P-ISSN: 2656-1743 | E-ISSN: 2656-1735 **JURNAL RISET INFORMATIKA**
DOI: https://doi.org/10.34288/jri.v7i1.354 Vol. 7, No. 1. December 2024
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

Explainable AI-Driven TabNet Model Enhanced with Bayesian Optimization for Lung Cancer Prediction and Interpretation

Ilham Maulana

Ilmu Komputer / Fakultas Teknologi Informasi
Universitas Nusa Mandiri
k4ilham@gmail.com

Abstract

This study aims to develop an accurate and explainable lung cancer risk prediction model using a TabNet approach optimized with Bayesian Optimization and applying Explainable AI (XAI) methods through LIME (Local Interpretable Model-Agnostic Explanations). TabNet was selected for its efficiency in processing tabular data and its ability to produce high-accuracy predictions. In the initial stage, the TabNet model was tested using a dataset that was preprocessed through standardization and split into training and testing sets. The performance evaluation of the model without optimization showed an accuracy of 95.83%, precision of 95.87%, recall of 95.76%, and F1-Score of 95.81%. Subsequently, Bayesian Optimization was applied using the Optuna library to find the best hyperparameter combination for the TabNet model. The optimization results demonstrated a significant improvement, achieving an accuracy of 98.33%, precision of 98.48%, recall of 98.21%, and F1-Score of 98.32%. After optimizing the TabNet model, LIME was implemented to provide interpretability for the generated predictions. LIME was used to identify the most influential features contributing to the predictions, enhancing the model's transparency in the lung cancer risk prediction process. Through the combination of TabNet, Bayesian Optimization, and Explainable AI, this study successfully developed a lung cancer prediction model that is not only accurate but also highly interpretable. This model can assist medical professionals in identifying key risk factors and providing transparent explanations for each prediction made.

Keywords: TabNet, Bayesian Optimization, Explainable AI, LIME, Lung Cancer, Risk Prediction.

Abstrak

Penelitian ini bertujuan untuk mengembangkan model prediksi risiko kanker paru-paru yang akurat dan dapat dijelaskan dengan menggunakan pendekatan TabNet yang dioptimalkan dengan Bayesian Optimization serta diterapkan metode Explainable AI (XAI) menggunakan LIME (Local Interpretable Model-Agnostic Explanations). Model TabNet dipilih karena kemampuannya dalam memproses data tabular dengan efisiensi tinggi dan menghasilkan prediksi yang akurat. Pada tahap awal, model TabNet diuji menggunakan dataset yang telah diproses melalui teknik standarisasi dan pembagian data menjadi training dan testing sets. Evaluasi performa model tanpa optimasi menunjukkan nilai akurasi 95.83%, precision 95.87%, recall 95.76%, dan F1-Score 95.81%. Selanjutnya, dilakukan proses Bayesian Optimization menggunakan pustaka Optuna untuk menemukan kombinasi hyperparameter terbaik pada model TabNet. Hasil optimasi menunjukkan peningkatan signifikan dengan nilai akurasi 98.33%, precision 98.48%, recall 98.21%, dan F1-Score 98.32%. Setelah model TabNet dioptimalkan, diterapkan metode LIME untuk memberikan interpretasi terhadap prediksi yang dihasilkan. LIME digunakan untuk mengidentifikasi fitur-fitur yang memiliki kontribusi terbesar terhadap hasil prediksi, sehingga meningkatkan transparansi model dalam proses prediksi risiko kanker paru-paru. Dengan kombinasi TabNet, Bayesian Optimization, dan Explainable AI, penelitian ini berhasil mengembangkan model prediksi kanker paru-paru yang tidak hanya akurat, tetapi juga dapat dijelaskan dengan baik. Model ini dapat membantu tenaga medis dalam mengidentifikasi faktor risiko utama serta memberikan penjelasan yang transparan terhadap setiap prediksi yang dihasilkan.

Kata Kunci: TabNet, Bayesian Optimization, Explainable AI, LIME, Kanker Paru-Paru, Prediksi Risiko.

INTRODUCTION

Early detection of lung cancer is a crucial component in improving patient survival rates. Lung cancer remains one of the most common and deadliest types of cancer worldwide, with statistics showing a high prevalence (Sung et al. 2021). Effective treatment depends on the ability to detect the disease at an early stage, where therapeutic intervention is more effective (Smith et al. 2022). Traditional methods, such as low-dose CT imaging, have proven helpful in detecting lung cancer in high-risk individuals; however, accessibility and cost remain significant challenges.

Over the past decade, the development of risk prediction models has become increasingly important as an alternative approach for early detection. These models analyze tabular data that includes risk factors such as age, smoking history, and underlying respiratory diseases (Chandran et al. 2023). For example, models like the Liverpool Lung Project version 3 (LLPv3) combine demographic and clinical factors to predict the risk of lung cancer (Zhang et al. 2022). Studies have shown that these factors not only contribute to risk assessment but can also be integrated with machine learning methods to enhance predictive accuracy.

Deep learning models have demonstrated significant potential across various predictive applications, including healthcare and cancer diagnosis. However, most traditional deep learning architectures are not designed to process tabular data, which is often the primary format used in medical records and epidemiological studies. Tabular data contains important structured information about patients, such as age, gender, medical history, and laboratory test results. Due to the limitations of existing architectures in handling such datasets effectively, the development of new models like TabNet has become highly relevant.

TabNet is a deep learning architecture that adopts an attention mechanism for efficient feature selection and optimized processing of tabular data. Unlike conventional neural networks, TabNet is specifically designed to handle the complexity and heterogeneity of tabular data without losing critical information (Nguyen and Byeon 2023). Studies have shown that TabNet not only improves predictive accuracy but also offers interpretability—an essential aspect in medical contexts where decisions often need to be explained to both patients and healthcare professionals.

For example, in the field of oncology, the use of TabNet can aid in building more accurate predictive models for lung cancer diagnosis. Research utilizing TabNet has demonstrated that this model can process data more efficiently, enhance diagnostic accuracy, and enable medical teams to carry out faster and more informed interventions (Lee, Chao, and Hsu 2021) (Tao et al. 2022). By leveraging the advantages of the attention mechanism, TabNet is capable of identifying patterns in tabular data that might be overlooked by traditional methods, thereby improving the predictive capabilities in diagnosing diseases such as lung cancer and beyond.

The development of the TabNet model, while promising in generating accurate predictions from tabular data, faces challenges related to effective hyperparameter tuning. Conventional hyperparameter tuning methods such as grid search and random search are often inefficient, particularly when applied to models with high complexity. In contrast, more advanced approaches like Bayesian Optimization—implementable through libraries such as Optuna—have proven to be more efficient solutions for identifying optimal hyperparameter combinations.

