

## Article

# XplainLungSHAP: Enhancing Lung Cancer Surgery Decision Making with Feature Selection and Explainable AI Insights

Flavia Costi <sup>\*,†</sup>, Emanuel Covaci <sup>\*,†</sup>  and Darian Onchis

Department of Computer Science, West University of Timisoara, 300223 Timisoara, Romania; darian.onchis@e-uvt.ro

\* Correspondence: flavia.costi@e-uvt.ro (F.C.); emmanuel.covaci98@e-uvt.ro (E.C.)

† These authors contributed equally to this work.

**Abstract: Background:** Lung cancer surgery often involves complex decision-making, where accurate and interpretable predictive models are crucial for assessing postoperative risks and optimizing outcomes. This study presents XplainLungSHAP, a novel framework combining SHAP (SHapley Additive exPlanations) and attention mechanisms to enhance both predictive accuracy and transparency. The aim is to support clinicians in preoperative evaluations by identifying and prioritizing key clinical features. **Methods:** The framework was developed using data from 470 patients undergoing lung cancer surgery. Key clinical features were identified through SHAP, ensuring alignment with medical expertise. These features were dynamically weighted using an attention mechanism in a neural network, enhancing their impact on survival predictions. The model's performance was evaluated through accuracy, confusion matrices, and ROC analysis, demonstrating its reliability and interpretability. **Results:** The XplainLungSHAP model achieved an accuracy of 91.49%, outperforming traditional machine learning models. SHAP analysis identified critical predictors, including pulmonary function, comorbidities, and age, while the attention mechanism prioritized these features dynamically. The combined approach ensured high accuracy and offered actionable insights into survival predictions. **Conclusions:** XplainLungSHAP addresses the limitations of black-box models by integrating explainability with state-of-the-art predictive techniques. This framework provides a transparent and clinically relevant tool for guiding surgical decisions, supporting personalized care, and advancing AI applications in thoracic oncology.



Academic Editor: Andrew A. Gumbs

Received: 27 November 2024

Revised: 27 December 2024

Accepted: 24 January 2025

Published: 31 January 2025

**Citation:** Costi, F.; Covaci, E.; Onchis, D. XplainLungSHAP: Enhancing Lung Cancer Surgery Decision Making with Feature Selection and Explainable AI Insights. *Surgeries* **2025**, *6*, 8. <https://doi.org/10.3390/surgeries6010008>

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** attention mechanism; explainable AI; machine learning in healthcare; postoperative survival; predictive modeling; precision medicine; SHAP; thoracic surgery

## 1. Introduction

Thoracic surgery plays a vital role in treating lung cancer, often offering the possibility of a cure through major lung resections [1]. However, predicting survival after surgery remains difficult, requiring precise models to guide decision making, assess risks, and use resources efficiently. These models must also be interpretable so clinicians can understand and trust their predictions. This study used data from the Wroclaw Thoracic Surgery Centre in Poland, involving 470 patients who underwent major lung resections for primary lung cancer between 2007 and 2011. The dataset focuses on a binary classification task: predicting whether a patient survives more than one year after surgery (class 2) or dies within a year (class 1). It includes information on preoperative health, demographics, procedural details, and survival outcomes, providing a valuable foundation for creating interpretable predictive models [2].

Post-surgery survival depends on comorbidities, lung function, and surgical complexity. While traditional machine learning (ML) models can be accurate, they often operate as “black boxes”, offering little explanation for their predictions. This lack of transparency limits their use in clinical settings, where interpretability is relevant. Explainable AI tools like SHAP overcome this by showing how each feature influences a model’s predictions. SHAP helps prioritize significant factors, such as tumor size and pulmonary function, aligning the model with clinical reasoning while also improving performance on challenging datasets [3]. Attention mechanisms, commonly used in fields like natural language processing and computer vision, can also enhance interpretability by focusing on key input features dynamically. In healthcare, they can highlight important clinical factors such as comorbidities and pulmonary function. However, traditional attention mechanisms also face challenges in being fully interpretable, which is important in high-stakes areas like medicine [4]. This study combines SHAP and attention mechanisms into a novel framework to predict survival outcomes after thoracic surgery. SHAP first identifies and ranks key features, while an attention mechanism implemented in PyTorch dynamically emphasizes these features during predictions. This hybrid approach improves both accuracy and efficiency. Post hoc XAI methods further validate the XplainLungSHAP model, ensuring insights are actionable for clinicians.

The approach addresses challenges in thoracic surgery by narrowing the focus to clinically significant features and dynamically adjusting their weight during prediction. This results in a model that is both high-performing and transparent. Clinicians gain a reliable tool for decision making in thoracic oncology, paving the way for better outcomes and wider use of ML in precision medicine. By integrating SHAP to identify important features and using a feature-based attention mechanism, the XplainLungSHAP model achieved improved accuracy and efficiency. SHAP’s feature ranking reduced noise, allowing the model to concentrate on meaningful data, which enhanced its ability to generalize. This led to better performance in predicting whether patients survived beyond one year [5]. The attention mechanism further optimized performance by prioritizing these features during prediction, reducing computational requirements and training time. This combination of accuracy and efficiency demonstrates the potential of this approach to deliver reliable, interpretable solutions for clinical decision making in thoracic surgery [6].

