Glass Box Explainability in AI on Lung Cancer Prediction

Vishakha Mistry

Head of Department, Department of Information Technology, 360 Research Foundation, Tumkaria, Bihar, India

---

Abstract - *According to the WHO, Lung cancer is the most common cause of cancer-related fatalities globally, accounting for the highest death rates among both men and women. As a result, identifying, diagnosing, and predicting lung cancer at an early stage is critical. This paper investigates the feasibility of predicting lung cancer illness using machine learning black box models and interpreting the findings using the machine learning package InterpreteML.*

Key Words: **Machine Learning, Lung cancer detection, explainable Artificial Intelligence (XAI), Explainable Boosting Machines (EBMs).**

1. INTRODUCTION

Cancer is more than simply an illness. There are several varieties of cancer, which can occur anywhere in the body. Lung cancer is defined as the uncontrolled growth of abnormal cells in the lungs. Lung cancer normally does not create apparent symptoms until it has progressed across the lungs. So this type of cancer is more serious than many other types of cancers. The survival rate is determined by the extent of the cancer's spread. As a result, early detection of cancer can make a significant impact.

Machine Learning has widespread applications in the real world. The uses of ML in healthcare are rising, and it benefits patients and professionals in various ways. Machine learning enables us to relate existing data to future illness estimates.

Though Machine Learning solves complicated issues, it may also be a black box, without explaining why or how judgments are made. This may confuse, especially when consequences relate to human lives or healthcare applications. InterpretML meets these demands. InterpretML supports both interpretable glassbox models and non-interpretable black box models. There are two basic categories of interpretability: global and local. The purpose of this research is to use the ML model- Explainable Boosting Machine under InterpreteML API to predict lung cancer at an early stage, highlighting both accuracy and transparency.

2. LITERATURE SURVEY

Various researchers have applied machine learning to predict lung cancer. Many researchers have not employed the interpretability technique with an Explainable Boosting machine. Here, I evaluated a variety of lung cancer prediction research publications and interpretable machine-learning research papers.

Authors, Gaoyang Liu and Bochao Sun in [1] have used EBM for compressive strength prediction. Concrete mix design data was collected from the UCI repository. EBM, Random forest, Decision Tree, XGBoost are performed over downloaded datasets. The authors evaluated the performance of EBM and other ML models and found that the EBM algorithm outperformed the other models.

The authors of [2] presented a case study to predict which patients are most likely to be readmitted to the hospital within 30 days of being released. The AUC of Logistic Regressor, Random Forest, and Generalized Additive Model with pairwise interaction (GA^2M) have been compared and they have shown that give the best accuracy and as well as maintain intelligibility.

Senthil and B. Ayshwarya [3] has presented Lung Cancer prediction using Feed Forward Back Propagation Neural Networks with Optimal Features. Lung cancer feature extraction is done by particle Swarm optimization (PSO) technology. Performance comparison of KNN, SVM, Bayes Network, and proposed NN-PSO has been shown. The proposed method has demonstrated remarkable accuracy, and NN-PSO can be used effectively by Lung Cancer oncologists.

Authors in [4] have proposed SVM classifier on the Lung cancer dataset. According to the assessment results, SVM with two rounds of SMOTE resampling is performed on the Lung Cancer dataset to achieve the greatest performance. The accuracy of the KNN method is 68.9%, but SVM reaches 98.8% accuracy.

MRI dataset is taken for the prediction of Alzheimer's disease by the Authors [5]. After dealing with missing values and categorical data, Chi-square and L1 regularization are used to choose features. The authors demonstrated that both strategies can generate superior outcomes. The suggested model is compared against ResNet-50, RF Classifier, Deep CNN, and VGG16-LIME. Overall, EBM achieves 94.35% accuracy.

3. LUNG CANCER DATASET DESCRIPTION AND PREPROCESSING

The data was obtained from the Kaggle website. The dataset has 309 instances and 16 attributes such as Gender, Age, Smoking, Yellow fingers, Anxiety, Peer pressure, Chronic Disease, Fatigue, Allergy, Wheezing, Alcohol, Coughing, Lung Cancer, Chest pain, Swallowing Difficulty, Shortness of

Breath. I needed to perform some preprocessing and adjustments so that the data is a more acceptable ML algorithm. Label encoding transforms categorical variables into numerical values. However, the age property has been adjusted to match the true one. The age normalization formula [6] is stated below:

$$\text{new value} = \frac{(old\ value - min(a_1.....a_n))}{(Max(a_1....a_n) - Min(a_1....a_n))}$$

Figure 1 depicts a significantly uneven target distribution. Before using ML algorithm, we need to handle this imbalance.