Bayesian Optimization leverages a probabilistic model to estimate the objective function being optimized, which helps identify hyperparameter values that are more likely to yield optimal model performance. In the context of TabNet, studies have shown that this method can significantly enhance model performance, as TabNet heavily relies on well-tuned hyperparameters to achieve the desired level of accuracy (Arık and Pfister 2021) (Nguyen and Byeon 2024). For instance, the use of Optuna for hyperparameter tuning has been reported to produce substantial improvements in performance metrics, making it a valuable tool for developing more effective models in healthcare applications and beyond.

A study by (Sun et al. 2024). demonstrated that applying Bayesian Optimization to TabNet leads to significantly better performance compared to traditional hyperparameter tuning methods. They reported that the model's performance improved markedly both before and after the tuning process, highlighting the effectiveness of optimization in enhancing the accuracy and reliability of TabNet for real-world applications (Sun et al. 2024). This approach also offers the added benefit of improved model interpretability, enabling researchers to better explain the model's decisions based on the features selected during training.

Data balancing techniques such as oversampling, SMOTE, and ADASYN are commonly

used to address class imbalance issues that frequently occur in various machine learning applications. For instance, the implementation of SMOTE and SMOTETomek has been proven effective in improving prediction accuracy on imbalanced datasets (Indra, Maulana, and Ernawati 2024). However, ADASYN, which employs an adaptive approach to generate synthetic samples, is considered superior in producing more representative examples of the minority class—especially in cases where the data distribution is highly skewed (Maulana, Ernawati, and Indra 2024)Consequently, the use of ADASYN can more effectively enhance the performance of predictive models, particularly in contexts involving datasets with significant class variation.

In the context of medical applications, the interpretability of deep learning models is critically important—particularly given that high predictive accuracy alone is not sufficient. Understanding how a model makes its decisions is equally vital, especially in the diagnosis and treatment of cancer kanker (Ahmed et al. 2021). Deep learning models are often perceived as "black boxes," meaning their decision-making processes are not always transparent. To address this challenge, Explainable AI (XAI) techniques such as LIME (Local Interpretable Model-agnostic Explanations) are applied. LIME provides insights into which features most influence the model's predictions, enabling clinicians and researchers to better understand the rationale behind each decision.

The LIME method works by generating a simple local model around the prediction made by a more complex model. This allows LIME to identify features that significantly contribute to a given prediction, producing explanations that are understandable to humans (Raptis, Ilioudis, and Theodorou 2024). Studies have shown that the application of LIME in analyzing medical imaging data, such as radiographs for lung cancer detection, has yielded results that clarify the model's decision-making process and enhance user trust in the system (Ahmed et al. 2021) (Gandhi et al. 2023). For example, in studies exploring variability in lung cancer mortality, XAI techniques—including LIME—have been used to demonstrate how specific factors contribute to the model's predictive outcomes.

This study aims to develop a lung cancer risk prediction model using TabNet, enhanced with Bayesian Optimization for hyperparameter tuning, and evaluated through Explainable AI techniques using the LIME method. The combination of these approaches is expected to produce a model that is highly accurate, efficient, and interpretable—enabling medical professionals to identify key risk factors and understand the rationale behind each prediction. By improving both accuracy and transparency, this approach seeks to enhance the reliability and practical applicability of predictive models for lung cancer risk assessment.

RESEARCH METHODS

This study was conducted through several key stages, including data collection, data preprocessing, predictive model development using TabNet, hyperparameter optimization with Bayesian Optimization, application of Explainable AI using LIME, and model evaluation. The dataset used contains various features related to lung cancer risk, which were processed using techniques such as standardization, encoding, and splitting into training and testing sets.

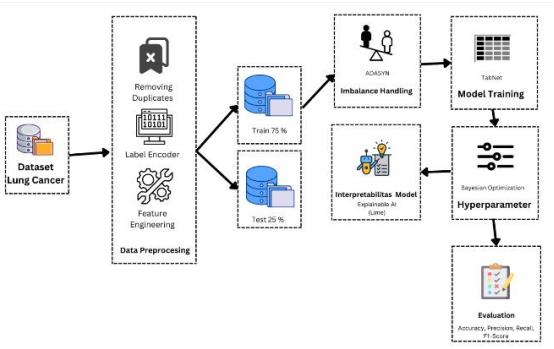

Figure 1. Research Stages

The TabNet model was built using the PyTorch TabNet library, incorporating various hyperparameters such as n_d, n_a, n_steps, gamma, lambda_sparse, optimizer, and learning rate. Subsequently, hyperparameter optimization was carried out using Bayesian Optimization through the Optuna library, aiming to maximize the model's accuracy. The best-performing model was then interpreted using Explainable AI methods with LIME to provide insights into its predictions, particularly by identifying the most influential features. Model performance was evaluated using accuracy, precision, recall, and F1-Score metrics, both before and after the hyperparameter optimization process.

P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/10.34288/jri.v7i1.354
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

Dataset

This study utilizes the "Lung Cancer Prediction Dataset" available on Kaggle, which consists of 309 samples with 16 descriptive features and one target column indicating lung cancer status. The dataset includes information reflecting risk factors related to lifestyle, medical history, and physical symptoms experienced by individuals, allowing for a holistic approach to lung cancer risk prediction.

Explicitly, previous studies using Kaggle datasets have shown that data structures containing a range of significant features per individual provide an effective foundation for building predictive models. For example, (Moozhippurath and Natarajan 2025) utilized the "Lung Cancer Prediction Dataset" from Kaggle, consisting of 309 instances and 16 attributes, to develop a graph neural network-based model. Their findings demonstrated that clustering risk factor and symptom information can improve predictive accuracy. This approach confirms the relevance of using such data in the context of in-depth evaluation of lung cancer risk.

In addition, (Ji, 2024) applied deep learning models such as bidirectional LSTM and GRU using a dataset with 16 similar features, providing insight that the variety of data types—ranging from clinical indicators to lifestyle patterns—greatly influences the design of algorithms capable of detecting lung cancer at an early stage. The combined findings from these studies support the selection of this dataset to identify critical factors correlated with the emergence of lung cancer, and also provide a strong foundation for further exploration through exploratory data analysis (EDA) and the testing of various classification methods, as also noted by (Aqila and Faisal 2023) in their implementation of Decision Trees. The detailed list of features used in this study is presented in Table 1.