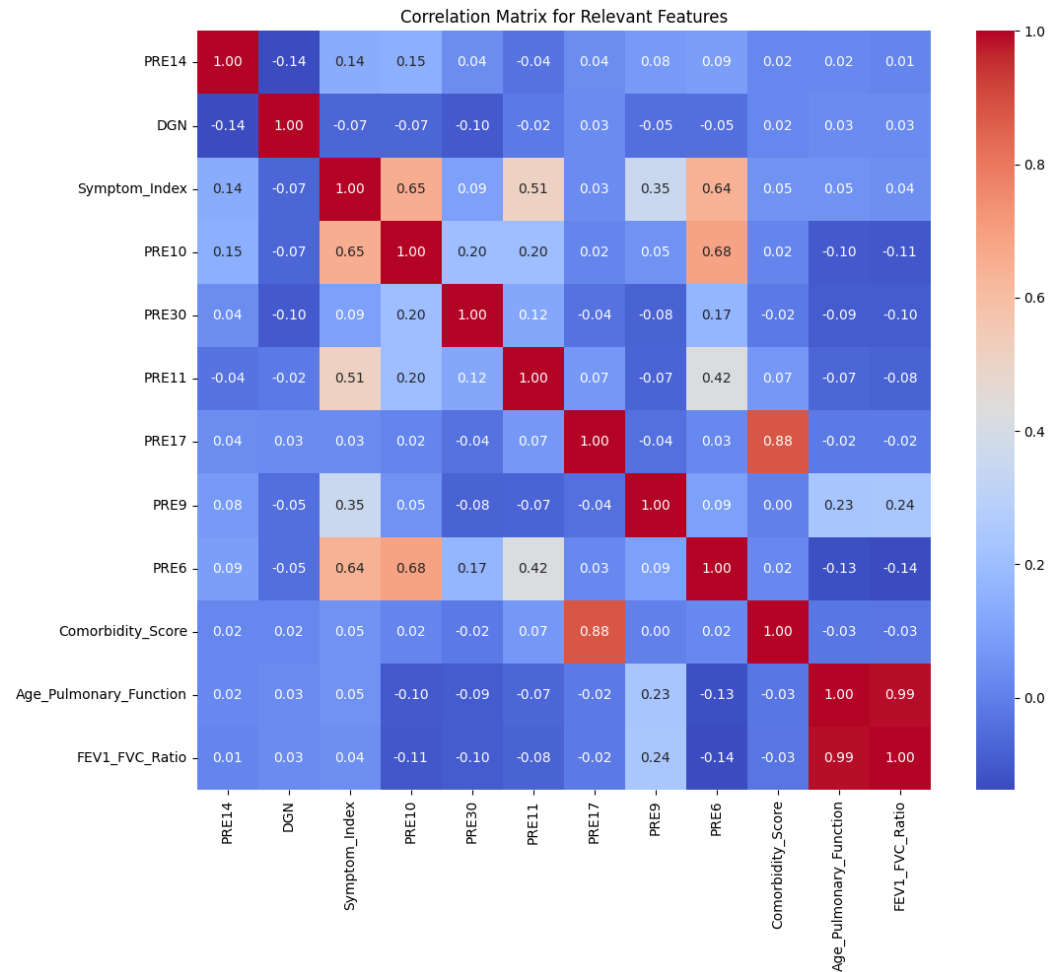
The novelty of this article lies in integrating SHAP for feature selection with a feature-based attention mechanism in a unified framework. This enables high predictive accuracy (91.49%) and clinical interpretability for postoperative survival in thoracic surgery, addressing the limitations of existing black-box models.

This paper is structured as follows: Section 2 provides an overview of the materials and methods, including feature engineering, SHAP analysis, and the implementation of the attention mechanism; Section 3 presents the experimental results and performance metrics of the model; Section 4 discusses the implications and significance of our findings; and Section 5 concludes with key insights and outlines future directions.

## 2. Materials and Methods

For this study, we utilized a public dataset available on Kaggle [2], which contains 17 characteristics in total. Of these, 16 are independent variables that represent a combination of clinical, demographic, and procedural factors. The dependent variable indicates whether the patient survived beyond one year after undergoing major lung resection surgery for primary lung cancer. This binary classification problem forms the core of the analysis, distinguishing between survival within one year (class 1) and survival beyond one year (class 2). In a correlation matrix, values range from -1 to 1, where 1 indicates a perfect positive linear relationship, meaning both variables increase together in exact proportion.

A value of 0 means there is no linear relationship between the variables, though they may still have a non-linear connection. A value of  $-1$  represents a perfect negative linear relationship, where one variable increases as the other decreases in perfect inverse proportion. Values between 0 and 1 indicate varying degrees of positive correlation, where one variable tends to increase as the other increases, while values between 0 and  $-1$  represent varying degrees of negative correlation, where one variable tends to increase while the other decreases. The closer the value is to 1 or  $-1$ , the stronger the linear relationship, and values near 0 suggest a weaker or no linear relationship. In Figure 1, the correlation matrix is presented exclusively for the relevant features extracted from the dataset.