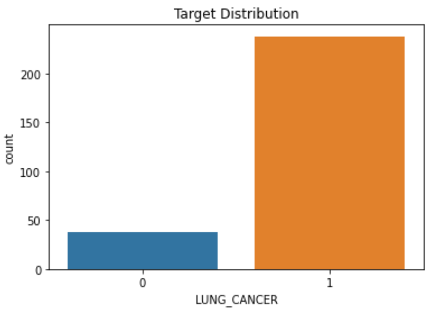

Fig -1: Target distribution

I employed the Adaptive Synthetic Sampling Approach (ADASYN) [7] for Imbalanced Learning. Based on the original data distribution, ADASYN can adaptively generate synthetic data samples for the minority class to reduce the bias introduced by the imbalanced data distribution ADASYN can generate synthetic data samples for the minority class based on the original distribution, reducing bias caused by imbalances [7]. ADASYN can focus the classifier decision boundary on challenging cases, leading to improved learning performance [7].

4. GLASS BOX VS BLACK BOX EXPLAINABILITY

Before presenting the findings, the clarity of the black box and the glass box explainability is essential.

Usually, we don't fully comprehend the reasoning behind the decisions or actions of AI systems. The kind of ML algorithms used will determine how explanations are produced. Glass box models include machine learning algorithms, including Bayesian classifiers, decision trees, and linear models. These algorithms provide interpretability of decision-making with a manageably small number of internal components.

However, all deep learning algorithms are referred to as "black box" models, which compromise transparency in favour of prediction accuracy. Neural nets, random forests, and Boosted trees are a few black-box model examples [8].

Techniques known as XAI are developed to explain the black box models. A few well-known XAI methods that tackle the trade-off between explanation and prediction in deep learning models are LIME, SHAP, and PDP.

InterpretML is Microsoft's open-source package that comprises ML interpretability techniques [8]. It exposes two types of interpretability- glassbox and blackbox. This package also has a visualization platform[8].

5. INTERPRETABLE LUNG CANCER PREDICTION USING EBM AND RESULTS

An explainable boosting machine classifier is used on the Lung Cancer dataset and the model has achieved an average accuracy of 0.9502. The following sections provide an explanation of the results and interpretable graphs.

5.1 Global explanation

The figure depicts the overall relevance of each feature. The overall relevance of each feature is calculated by averaging the absolute predicted values of each feature in the training set.

The behavior of a model can be explained by using feature importance. It shows the number of records that a feature has a significant impact on. By averaging the absolute values of a feature's influences across training datasets, feature significance is determined. Sample weights (if any) and the quantity of samples in each bin are used to weight contributions. The top 15 terms are displayed in sFigure 2.

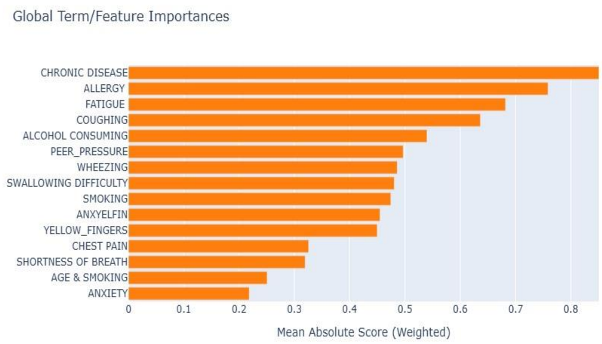

Fig -2: Overall relevance of each feature

The graphic illustrates the contribution of term age to model prediction is indicated in Figure 3. It shows that age has no effect till 50. Following this, age has the greatest impact on prediction.

Fig -3: Contribution of the term Age to the model prediction

Figure 4 depicts the contribution of the two features, smoking, and allergy, to the model's predictions. Similarly, the influence of numerous other features may be observed using similar graphs.