Table 1. Features and Target Used in the Experiment

Type	Name	Description
Fitur	AGE	Age of the respondent
Fitur	SMOKING	Whether the respondent smokes
Fitur	YELLOW_FINGERS	Presence of yellow-stained fingers
Fitur	ANXIETY	Whether the respondent experiences anxiety
Fitur	PEER_PRESSURE	Social pressure to smoke
Fitur	CHRONIC DISEASE	History of chronic illness
Fitur	FATIGUE	Frequent tiredness or fatigue
Fitur	ALLERGY	Presence of allergies
Fitur	WHEEZING	Wheezing or high-pitched breathing sound
Fitur	ALCOHOL CONSUMING	Whether the respondent consumes alcohol
Fitur	COUGHING	Whether the respondent experiences frequent coughing
Fitur	SHORTNESS OF BREATH	Experience of breathlessness or shortness of breath
Fitur	SWALLOWING DIFFICULTY	Difficulty in swallowing
Fitur	CHEST PAIN	Experience of chest pain
Target	LUNG_CANCER	Diagnosis indicating presence or absence of lung cancer

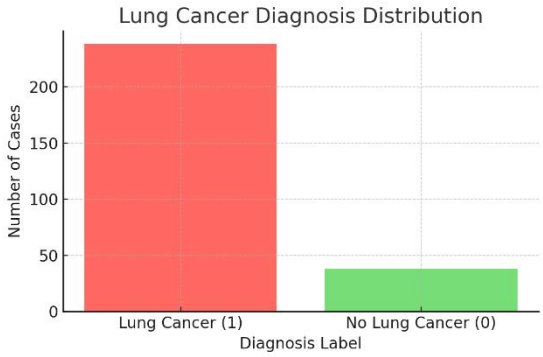

Figure 2. Data Distribution graph Based on Target

The visualization in Figure 2 displays the distribution of data based on the target labels in

the Survey Lung Cancer dataset. It is evident that the number of cases labeled as "Lung Cancer (1)" is significantly higher than those labeled as "No Lung Cancer (0)", with 238 and 38 samples, respectively. This imbalance indicates that the majority of individuals in the dataset are classified as having lung cancer, while only a small portion are not.

Such an imbalanced distribution is crucial to address, as it can lead to predictive model bias toward the majority class. If left unaddressed, the model may become highly accurate for the dominant class while performing poorly on the minority class. Therefore, specialized strategies such as resampling techniques, threshold adjustments, or the use of appropriate evaluation metrics (e.g., F1-Score or ROC-AUC) are necessary to ensure that the model remains fair and accurate across both classes.

Data Preprocessing

The Data Preprocessing stage is a crucial process carried out to ensure that the data used for model training is of high quality and ready for analysis. This process involves several key steps, including removing duplicate entries, checking for missing values, performing label encoding, analyzing the distribution of the target variable, visualizing data, eliminating irrelevant features, examining feature correlations, and conducting feature engineering.

The first step, Removing Duplicates, ensures that each row in the dataset is unique and that no repeated data skews the analysis. This is followed by Checking for Missing Values, which identifies any null entries across rows or columns. If missing data is found, appropriate action is taken—either by imputing the values (e.g., using mean or median) or by removing the affected entries. Next, Label Encoding is performed to convert categorical data into a numeric format that machine learning models can interpret. For instance, the column GENDER with values 'M' and 'F' is converted into 0 and 1, and the target variable LUNG_CANCER with 'YES' and 'NO' becomes 1 and 0.

After encoding, the Distribution of the Target Variable is analyzed to observe the balance between positive and negative classes. If the data is found to be imbalanced, techniques such as oversampling or undersampling may be considered. To gain deeper insight into relationships between features, Visualizations (Plotting) and Correlation Analysis are carried out. A correlation matrix helps identify features that are strongly related to each other. If high correlation is detected between two or more features, it may be necessary to remove or combine them through Feature Engineering, to reduce redundancy and improve model performance.

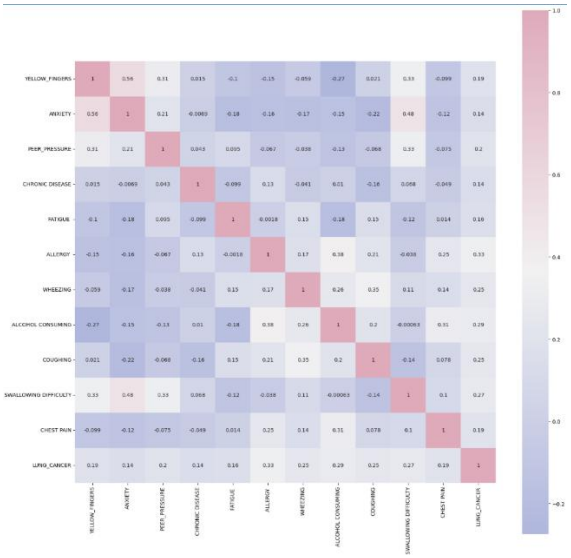

Figure 3. Heatmap of Feature Correlation Matrix in Lung Cancer Dataset

Figure 3 presents a heatmap of the correlation matrix among features in the Survey Lung Cancer Dataset. The colors in the heatmap represent the strength and direction of relationships between feature pairs—where light red approaching +1 indicates a strong positive correlation, and blue approaching -1 represents a strong negative correlation. From this visualization, it is evident that features such as SMOKING, YELLOW_FINGERS, and CHEST_PAIN have relatively higher positive correlations with the target label LUNG_CANCER, suggesting that these features may play a significant role in the classification process.

Conversely, features with correlation values close to zero exhibit weak relationships with the target variable, indicating they may be less relevant for prediction tasks. This correlation analysis is critical for feature selection, as it helps identify which variables contribute meaningfully to the model and which might be redundant or non-informative. Eliminating or transforming weakly correlated features can lead to more

efficient model training and improved performance.

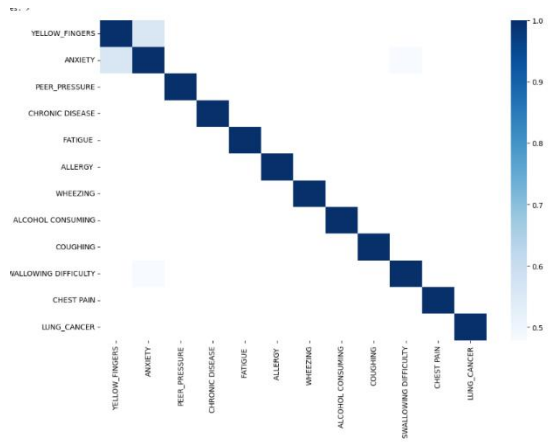

Figure **4**. Heatmap of Feature Multicollinearity Matrix in Lung Cancer Dataset

Figure 4 displays a heatmap of the multicollinearity matrix among features in the Survey Lung Cancer Dataset. This visualization focuses on the lower triangle of the correlation matrix to highlight pairwise relationships without visual redundancy. Darker colors indicate higher correlation values, while lighter colors represent weaker correlations.