**Figure 1.** Correlation matrix of relevant features from the dataset.

We summarize the most relevant characteristics in the dataset:

- **FEV1 (Forced Expiratory Volume in One Second):** A measure of lung function, FEV1 evaluates the volume of air exhaled in one second. Lower values indicate impaired pulmonary function, which increases the risk of complications and affects recovery post-surgery.
- **Before Surgery:** This binary variable reflects whether the patient experienced significant preoperative pain, often associated with advanced disease or tumor involvement in sensitive areas, which can negatively impact recovery.
- **Age:** Older patients face higher surgical risks due to reduced physiological reserves and comorbidities, while younger patients generally experience better outcomes.

- **Dyspnea (Shortness of Breath):** This symptom signals compromised respiratory function, increasing the likelihood of perioperative complications such as hypoxia or ventilation issues.
- **Cough:** Chronic cough, a symptom of lung conditions like cancer or bronchitis, can exacerbate discomfort and indicate advanced airway involvement.
- **Hemoptysis:** Coughing up blood is an important symptom often linked to severe underlying pathologies like advanced lung cancer or airway damage, increasing surgical risks.
- **Tumor Size and Extent:** Larger tumors or those involving relevant structures require more extensive resections, elevating surgical complexity and affecting recovery.
- **Surgical Approach:** The type of procedure (e.g., lobectomy or pneumonectomy) significantly impacts postoperative outcomes. Minimally invasive methods like VATS reduce complications and recovery time.
- **Air Leak and Pneumothorax:** Postoperative complications such as prolonged air leaks or collapsed lungs delay recovery and may require additional interventions.
- **Performance Status Scores:** These scores evaluate the patient's overall ability to perform daily activities, predicting their tolerance to surgical stress and recovery prospects.
- **Comorbidity Indices:** Chronic conditions like diabetes or hypertension are important factors that influence recovery and overall surgical risk.
- **Dependent Variable: Survival Outcome:** The binary outcome indicates whether the patient survived beyond one year (class 2) or not (class 1). This serves as the central variable for modeling postoperative survival probabilities [2].

In this study, we employed a combination of feature engineering, scaling, and machine learning techniques enhanced by XAI using SHAP to interpret the predictions of our XplainLungSHAP model. Below, we describe each step in detail, starting from feature engineering, data preparation, and model development and ending with applying SHAP for explainability.

### 2.1. Feature Engineering

Feature engineering is relevant in preparing data for machine learning, especially in clinical applications where expert knowledge can greatly enhance a dataset's predictive capabilities. In this study, we worked with the thoracic surgery dataset and developed new features by exploring relationships and interactions between existing variables. These features were crafted to emphasize meaningful clinical patterns and improve the model's ability to make accurate predictions. Below is an explanation of the logic, mathematical approach, and clinical importance of the derived features [7].

One of the key engineered features is the FEV1/FVC ratio, calculated by dividing Forced Expiratory Volume in One Second (FEV1) by Forced Vital Capacity (FVC). This ratio is a measure in pulmonary medicine and is commonly used to diagnose obstructive lung conditions, such as chronic obstructive pulmonary disease (COPD). The formula for this ratio is as follows:

$$f(\text{FEV1}, \text{FVC}) = \frac{\text{FEV1}}{\text{FVC}}, \quad (1)$$

where the following are true:

- FEV1 (PRE5) represents the volume of air a patient can forcefully exhale in one second.
- FVC (PRE4) represents the total volume of air exhaled during a forced breath.

In this study, the FEV1/FVC ratio provides an aggregated measure of lung function, which is important in assessing whether a patient's respiratory capacity can tolerate surgical

interventions. Patients with a lower ratio are likely to have compromised pulmonary function, increasing the risk of postoperative complications.

Next, we derived the Symptom Index, which quantifies the burden of preoperative symptoms reported by each patient. Symptoms such as dyspnea, cough, and hemoptysis are not only indicative of disease severity but also predictive of surgical outcomes. To capture this symptom burden, we summed five binary variables:

$$\text{Symptom\_Index} = \text{PRE7} + \text{PRE8} + \text{PRE9} + \text{PRE10} + \text{PRE11}, \quad (2)$$

where each term represents the presence (1) or absence (0) of specific symptoms. This composite score reflects the overall physiological distress of the patient, helping the model to incorporate non-invasive indicators of disease progression.

The comorbidity score was introduced to capture the cumulative effect of chronic conditions on surgical outcomes. Pre-existing conditions such as diabetes, cardiovascular disease, and hypertension can significantly influence postoperative recovery and mortality risk. This score is defined as:

$$\text{Comorbidity\_Score} = \text{PRE17} + \text{PRE25} + \text{PRE32}, \quad (3)$$

where each binary variable represents the presence (1) or absence (0) of a specific comorbidity. This aggregation ensures that the model accounts for the holistic impact of a patient's overall health status on their survival probability.

Recognizing the interplay between age and pulmonary function, we created the Age–Pulmonary Function Interaction feature. This interaction term is designed to capture how age-related declines in respiratory capacity exacerbate surgical risks. It is mathematically expressed as:

$$\text{Age\_Pulmonary\_Function} = \text{AGE} \times \text{FEV1\_FVC\_Ratio}. \quad (4)$$

where AGE represents the chronological age of the patient. Multiplying it with the FEV1/FVC ratio highlights patients who, due to advanced age and poor lung function, are at heightened risk of complications or mortality.