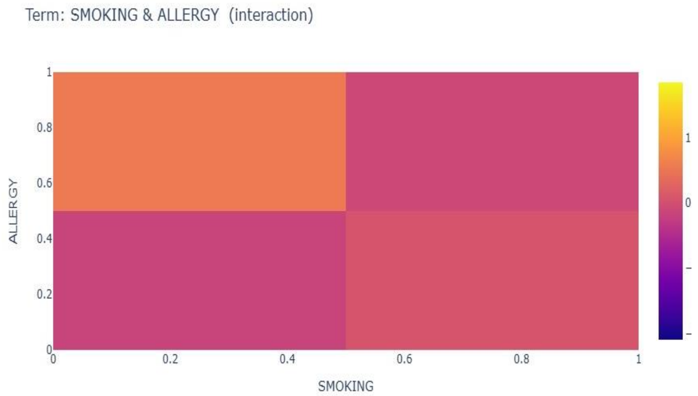

Fig -4: Contribution of the term Smoking and Allergy to the model prediction

5.2 Local explanation

Global feature explanation indicates which columns had the biggest influence on the forecast. However, a local explanation is required if I wish to examine each forecast separately. Local explanation graph will show actual class as well as predicted class. When examining columns, each row will have a distinct graph illustrating its reasons for making decisions.

These graphs provide an explanation for a person's classification as either having lung cancer or not. For the observation shown in Figure 5, the actual class is 1 and the predicted class is 1, and the prediction probability is 0.991. The orange feature in the illustration is positive, whereas the blue feature is negative. Features that cause negative effects include coughing, allergies, exhaustion, wheezing, and shortness of breath. But the forecast is positive when all the features are added together.

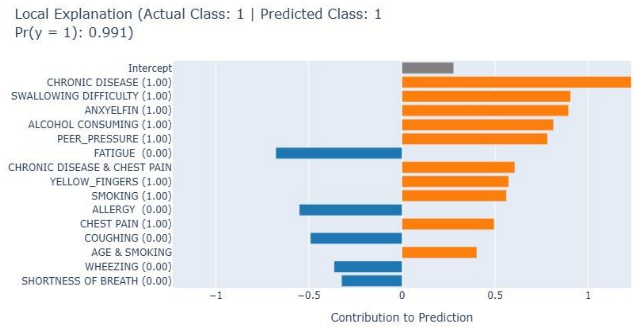

Fig -5: Local explanation for Actual Class=1 and Predicted Class=1

Additionally, as can be seen in Figure 6, the predicted class is zero, the actual class is zero, and the prediction probability is 0.994. Two features that predict to the inappropriate class are smoking and chest pain. However, the forecast turns out to be negative after all features are added.

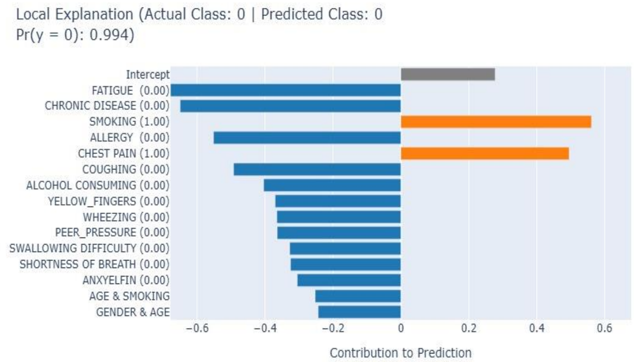

Fig -6: Local explanation for Actual Class=0 and Predicted Class=0

A misclassification observation is shown in Figure 7, where the predicted class is 1 and the actual class is 0. Chronic disease, swallowing difficulty, wheezing, and shortness of breath are the features that lead to incorrect classification. This aids in data debugging and coding enhancement.

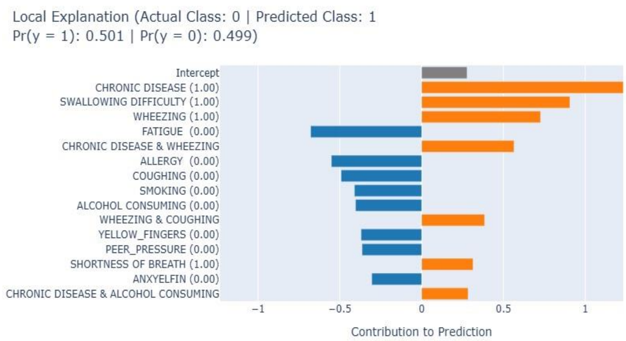

Fig -7: Local explanation for Actual Class=0 and Predicted Class=1

Likewise, another instance of misclassification is shown in Figure 8, where the predicted class is zero and the actual class is one. The primary characteristics that lead to misclassification include fatigue, chronic illness, and allergies.