The results reveal that most features exhibit low correlations with each other, indicating a low level of multicollinearity in the dataset. This condition is ideal for machine learning models, as it reduces the risk of redundant information across features and helps maintain both the stability and accuracy of the predictive model. Low multicollinearity ensures that each feature contributes uniquely to the learning process, improving interpretability and overall model reliability.

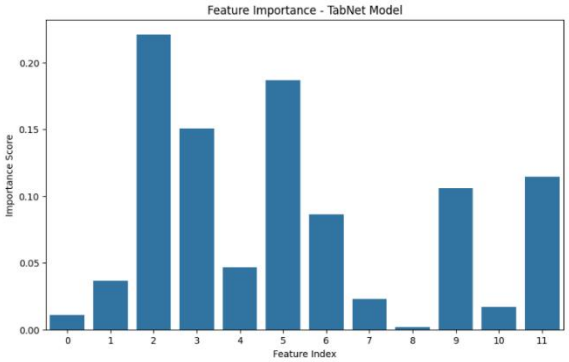

Figure 5. Feature Importance

Figure 5 presents a feature importance graph that illustrates the relative contribution of each feature to the prediction outcomes. The vertical axis represents the importance scores, while the horizontal axis indicates the feature indices. It is evident that certain features exert significantly greater influence than others, as indicated by importance scores exceeding 0.15.

Features with the highest importance are likely to have a strong relationship with the target variable (LUNG_CANCER) and thus become central to the model's decision-making process. This information is particularly valuable for model interpretation, feature selection, and validation of results in a way that is understandable to medical practitioners and decision-makers. By identifying the most impactful features, stakeholders can focus on the most relevant risk factors in both diagnostic processes and policy development.

Clasification Model

The classification model used in this study is TabNet, a deep learning architecture specifically designed for tabular data that incorporates an attention mechanism to dynamically select features during training. TabNet was chosen due to its strong capability to efficiently process structured data while maintaining a high degree of interpretability. The model produces accurate results by optimizing data processing through its integrated decision steps mechanism.

Prior to model training, a Handling Imbalance Data step was performed using the ADASYN (Adaptive Synthetic Sampling Approach for Imbalanced Learning) technique. This method addresses class imbalance in the target variable (LUNG_CANCER) between the positive and negative classes. ADASYN works by synthesizing new examples from the minority class based on the spatial proximity of existing samples. As a result, this process enhances the model's ability to detect minority class instances and reduces potential bias during training, ultimately leading to more balanced and reliable predictions.

The training and testing process was carried out by splitting the balanced dataset into two parts: 75% for training and 25% for testing. This division aims to objectively evaluate the model's performance using data that the model has not previously seen, thereby ensuring a more realistic assessment of its generalization capability.

The training of the TabNet model involved tuning several key hyperparameters, including n_d (decision representation size), n_a (attention representation size), n_steps (number of decision steps), gamma (relaxation parameter), lambda_sparse (sparsity regularization), learning rate, and mask type (e.g., sparsemax or entmax). To optimize the model, Bayesian Optimization was employed using the Optuna library, aiming to identify the best combination of hyperparameters by maximizing the model's accuracy metric. This approach ensures that the final model configuration is both efficient and tailored for the specific characteristics of the lung cancer dataset.

Bayesian Optimization applies Bayes' Theorem to select the best values in an optimization process. The equation for Bayes' Theorem is:

$$P(Z \mid Y) = \frac{(P(Y \mid Z)P(Z))}{P(Y)} \quad (1)$$

Where P(Z|Y) is the posterior probability. P(Y|Z) is the likelihood, which represents the probability of obtaining YYY given that ZZZ is known. P(Z) is the prior probability, which is the initial probability before observing the data YYY. P(Y) is the marginal probability.

Explainable AI

Explainable AI (XAI) is an approach in artificial intelligence that aims to make the prediction or decision-making process of models more transparent and interpretable. In the context of traditional machine learning and deep learning, complex models such as deep neural networks, TabNet, or ensemble methods are often viewed as "black boxes"—models whose internal logic is difficult for humans to understand. By implementing XAI, models can provide human-readable explanations that help increase trust and understanding of their outputs.

In this study, XAI is applied using the LIME (Local Interpretable Model-Agnostic Explanations) method to explain the predictions made by the optimized TabNet model. LIME works by generating a simple and interpretable model (such as linear regression or decision trees) around a specific prediction made by a complex model. By evaluating the influence of each feature on a local level, LIME provides clear and accurate explanations of how the model arrived at a particular decision. This enables users—particularly in medical and clinical settings—to better understand the rationale behind predictions, ultimately supporting more informed decision-making.

The LIME method provides two types of explanations: global explanations and local explanations. Global explanations focus on how the overall model functions and which features generally have the most influence on predictions. In contrast, local explanations concentrate on why a particular sample is classified in a specific way by the model. In this study, LIME is used to generate local explanations, allowing researchers to identify the most influential features contributing to the lung cancer risk prediction for each individual.

The visualizations generated by LIME highlight which features contribute positively or negatively to the model's prediction. In the resulting charts, green bars indicate features that support the prediction of a specific class (e.g., 'YES' for lung cancer), while red bars represent features that oppose it. The length of each bar reflects the magnitude of the feature's influence. By using LIME, researchers can pinpoint critical features for further analysis and provide clearer explanations to medical professionals or end users.

The application of Explainable AI using LIME to the TabNet model addresses the challenge of interpreting complex models. As a result, the model not only achieves high predictive accuracy but also offers clear justifications for each prediction it makes. This is especially crucial in the medical field, where transparency and interpretability are essential for making informed and trustworthy decisions.

Evaluation

Evaluation is a critical phase in this study to ensure that the developed model delivers accurate and interpretable predictions. The TabNet model in this research was evaluated using several key metrics: Accuracy, Precision, Recall, F1-Score, and AUC-ROC. Accuracy measures the proportion of correct predictions out of all predictions made by the model. While accuracy provides an overall picture of model performance, it can be misleading when the dataset is imbalanced between positive and negative classes. Therefore, additional evaluation metrics like Precision and Recall were employed. Precision assesses the model's ability to avoid false positives, which is crucial to ensure that lung cancer risk detection does not result in unnecessary alarms. Recall (or Sensitivity), on the other hand, evaluates the model's ability to correctly identify positive

cases, which is vital in medical applications where missing a true positive case could have serious consequences.