Once these engineered features were generated, the dataset underwent preprocessing to ensure uniformity across variables. Specifically, numerical features, including AGE, PRE4, PRE5, FEV1\_FVC\_Ratio, and Age\_Pulmonary\_Function, were standardized using z-score normalization:

$$X' = \frac{X - \mu}{\sigma}, \quad (5)$$

where  $X$  is the original value of the feature,  $\mu$  is the mean, and  $\sigma$  is the standard deviation. This scaling ensures that all features are on the same scale, facilitating faster model convergence and improving numerical stability during training [8].

## 2.2. SHAP for Feature Relevance Extraction

SHAP is a state-of-the-art XAI technique grounded in cooperative game theory, designed to attribute the contribution of each feature to the predictions made by a machine learning model. In the context of this study, SHAP was employed to identify the most relevant features influencing the postoperative survival outcomes of thoracic surgery patients. By quantifying the importance of each feature, SHAP provides a transparent and interpretable framework for understanding XplainLungSHAP model behavior, ensuring that the predictions align with clinical reasoning and domain knowledge [9].

The fundamental concept of SHAP is the Shapley value, which originates from game theory and represents a fair distribution of a collective payoff among contributors. In ma-

chine learning, the collective payoff corresponds to the model's prediction, and the contributors are the features involved in generating that prediction. Mathematically, the SHAP value for a feature  $j$  is calculated as:

$$\phi_j = \sum_{S \subseteq N \setminus \{j\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} [f(S \cup \{j\}) - f(S)], \quad (6)$$

where the following are true:

- $N$  is the set of all features.
- $S$  is a subset of features excluding  $j$ .
- $f(S)$  is the model prediction when only features in  $S$  are considered.

This formula ensures that each feature's contribution is assessed across all possible subsets of features, capturing both its individual and interactive effects on the prediction. The weighting term, derived from the factorial of subset sizes, guarantees fairness by accounting for all permutations of feature inclusion [10].

In this study, SHAP was applied to the thoracic surgery dataset to identify the most significant features influencing patient survival. Using a trained neural network as the predictive model, SHAP values were computed for all features to rank their contributions to the model's predictions. This analysis highlighted the importance of core variables like age and pulmonary function measures, as well as engineered features such as the FEV1/FVC ratio and comorbidity score. Features with high SHAP values consistently shifted predictions toward either mortality within one year (class 1) or survival beyond one year (class 2), showcasing their role in the decision making process.

One of SHAP's strengths is its ability to provide both local and global explanations. Locally, SHAP values illustrate how each feature affects a patient's survival prediction, helping clinicians understand the reasoning behind specific outcomes [11]. Globally, aggregated SHAP values uncover overall trends and rank features by importance across the dataset, ensuring the model's predictions remain interpretable and clinically meaningful. By integrating SHAP into the feature selection process, we balanced clinical relevance with data-driven insights [12]. This approach allowed us to prioritize the most influential predictors, such as age, pulmonary function, and comorbidities, while reducing the impact of less relevant variables. As a result, SHAP not only enhanced the interpretability of the model but also improved its predictive accuracy by focusing on the features most strongly associated with thoracic surgery outcomes [13].

The pseudocode for computing SHAP values is outlined in Algorithm 1. It iteratively calculates the marginal contribution of each feature for all possible subsets of the dataset, summing these contributions with appropriate weights [11].

In this study, SHAP was applied to interpret predictions made by a neural network trained on the thoracic surgery dataset. A subset of the data ( $B$ ) was used as background data to compute the baseline predictions, providing a reference point for feature contributions. The SHAP values were calculated for both original and engineered features, including:

- The FEV1/FVC ratio, capturing pulmonary function.
- The Symptom Index, reflecting the cumulative symptom burden.
- The comorbidity score, indicating the impact of chronic conditions.
- The Age-Pulmonary Function Interaction, highlighting the interplay between age and respiratory capacity [7].

After identifying the most relevant features using SHAP, we integrated these features into an attention-based deep learning framework to predict postoperative survival outcomes. Attention mechanisms are widely used in modern machine learning architectures,

particularly for tasks requiring dynamic weighting of input features. In this study, the attention mechanism enabled the XplainLungSHAP model to focus on the most important features, as determined by SHAP, while dynamically adjusting their contributions based on the input data.

---

**Algorithm 1** SHAP for Feature Relevance Extraction
 

---

```

1: Input: Trained model  $f$ , dataset  $X$  with features  $N$ , background data  $B \subseteq X$  for SHAP baseline
2: Output: SHAP values  $\phi_j$  for each feature  $j \in N$ 
3: procedure COMPUTE SHAP VALUES
4:   Initialize SHAP value vector  $\phi = [0, \dots, 0]$  of size  $|N|$ 
5:   for all features  $j \in N$  do
6:      $\phi_j \leftarrow 0$  //Initialize SHAP value for feature  $j$ 
7:     for all subsets  $S \subseteq N \setminus \{j\}$  do
8:       Compute model predictions:
            $f(S \cup \{j\})$  and  $f(S)$ 
9:       Compute marginal contribution:
            $\Delta = f(S \cup \{j\}) - f(S)$ 
10:      Update SHAP value:
            $\phi_j \leftarrow \phi_j + \frac{|S|!(|N| - |S| - 1)!}{|N|!} \cdot \Delta$ 
11:     end for
12:   end for
13: end procedure
14: Return: SHAP values  $\phi = [\phi_1, \phi_2, \dots, \phi_{|N|}]$ 

