retrieved October 7, 2024, from www.who.int/news-room/fact-sheets/detail/lung-cancer