In addition, the F1-Score is used as a balanced measure that considers both Precision and Recall, especially in imbalanced datasets. The F1-Score provides a harmonic mean of Precision and Recall, offering a fairer representation of model performance under class imbalance conditions. To assess the model's overall ability to distinguish between positive and negative classes, the AUC-ROC metric is utilized. This metric visualizes the trade-off between the True Positive Rate and False Positive Rate across different thresholds, with an AUC value approaching 1 indicating excellent classification capability.

Evaluations were conducted both before and after hyperparameter optimization using Bayesian Optimization with Optuna, and the results showed a significant improvement in model performance after optimization. Furthermore, the integration of Explainable AI using LIME enhanced understanding of how each feature contributes to the model's predictions, thereby improving transparency and interpretability. Thus, this evaluation process ensures that the resulting TabNet model is not only highly accurate but also explainable and practical for effective lung cancer risk detection.

RESULTS AND DISCUSSION

This study proposes the implementation of an Explainable AI-driven TabNet as a predictive model for detecting lung cancer risk, with enhanced performance through hyperparameter optimization using Bayesian Optimization via Optuna. TabNet was chosen for its advantages in handling complex structured tabular data, as well as its interpretable learning capabilities, allowing it to identify the most influential features in decision-making processes. To achieve optimal performance, an automatic and efficient hyperparameter tuning process was conducted using the Bayesian approach, which is capable of exploring the parameter space more intelligently compared to conventional methods such as grid search or random search.

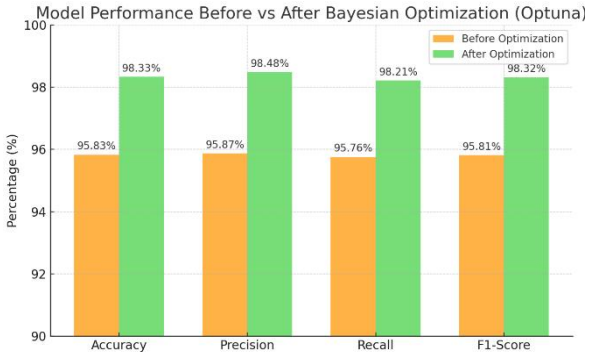

Figure 6. Model Performance Before vs After Bayesian Optimization

Figure 6 illustrates the comparison of TabNet model performance before and after the hyperparameter optimization process using Optuna. The initial model, which employed default hyperparameters, achieved an accuracy of 95.83%, precision of 95.87%, recall of 95.76%, and an F1-Score of 95.81%. After optimization, all four metrics showed significant improvement: accuracy increased to 98.33%, precision to 98.48%, recall to 98.21%, and F1-Score to 98.32%.

These results indicate that the hyperparameter optimization process using Bayesian Optimization effectively enhanced the model's classification performance. Furthermore, the improvement underscores the reliability of the explainable AI approach in delivering both accurate and accountable lung cancer predictions—making it a promising tool for real-world medical applications where both performance and interpretability are essential.

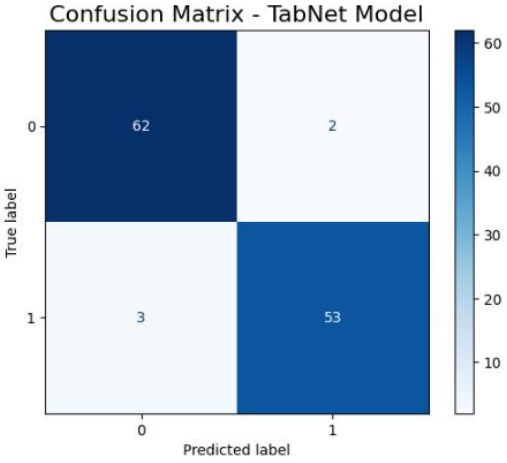

Figure 7. Confusion Matrix

Figure 7 displays the confusion matrix resulting from the predictions of the TabNet model after hyperparameter optimization. This matrix provides a detailed view of the model's classification performance in distinguishing between two classes: patients diagnosed with lung cancer (label 1) and those without (label 0). From the test dataset, the model correctly classified 62 negative samples (true negatives) and 53 positive samples (true positives). Meanwhile, there were only 2 false positives and 3 false negatives.

These values indicate that the model has a very low error rate and high accuracy, especially in correctly identifying lung cancer cases. The strong ratio of correct to incorrect predictions reinforces previous findings that the integration of TabNet with Bayesian Optimization produces a model that is not only accurate but also reliable—particularly in critical diagnostic contexts such as lung cancer detection.

Table 3. Comparison of Lung Cancer Prediction Models Based on Accuracy

Authors	Dataset	Best Algoritm	Accuracy(%)
(Zamzam et al. 2024)	Survey Lung Cancer (Kaggle)	CatBoost (with Bayesian Optimization)	95.7%
(Zamzam et al. 2024)	Survey Lung Cancer (Kaggle)	Random Forest (with Bayesian Optimization)	97.1%
(Nemlander et al. 2022)	Survey Lung Cancer (Kaggle)	Stochastic Gradient Boosting (SGB)	81.5%
Proposed Model	**Survey Lung Cancer**	**TabNet (with Bayesian Optimization)**	**98.33%**

Table 3 presents a comparison of accuracy scores from various lung cancer prediction models applied to the Survey Lung Cancer Dataset from Kaggle. The models compared in this study include CatBoost, Random Forest, and Stochastic Gradient Boosting (SGB), all of which were previously optimized using Bayesian Optimization in prior research. According to Zamzam et al. (2024), CatBoost achieved an accuracy of 95.7%, while Random Forest performed slightly better with 97.1%. In contrast, a study by Nemlander et al. (2022) reported that SGB only reached 81.5% accuracy.

The proposed model in this study—TabNet with Bayesian Optimization—demonstrated the highest performance, achieving an accuracy of 98.33%. These findings confirm that the TabNet approach, when combined with advanced optimization methods like Optuna, not only competes with but significantly outperforms conventional models. The strength of TabNet lies not only in its high accuracy but also in its ability to provide clear feature interpretability, making it an ideal choice for medical applications that demand transparency and clarity in decision-making processes.

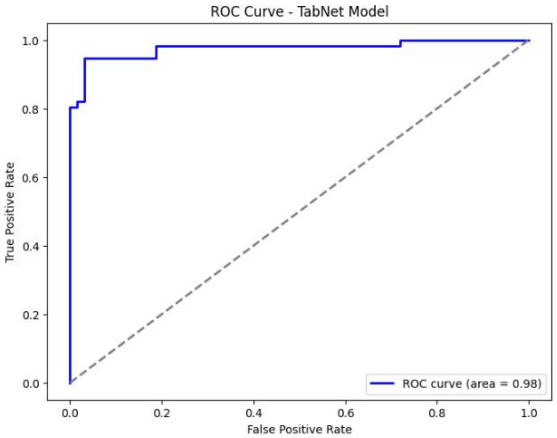

Figure 8. ROC Curve

Figure 8 illustrates the Receiver Operating Characteristic (ROC) curve of the TabNet model used for lung cancer prediction. The ROC curve represents the trade-off between the True Positive Rate (TPR) and the False Positive Rate (FPR) across various classification thresholds. The blue curve depicts the actual performance of the model, while the dashed gray line serves as a reference line for random guessing (AUC = 0.5). The Area Under the Curve (AUC) value of 0.98 indicates that the model possesses excellent—and nearly perfect—classification capability in distinguishing between patients with and without lung cancer.