```

---

The attention mechanism works by assigning a weight  $\alpha_i$  to each feature  $x_i$ , which represents its relative importance for the given prediction. These weights are computed as follows:

$$\alpha_i = \frac{\exp(e_i)}{\sum_{j=1}^N \exp(e_j)}, \quad (7)$$

where the following are true:

- $e_i$  is the relevance score of feature  $x_i$ , calculated using a trainable scoring function  $e_i = W^T x_i + b$ , where  $W$  and  $b$  are learnable parameters.
- $\sum_{j=1}^N \exp(e_j)$  ensures that the weights  $\alpha_i$  sum to 1, making them interpretable as probabilities.

The weighted sum of the feature embeddings is then calculated as:

$$z = \sum_{i=1}^N \alpha_i x_i, \quad (8)$$

where  $z$  is the attention-weighted representation of the input features. This representation is passed through the subsequent layers of the neural network to make the final prediction.

Incorporating the attention mechanism significantly improves the model's interpretability and adaptability by emphasizing the features that have the greatest influence on individual predictions [4]. For example, in predicting survival outcomes, factors like age, comorbidity score, or FEV1/FVC ratio may be assigned higher attention weights for patients with compromised pulmonary function, whereas different features might domi-

nate in the case of younger, healthier individuals. This mechanism was implemented in PyTorch using a structured approach:

1. The input features, including those identified as important by SHAP, were transformed into high-dimensional embeddings through a linear layer.
2. Attention scores  $e_i$  were calculated for each feature and normalized using the softmax function to generate weights  $\alpha_i$ .
3. A weighted sum  $z$  of the features was then computed, and this output was fed into subsequent fully connected layers to perform the binary classification (survival or mortality).

The attention mechanism works in tandem with SHAP, leveraging SHAP's insights into feature relevance to inform the model's focus during predictions. This creates a unified, interpretable framework for clinical decision making. By prioritizing the most relevant features as identified by SHAP, the model dynamically adjusts to the specific attributes of each patient. This approach not only achieves strong predictive accuracy but also ensures transparency, making it particularly suitable for high-stakes medical applications [14].

### 3. Results

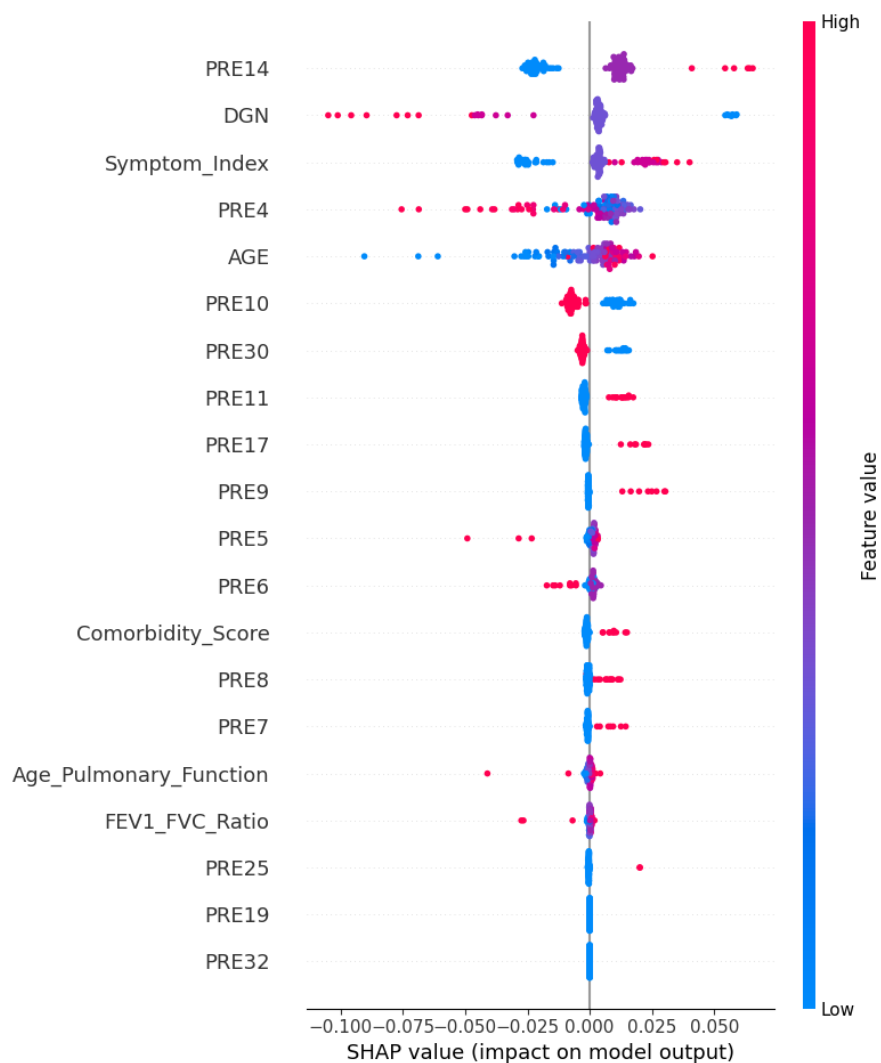
After applying the SHAP explainability method to the dataset, we identified the most relevant features to predict postoperative survival outcomes. These features, presented in the figure below, highlight the variables with the highest impact on the XplainLungSHAP model's decision making process. The identified features include the following: PRE14, DGN, Symptom\_Index, PRE10, PRE30, PRE11, PRE17, PRE9, PRE6, Comorbidity\_Score, Age\_Pulmonary\_Function, and FEV1\_FVC\_Ratio (see Figure 2).