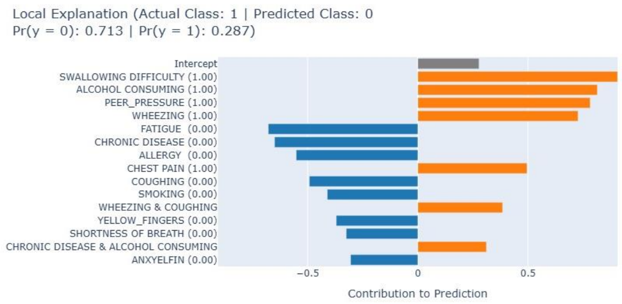

Fig -8: Local explanation for Actual Class=1 and Predicted Class=0

6. CONCLUSION AND FUTURE WORK

Main goal of current research is to accurately diagnose and classification of lung cancer using EBM. The model gives accurate results and provides insights into the factors affecting the prediction results. This allows human users to comprehend and trust the results and output created by machine learning. This model helps clinicians to make decisions confidently. With the help of local and global explanations, further insights into the data are possible.

In the future, this work can be extended for other disease detection. We will develop a more comprehensive dataset with real-world clinical features. Also, we will translate the algorithm into a user-friendly interface where doctors and healthcare professionals can utilize it easily.

REFERENCES

[1] Gaoyang Liu, Bochao Sun, “Concrete compressive strength prediction using an explainable boosting machine model”, Case Studies in Construction Materials, Volume 18, July 2023, Elsevier, e01845

[2] R. Caruana, Y. Lou, J. Gehrke, P. Koch, M. Sturm, and N. Elhadad, “Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission”, in Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 1721–1730. ACM, 2015.

[3] Dr. S. Senthil, B. Ayshwarya, “Lung Cancer Prediction using Feed Forward Back Propagation Neural Networks with Optimal Features”, International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 1 (2018) pp. 318-325

[4] C. Anil Kumar, S. Harish, Prabha Ravi, Murthy SVN, B. P. Pradeep Kumar, V. Mohanavel, Nouf M. Alyami, S. Shanmuga Priya, Amare Kebede Asfaw, “Lung Cancer Prediction from Text Datasets Using Machine Learning”, BIOMED RESEARCH INTERNATIONAL, Wiley, https://doi.org/10.1155/2022/6254177

[5] Archana Menon p, r. Gunasundari, “Intelligent Alzheimer’s disease prediction using Explainable Boosting Machine”, Journal of Theoretical and Applied Information Technology, 31st March 2024, Vol.102. No 6

[6] Ibrahim M. Nasser, Samy S. Abu-Naser, “Lung Cancer Detection Using Artificial Neural Network”, International Journal of Engineering and Information Systems (IJEAIS), Vol. 3 Issue 3, March – 2019, Pages: 17-23

[7] Haibo He, Yang Bai, Edwardo A. Garcia, and Shutao Li, “ADASYN: Adaptive Synthetic Sampling Approach for Imbalanced Learning”, 2008 International Joint Conference on Neural Networks (IJCNN 2008)

[8] Nori, Harsha and Jenkins, Samuel and Koch, Paul and Caruana, Rich, “InterpretML: A Unified Framework for Machine Learning Interpretability.” ArXiv abs/1909.09223 (2019)

BIOGRAPHY

Vishakha is working as a HOD of the Department of IT at 360 Research Foundation, she establishes a research agenda in **AI/ML**, helps researchers in the field of empowerment and livelihood, and guides research work to engineering and computer science students. She earned her M.Tech. from (NIT), Surat, Gujarat. She is an experienced Assistant Professor with a variety of engineering educational institutes including NIT-Surat and R V College of Engineering, Bangalore. She is passionate about **AI/ML, Explainable AI(XAI),** speech and audio signal processing, and Digital Signal Processing.