This high AUC score reflects that TabNet excels not only in terms of accuracy but also in maintaining a strong balance between sensitivity (recall) and specificity. This balance is crucial in medical applications, where failing to detect true positive cases (false negatives) or incorrectly classifying healthy individuals (false positives) can have serious consequences. The ROC curve in Figure 8 reinforces the conclusion that the optimized Explainable AI approach using TabNet delivers highly reliable performance in medical classification tasks.

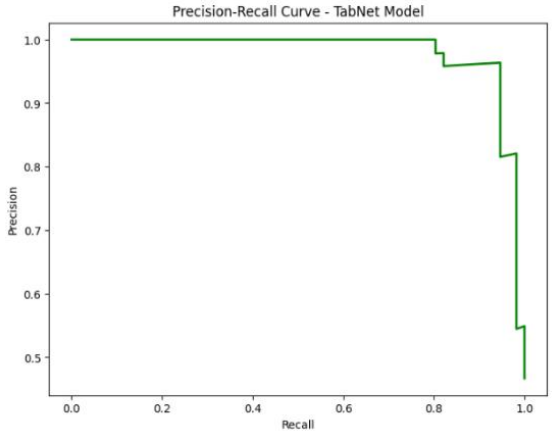

Figure 9. Precisian-Recall Curve

Figure 9 presents the Precision-Recall Curve of the TabNet model used for lung cancer prediction. This curve illustrates the relationship between recall (sensitivity) and precision across different prediction thresholds. A curve that trends closer to the top-left corner indicates excellent classification performance, especially in scenarios involving imbalanced datasets, such as this one where the number of cancer-positive cases significantly outweighs the negatives.

From the curve, it is evident that precision remains consistently high even as recall increases, up to a certain point before slightly declining. This pattern suggests that the model is capable of maintaining a high level of accuracy in identifying positive cases without producing an excessive number of false positives. In other words, the model is not only sensitive in detecting lung cancer but also selective in ensuring the relevance of its positive predictions. This Precision-Recall Curve complements the previous ROC Curve evaluation, offering additional evidence of the reliability and effectiveness of the TabNet model in medical classification tasks.

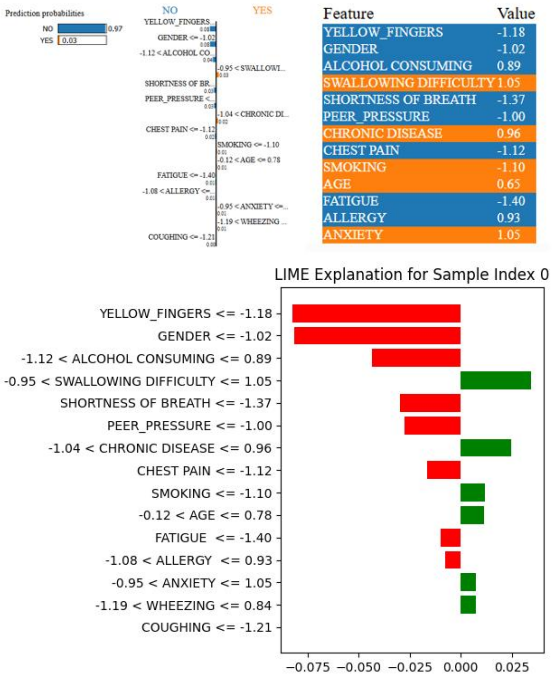

Figure 10. LIME Explanation for Feature Contributions in Lung Cancer Prediction

Figure 10 presents a LIME (Local Interpretable Model-agnostic Explanations) analysis of a single sample from a lung cancer prediction model. The prediction outcome shows a strong confidence toward the "NO" class, with a probability of 0.97, indicating that the model believes the patient does not have lung cancer. The top portion of the figure displays the actual values of the patient's features, such as YELLOW_FINGERS = -1.18, GENDER = -1.02, and ANXIETY = 1.05, providing context for how the model interpreted this individual input.

The lower part of Figure 10 visualizes the contribution of each feature toward the prediction using horizontal bars. Red bars represent features that pushed the prediction toward "NO", while green bars show features that supported a "YES" prediction. The most influential negative contributors were YELLOW_FINGERS, GENDER, and ALCOHOL CONSUMING, all of which heavily pulled the prediction away from a lung cancer diagnosis. On the other hand, features such as SWALLOWING DIFFICULTY and CHRONIC DISEASE had a positive impact, slightly nudging the prediction toward "YES", though not enough to overcome the dominant negative factors.

Figure 10 effectively demonstrates how interpretable machine learning can offer transparency in healthcare models. By breaking down the prediction into individual feature contributions, clinicians and researchers can

understand not only what the model predicted, but also why it made that decision. This level of interpretability is crucial when applying AI models to sensitive domains such as medical diagnosis, where trust and accountability are essential.

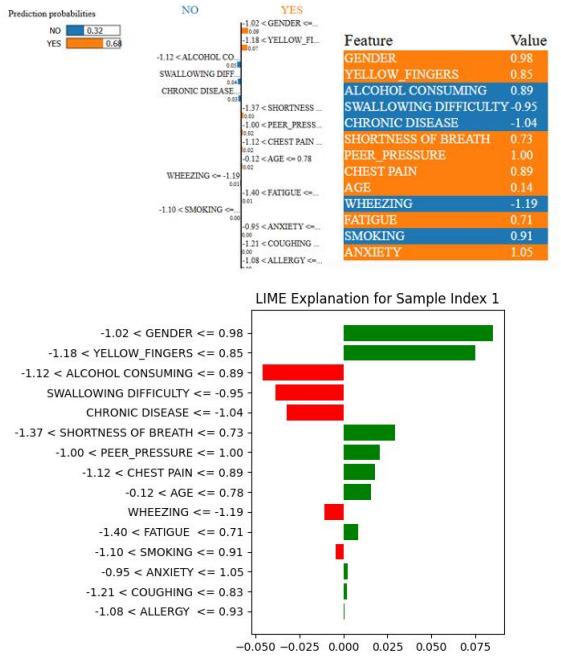

Figure 11. LIME Explanation for Feature Contributions in Lung Cancer Prediction

Figure 11 displays a LIME-based explanation for Sample Index 1 in a lung cancer prediction task. According to the model's output, there is a 68% probability that the patient has lung cancer ("YES"), while the probability for the "NO" class is 32%. The top section presents the actual values of the patient's features, including variables such as GENDER = 0.98, YELLOW_FINGERS = 0.85, and ANXIETY = 1.05. These inputs feed into the model and are used by LIME to interpret how each feature contributes to the prediction outcome.