Following the application of the attention mechanism to the relevant features identified by SHAP, the model achieved a notable accuracy of 0.9149 (see Figure 3). This demonstrates the effectiveness of dynamically prioritizing the most significant features during prediction, enhancing the model's performance in distinguishing between postoperative survival outcomes.

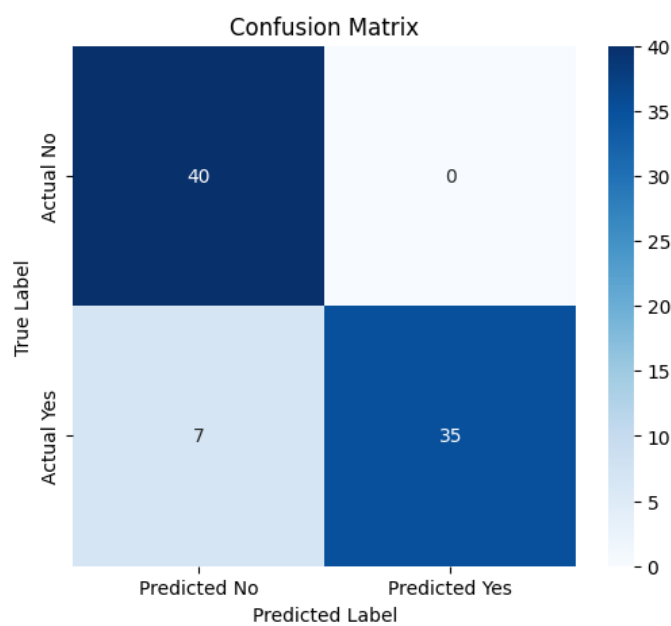
The results are further validated by the confusion matrix, which provides a detailed breakdown of true positives, true negatives, false positives, and false negatives. This matrix highlights the model's ability to accurately classify survival probabilities, ensuring its robustness and reliability for clinical applications. At the same time, Figure 4 shows the ROC curve, which also attests to the performance of the developed model.

Logistic regression, Random Forest, and XGBoost are widely used and effective machine learning models for classification problems. However, they failed to outperform the performance of XplainLungSHAP for several reasons. Logistic regression, being a linear model, has limitations in capturing complex non-linear relationships between features and outcomes. In our dataset, the non-linear relationships between factors such as age, pulmonary function, and comorbidities are important for prediction [15]. Although Random Forest is strong in capturing non-linear relationships, it is not optimal for small datasets with complex dependencies. The model tends to overfit noisy data and faces challenges in clinical interpretation [16]. XGBoost achieved the highest accuracy among traditional models due to its advanced techniques for loss reduction and feature weighting [17]. However, it lacks intrinsic interpretability mechanisms comparable to SHAP and feature-based attention.

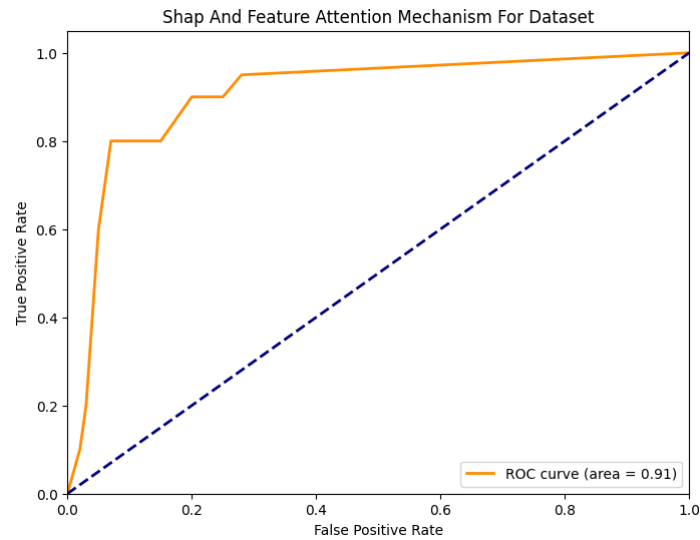




**Figure 2.** Relevant features identified by SHAP. The figure highlights the most impactful features contributing to the prediction of postoperative survival outcomes.