The bottom part of Figure 11 shows a bar chart that visualizes feature contributions. Green bars indicate features that pushed the prediction toward "YES" (lung cancer), while red bars reflect those that contributed toward a "NO" prediction. In this case, features like GENDER, YELLOW_FINGERS, and SHORTNESS OF BREATH were strong positive contributors toward predicting lung cancer. Conversely, ALCOHOL CONSUMING, SWALLOWING DIFFICULTY, and CHRONIC DISEASE acted against the lung cancer prediction, pulling the model's confidence downward, though not strongly enough to change the final prediction.

Figure 11 highlights how LIME can identify both supporting and contradicting factors in a single prediction. By providing insight into how individual features impact the model's decision, this figure supports transparency and trust in machine learning predictions, especially in critical applications such as healthcare. This detailed view helps practitioners and researchers validate whether the model's logic aligns with medical understanding or requires further refinement.

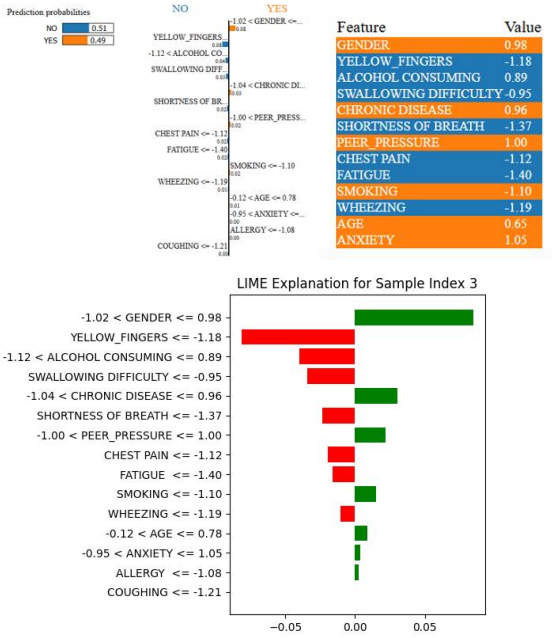

Figure 12. LIME Explanation for Feature Contributions in Lung Cancer Prediction

Figure 12 illustrates a LIME explanation for Sample Index 3 in the lung cancer prediction model. The model's prediction in this case is nearly evenly split, with a slight lean toward "NO" (no lung cancer) at 51%, and 49% toward "YES". This close probability suggests the model finds the input data ambiguous, with both classes receiving substantial support. The top section of the figure displays the input feature values for this individual, showing attributes like GENDER = 0.98, YELLOW_FINGERS = -1.18, and ANXIETY = 1.05.

The bar chart at the bottom of Figure 12 provides a visual representation of how each feature influenced the model's prediction. Green bars indicate support for the "YES" class, while red bars indicate support for "NO". Notably, GENDER, SWALLOWING DIFFICULTY, and CHRONIC DISEASE positively contributed toward a lung cancer prediction. However, features like

YELLOW_FINGERS, ALCOHOL CONSUMING, and CHEST PAIN significantly pushed the model's confidence toward predicting "NO", which ultimately swayed the final decision despite strong opposing signals.

Figure 12 highlights the interpretability power of LIME in understanding edge cases where the prediction is not definitive. By breaking down and quantifying the impact of each input feature, LIME enables analysts and medical professionals to investigate not only which decision the model made, but also the reasoning behind it. In critical applications like healthcare, this transparency is crucial for building trust and ensuring that model outputs are medically reasonable.

CONCLUSIONS AND SUGGESTIONS

Conclusion

This study successfully developed a highly accurate and interpretable lung cancer risk prediction model by combining TabNet, Bayesian Optimization, and Explainable AI (LIME). Hyperparameter optimization using Optuna proved effective in enhancing model performance—raising accuracy from 95.83% to 98.33%, along with significant improvements in precision, recall, and F1-Score. The application of the ADASYN technique to address class imbalance also contributed positively, improving the model's ability to identify minority class cases more effectively.

Furthermore, the implementation of LIME as an Explainable AI method provided transparent explanations for each prediction, enabling clear identification of key features influencing lung cancer risk. With these results, the proposed model demonstrates strong potential for use in medical settings, particularly for supporting early detection of lung cancer with predictions that are both accurate and understandable to healthcare professionals. Nonetheless, further research is needed to evaluate the model's generalizability across larger and more diverse datasets.

Suggestion

Future research is recommended to evaluate the developed model on larger and more diverse datasets to ensure its generalizability and robustness under various conditions. Additionally, further exploration of other Explainable AI methods such as SHAP (SHapley Additive exPlanations) or Integrated Gradients is necessary to compare their interpretive capabilities with LIME. Implementing more advanced explainability techniques and conducting global explanation analysis would provide a more comprehensive understanding of the model's behavior.

This research can also be extended by integrating ensemble learning techniques or combining TabNet with other models to further enhance predictive performance. Moreover, testing the model's effectiveness in other medical applications could serve as a promising area of study and significantly contribute to the advancement of AI-driven solutions in healthcare.

.

REFERENCES

Ahmed, Zia U., Kang Sun, Michael Shelly, and Lina Mu. 2021. "Explainable Artificial Intelligence (XAI) for Exploring Spatial Variability of Lung and Bronchus Cancer (LBC) Mortality Rates in the Contiguous USA." *Scientific Reports* 11(1):1–15. doi: 10.1038/s41598-021-03198-8.

Aqila, Aqila, and Muhammad Faisal. 2023. "Lung Cancer EDA Classification Using the Decision Trees Method in Python." *Informatics and Software Engineering* 1(1):8–13. doi: 10.58777/ise.v1i1.56.

Arık, Sercan, and Tomas Pfister. 2021. "TabNet: Attentive Interpretable Tabular Learning." *35th AAAI Conference on Artificial Intelligence, AAAI 2021* 8A:6679–87. doi: 10.1609/aaai.v35i8.16826.

Chandran, Urmila, Jenna Reps, Robert Yang, Anil Vachani, Fabien Maldonado, and Iftekhar Kalsekar. 2023. "Machine Learning and Real-World Data to Predict Lung Cancer Risk in Routine Care." *Cancer Epidemiology Biomarkers and Prevention* 32(3):337–43. doi: 10.1158/1055-9965.EPI-22-0873.

Gandhi, Zainab, Priyatham Gurram, Birendra Amgai, Sai Prasanna Lekkala, Alifya Lokhandwala, Suvidha Manne, Adil Mohammed, Hiren Koshiya, Nakeya Dewaswala, Rupak Desai, Huzaifa Bhopalwala, Shyam Ganti, and Salim Surani. 2023. "Artificial Intelligence and Lung Cancer: Impact on Improving Patient Outcomes." *Cancers* 15(21):1–16. doi: 10.3390/cancers15215236.

Indra, Muhamad, Ilham Maulana, and Siti Ernawati. 2024. "Machine Learning for Stroke Prediction : Evaluating the Effectiveness of Data Balancing Approaches." 6(4).

Lee, Hsiu An, Louis R. Chao, and Chien Yeh Hsu. 2021. "A 10-Year Probability Deep Neural Network Prediction Model for Lung Cancer." *Cancers* 13(4):1–15. doi: 10.3390/cancers13040928.

Maulana, Ilham, Siti Ernawati, and Muhammad Indra. 2024. "IMPROVING IMAGE CLASSIFICATION ACCURACY WITH OVERSAMPLING AND DATA AUGMENTATION USING DEEP LEARNING : A CASE STUDY ON." 6(4).

Moozhippurath, Bineesh, and Jayapandian Natarajan. 2025. "Lung Cancer Prediction with Advanced Graph Neural Networks." *Indonesian Journal of Electrical Engineering and Computer Science* 37(2):1077–84. doi: 10.11591/ijeecs.v37.i2.pp1077-1084.

Nemlander, Elinor, Andreas Rosenblad, Eliya Abedi, Simon Ekman, Jan Hasselström, Lars E. Eriksson, and Axel C. Carlsson. 2022. "Lung Cancer Prediction Using Machine Learning on Data from a Symptom E-Questionnaire for Never Smokers, Formers Smokers and Current Smokers." *PLoS ONE* 17(10 October):1–11. doi: 10.1371/journal.pone.0276703.

Nguyen, Hung Viet, and Haewon Byeon. 2023. "Predicting Depression during the COVID-19 Pandemic Using Interpretable TabNet: A Case Study in South Korea." *Mathematics* 11(14). doi: 10.3390/math11143145.

Nguyen, Hung Viet, and Haewon Byeon. 2024. "A Hybrid Self-Supervised Model Predicting Life Satisfaction in South Korea." *Frontiers in Public Health* 12(October):1445864. doi: 10.3389/fpubh.2024.1445864.

Raptis, Sotiris, Christos Ilioudis, and Kiriaki Theodorou. 2024. "From Pixels to Prognosis: Unveiling Radiomics Models with SHAP and LIME for Enhanced Interpretability." *Biomedical Physics and Engineering Express* 10(3). doi: 10.1088/2057-1976/ad34db.

Smith, Richard J., Thurairajah Vijayaharan, Victoria Linehan, Zhuolu Sun, Jean Hai Ein Yong, Scott Harris, Hensley H. Mariathas, and Rick Bhatia. 2022. "Efficacy of Risk Prediction Models and Thresholds to Select Patients for Lung Cancer Screening." *Canadian Association of Radiologists Journal* 73(4):672–79. doi: 10.1177/08465371221089899.

Sun, Yilan, Guozhen Cheng, Dongliang Wei, Jiacheng Luo, and Jiannan Liu. 2024. "Integrating Omics Data and Machine Learning Techniques for Precision Detection of Oral Squamous Cell Carcinoma: Evaluating Single Biomarkers." *Frontiers in Immunology* 15(December):1–14. doi: 10.3389/fimmu.2024.1493377.

Sung, Hyuna, Jacques Ferlay, Rebecca L. Siegel, Mathieu Laversanne, Isabelle Soerjomataram, Ahmedin Jemal, and Freddie Bray. 2021. "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries." *CA: A Cancer Journal for Clinicians* 71(3):209–49. doi: 10.3322/caac.21660.

Tao, Guangyu, Li Zhu, Qunhui Chen, Lekang Yin, Yamin Li, Jiancheng Yang, Bingbing Ni, Zheng Zhang, Chi Wan Koo, Pradnya D. Patil, Yinan Chen, Hong Yu, Yi Xu, and Xiaodan Ye. 2022. "Prediction of Future Imagery of Lung Nodule as Growth Modeling with Follow-up Computed Tomography Scans Using Deep Learning: A Retrospective Cohort Study." *Translational Lung Cancer Research* 11(2):250–62. doi: 10.21037/tlcr-22-59.

Zamzam, Yra Fatria, Triando Hamonangan Saragih, Rudy Herteno, Muliadi, Dodon Turianto Nugrahadi, and Phuoc Hai Huynh. 2024. "Comparison of CatBoost and Random Forest Methods for Lung Cancer Classification Using Hyperparameter Tuning Bayesian Optimization-Based." *Journal of Electronics, Electromedical Engineering, and Medical Informatics* 6(2):125–36. doi: 10.35882/jeeemi.v6i2.382.

Zhang, Ruyang, Sipeng Shen, Yongyue Wei, Ying Zhu, Yi Li, Jiajin Chen, Jinxing Guan, Zoucheng Pan, Yuzhuo Wang, Meng Zhu, Junxing Xie, Xiangjun Xiao, Dakai Zhu, Yafang Li, Demetrios Albanes, Maria Teresa Landi, Neil E. Caporaso, Stephen Lam, Adonina Tardon, Chu Chen, Stig E. Bojesen, Mattias Johansson, Angela Risch, Heike Bickeböller, H. Erich Wichmann, Gadi Rennert, Susanne Arnold, Paul Brennan, James D. McKay, John K. Field, Sanjay S. Shete, Loic Le Marchand, Geoffrey Liu, Angeline S. Andrew, Lambertus A. Kiemeney, Shan Zienolddiny-Narui, Annelie Behndig, Mikael Johansson, Angela Cox, Philip Lazarus, Matthew B. Schabath, Melinda C. Aldrich, Juncheng Dai, Hongxia Ma, Yang Zhao, Zhibin Hu, Rayjean J. Hung, Christopher I. Amos, Hongbing Shen, Feng Chen, and David C. Christiani. 2022. "A Large-Scale Genome-Wide Gene-Gene Interaction Study of Lung Cancer Susceptibility in Europeans With a Trans-Ethnic Validation in Asians." *Journal of Thoracic Oncology* 17(8):974–90. doi: 10.1016/j.jtho.2022.04.011.