**Figure 3.** Confusion matrix obtained from XplainLungSHAP.

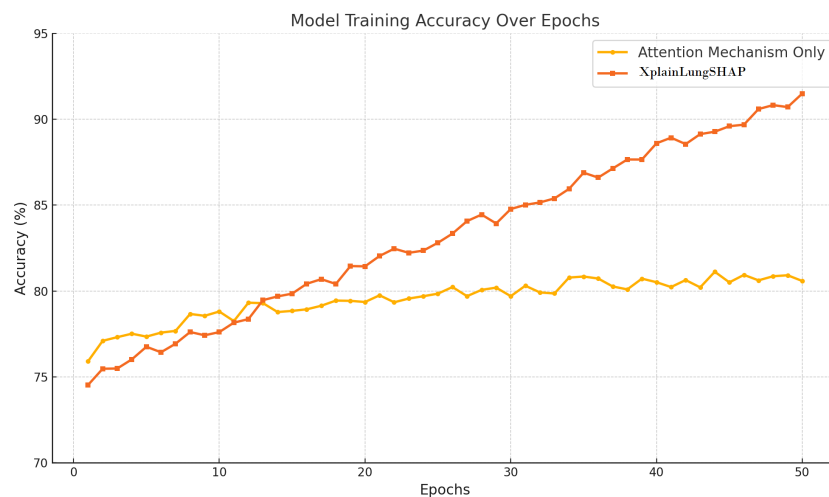


**Figure 4.** ROC curve obtained from XplainLungSHAP.

In contrast, XplainLungSHAP combines the explanation of important features (using SHAP) with the attention mechanism to prioritize the dynamics of relevant factors, delivering both high precision and interpretability—relevant for medical applications. This approach enables clinicians to understand why and how a particular feature influences prediction. As shown in the table above (see Table 1), the results obtained before implementing XplainLungSHAP were 11% lower. In Figure 5, another result is presented, obtained after training the model using an attention mechanism and when using the custom model. A consistent increase in the accuracy of the developed model can be observed, reaching over 90% at some point.

**Table 1.** Models’ accuracy comparison.

Model	Accuracy (%)
Logistic regression	78.65
Random Forest	79.12
XGBoost	80.49
Attention mechanism only	80.97
XplainLungSHAP model	91.49



**Figure 5.** Training accuracy progression for attention mechanism and XplainLungSHAP models.

#### 4. Discussion

The incorporation of an attention mechanism into the neural network greatly improved both the accuracy and interpretability of the model. By dynamically assigning importance to features based on their relevance to each prediction, the mechanism effectively highlighted key clinical variables such as pulmonary function, comorbidity scores, and age. This dynamic weighting process aligned the model's functionality with clinical reasoning, supported by a carefully designed architecture and thoughtful selection of model components.

The neural network's structure begins with an input layer that processes 12 essential features identified through SHAP, including variables like the FEV1/FVC ratio, Symptom Index, and comorbidity score. This ensures the input data are compact yet clinically relevant. Following the input layer, two fully connected hidden layers refine the data. The first hidden layer, comprising 64 neurons, applies a ReLU activation function, enabling the model to capture complex, non-linear relationships between the features. This step is particularly important for understanding the intricate patterns present in medical datasets. The second hidden layer, with 32 neurons and also employing a ReLU activation function, narrows the model's focus further, emphasizing the most relevant patterns uncovered in the previous layer. This progression enables the model to extract and prioritize information effectively. The design not only enhances the model's predictive performance but also ensures interpretability. The attention mechanism allows for the dynamic prioritization of significant features, resulting in a reliable and transparent model tailored to clinical applications. This structure offers a robust framework for decision making, aligning with medical requirements while delivering valuable insights.

The attention mechanism refines the predictive process by assigning different levels of importance to each feature based on its relevance to the specific case. It evaluates how much each feature contributes to the overall prediction and adjusts their influence accordingly. This process creates a weighted combination of the features, emphasizing the most relevant ones. The resulting representation is then used by the model to make predictions, offering a clear and interpretable way to understand which factors are most significant in determining postoperative survival outcomes [4].

To train the model effectively, we employed binary cross-entropy loss, a standard choice for binary classification tasks [18]. This loss function calculates the difference between the predicted probabilities and the actual binary outcomes, penalizing the model proportionally to its confidence in incorrect predictions. The binary cross-entropy loss is defined as:

$$\mathcal{L} = -\frac{1}{m} \sum_{i=1}^m [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)], \quad (9)$$

To optimize the model, we used the Adam optimizer, which combines the advantages of momentum and adaptive learning rates. Adam adjusts the learning rate for each parameter dynamically, improving convergence speed and stability, particularly in sparse or noisy datasets. The optimizer updates the parameters using the following equations:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t, \quad v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2, \quad (10)$$

$$\theta_t = \theta_{t-1} - \frac{\alpha}{\sqrt{v_t} + \epsilon} m_t, \quad (11)$$

where  $m_t$  and  $v_t$  are the first and second moment estimates,  $g_t$  is the gradient, and  $\theta_t$  represents the updated parameters. This adaptive nature of Adam makes it particularly suitable for the complex optimization landscape of neural networks with attention mechanisms [19].

The attention mechanism not only improved the model's predictive accuracy but also enhanced its interpretability. By providing insight into the relative importance of each feature for individual predictions, the model allowed clinicians to validate its reasoning and integrate its outputs into clinical decision making with confidence. This combination of accuracy, interpretability, and computational efficiency highlights the potential of attention-based approaches in medical prediction tasks, offering a robust and transparent framework for improving outcomes in thoracic surgery [5].

## 5. Conclusions

This study introduces an innovative framework leveraging AI for predicting postoperative survival outcomes in lung cancer surgery, demonstrating a high predictive accuracy of 91.49%. By combining SHAP-based feature selection and an attention mechanism, the model enhances interpretability, aligning with clinical reasoning and focusing on important features like pulmonary function and comorbidities. These advancements showcase how AI can go beyond traditional statistical methods, offering a deeper understanding of complex medical data while preserving transparency and trust in the decision making process. Integrating AI in this framework is pivotal to precision medicine, enabling clinicians to make more informed decisions tailored to individual patients. The model's ability to pinpoint key prognostic factors allows for personalized preoperative assessments and risk stratifications. This supports optimizing treatment plans, improving postoperative care and potentially leading to better survival rates for lung cancer patients.

**Author Contributions:** Conceptualization, F.C. and E.C.; methodology, F.C. and E.C.; software, F.C. and E.C.; validation, F.C. and E.C.; formal analysis, F.C. and E.C.; investigation, F.C. and E.C.; data curation, F.C. and E.C.; writing—original draft preparation, F.C. and E.C.; visualization, F.C. and E.C.; writing—review and editing, D.O.; supervision, D.O. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable. The study used publicly available datasets, and no new data were collected directly from individuals.

**Informed Consent Statement:** Not applicable. The study used publicly available datasets, and no new data were collected directly from individuals.

**Data Availability Statement:** The original data presented in the study are openly available in Kaggle repository: Thoracic Surgery Dataset at <https://www.kaggle.com/datasets/sid321axn/thoracic-surgery>.

**Acknowledgments:** We thank our three peer reviewers for pointing out areas where the text needed clarification.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Ancel, J.; Bergantini, L.; Mendogni, P.; Hu, Z. Thoracic Malignancies: From Prevention and Diagnosis to Late Stages. *Life* **2025**, *15*, 138. <https://doi.org/10.3390/life15020138>.
2. Kaggle. Thoracic Surgery Data Set. Available online: <https://www.kaggle.com/datasets/sid321axn/thoracic-surgery> (accessed on 21 November 2024).
3. Van den Broeck, G., Lykov, A., Schleich, M., and Suciu, D. On the tractability of SHAP explanations. *J. Artif. Intell. Res.* **2022**, *74*, 851–886.
4. Niu, Z.; Zhong, G.; Yu, H. A review on the attention mechanism of deep learning. *Neurocomputing* **2021**, *452*, 48–62.
5. Onchis, D.M.; Costi, F.; Istin, C.; Secasan, C.C.; Cozma, G.V. Method of Improving the Management of Cancer Risk Groups by Coupling a Features-Attention Mechanism to a Deep Neural Network. *Appl. Sci.* **2024**, *14*, 447.

6. Rahman, M.M.; Munir, M.; Marculescu, R. EMCAD: Efficient Multi-scale Convolutional Attention Decoding for Medical Image Segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 16–22 June 2024; pp. 1–14.
7. Biosymetrics. Feature Engineering of Electronic Medical Records. Available online: <https://www.biosymetrics.com/blog/emr-feature-engineering> (accessed on 21 November 2024).
8. Liu, Y.; Wang, H.; Zhang, X. A Comparative Study of Normalization Techniques in Deep Learning Models for Medical Image Analysis. *Appl. Sci.* **2023**, *13*, 1234.
9. Kor, C.-T.; Li, Y.-R.; Lin, P.-R.; Lin, S.-H.; Wang, B.-Y.; Lin, C.-H. Explainable Machine Learning Model for Predicting First-Time Acute Exacerbation in Patients with Chronic Obstructive Pulmonary Disease. *J. Pers. Med.* **2022**, *12*, 228.
10. Molnar, C. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*, 2nd ed.; Leanpub: Victoria, BC, Canada, 2022.
11. Lundberg, S.M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4765–4774.
12. Costi, F.; Onchis, D.M.; Istin, C.; Cozma, G.V. Explainability-Enhanced Neural Network for Thoracic Diagnosis Improvement. In Proceedings of the International Conference on Computer Analysis of Images and Patterns, Limassol, Cyprus, 28–30 September 2023; pp. 35–44.
13. Liu, Y.; Zhang, X.; Wang, H. Application of SHAP Values in Predicting Postoperative Survival in Thoracic Surgery Patients. *J. Clin. Med.* **2023**, *12*, 1234.
14. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is All You Need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5998–6008.
15. Livadiotis, G. Linear Regression with Optimal Rotation. *Stats* **2019**, *2*, 416–425. <https://doi.org/10.3390/stats2040028>
16. Belgiu, M.; Drăguț, L. Comparison of Random Forest, Support Vector Machines, and Neural Networks for Forest Species Classification Using Airborne Hyperspectral Data. *Remote Sens.* **2021**, *13*, 2581. <https://doi.org/10.3390/rs13132581>
17. Liu, Y.; Zhang, W.; Chen, J. Advanced Short-Term Load Forecasting with XGBoost-RF Feature Selection and CNN-GRU Neural Network. *Processes* **2024**, *12*, 2466. <https://doi.org/10.3390/pr12112466>
18. GeeksforGeeks. Binary Cross Entropy/Log Loss for Binary Classification. 2023. Available online: <https://www.geeksforgeeks.org/binary-cross-entropy-log-loss-for-binary-classification/> (accessed on 21 November 2024).
19. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. In Proceedings of the 3rd International Conference on Learning Representations (ICLR), San Diego, CA, USA, 7–9 May 2015. arxiv:1412.6980.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.