Journal of Applied Informatics and Computing (JAIC)
Vol.9, No.6, December 2025, pp. 3454~3463
e-ISSN: 2548-6861

Enhancing Interpretable Multiclass Lung Cancer Severity Classification using TabNet

Maria Bernadette Chayeenee Norman [1]*, Ika Novita Dewi [2]*, Abu Salam [3], Danang Wahyu Utomo [4]**, Sindhu Rakasiwi [5]**, Asih Rohmani [6]***

* Department of Information System, Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang
** Research Center for Intelligent Distributed Surveillance and Security (IDSS), Universitas Dian Nuswantoro, Semarang
112202206930@mhs.dinus.ac.id [1], ikadewi@dsn.dinus.ac.id [2], abu.salam@dsn.dinus.ac.id [3], danang.wu@dsn.dinus.ac.id [4], sindhu.rakasiwi@dsn.dinus.ac.id [5], aseharsoyo@dsn.dinus.ac.id [6]

Article Info

Article history:

Received 2025-10-02
Revised 2025-10-30
Accepted 2025-11-05

Keyword:

Cross Validation,
Deep Learning,
Lung Cancer,
TabNet,
Tabular Data.

ABSTRACT

Lung cancer poses a significant global mortality challenge, with early clinical detection hindered by non-specific symptoms making accurate diagnosis dependent on extracting subtle patterns from often complex medical tabular data. Traditional machine learning approaches often fall short in capturing intricate patterns within such heterogeneous datasets, hindering effective clinical decision support. This research introduces TabNet, an interpretable deep learning architecture, for multiclass lung cancer severity prediction (low, medium, high). Utilizing the Kaggle Lung Cancer dataset, our methodology leverages TabNet's unique attention-based feature selection for end-to-end processing of tabular data, enabling adaptive identification of key predictors and crucial model interpretability. To effectively assess its predictive capabilities and ensure robust performance, the model was trained with default configurations and validated through stratified 5-fold cross-validation, achieving outstanding performance on the test set: 98.50% accuracy, a 0.98 F1-score, and a 0.9996 macro-AUC-ROC. Beyond its robustness, confirmed by stable learning curves, interpretability analysis highlighted 'Genetic Risk' and 'Shortness of Breath' as dominant factors. Our results underscore TabNet's efficacy as a reliable, robust, and inherently interpretable solution, offering significant potential to improve the precision and transparency of lung cancer severity assessment in clinical practice.

I. INTRODUCTION

Lung cancer is one of the most prevalent diseases worldwide. It is the leading cause of cancer-related mortality, with an increasing incidence rate annually in both developed and developing countries [1]. Biologically, lung cancer is caused by genetic alterations in epithelial cells of the respiratory tract [2]. These alterations lead to abnormal and uncontrolled proliferation, which may originate directly from the lung tissue (primary) or from other organs (metastasis) [3]. In its early stages, lung cancer symptoms are generally nonspecific and are often misinterpreted as minor respiratory disorders, resulting in delayed diagnosis and treatment. Additionally, the low level of awareness among medical personnel regarding the need for further examination of suspicious symptoms is also a factor that can worsen the condition [4]. These issues underscore the urgency of more effective strategies for prevention and early detection [5]. Early detection plays a crucial role in reducing mortality rates by enabling more appropriate control measures and significantly improving lung cancer patients survival rates [6]. According to data from the Global Cancer Observatory (GLOBOCAN) in 2022, lung cancer accounts for 12.4% of all cancer cases worldwide, representing approximately 2.4 million cases, with a mortality rate is 18.7%, resulting in 1,817,469 deaths [7]. In Indonesia, lung cancer is the second most prevalent case, accounting for 38,904 or 9.5% of cases, with a mortality rate of 34,339 or 14.1% of all deaths from

lung cancer [8]. The high incidence rate is influenced by various risk factors, with smoking being the main cause in 85% of cases for both active and passive smokers [9]. Other factors such as air pollution, exposure to carcinogens in the workplace, alcohol consumption, obesity, and genetic predisposition also increase the risk of lung cancer.

Various research has applied machine learning algorithms to support the early detection of lung cancer. Deepa Yadav's research used supervised machine learning algorithms and achieved an accuracy rate of 87% using logistic regression [10]. Sinaga et al.'s research, which combined AdaBoost and Random Forest, resulted in an accuracy value of 95.4%, with a precision of 96%, and a recall of 96.3% [11]. Research conducted by Septhya et al., using the Decision Tree and Support Vector Machine with Forward Selection the accuracy was 62.3% [12]. Meanwhile, Marzuq et al.'s research, which used a Random Forest Decision Tree with 5-fold cross validation, obtained an accuracy of 88.9% [13]. Analysis of previous research shows that, despite significant progress, there are still limitations that indicating a research gap that needs to be addressed. Most previous research relied on conventional algorithms such as Random Forest and Decision Tree, which are less than optimal in handling highly complex medical data [4], [12]. Although other approaches using combination of algorithms are able to improve accuracy, but most are still limited to binary classification and have not explored more specific predictions of lung cancer severity or stage [11].

To overcome the limitations of existing lung cancer diagnostic methods, this research introduces the TabNet deep learning architecture, which is specifically designed for tabular data. Unlike conventional models, which treat features uniformly, TabNet uses an attention-based sparse feature selection mechanism that adapts to identify the most relevant predictors at each decision step. This mechanism enhances predictive performance and improves interpretability, which is critical in medical applications. Previous research have demonstrated of Chronic Kidney Disease (CKD), achieving accuracy rates above 94% in multiclass CKD stage classification [14]. Similarly, in fetal health analysis, TabNet achieved 94.36% accuracy, surpassing classical machine learning algorithms while maintaining interpretability in feature importance [15]. These findings validate TabNer's ability to handle complex medical tabular data and highlight its potential for broader clinical applications.

Building upon this foundation, the novelty of this research lies in applying TabNet to multiclass classification of lung cancer severity (low, medium, high). To further strengthen model robustness, Stratified K-fold Cross Validation is employes, ensuring balanced class proportions across folds and reducing evaluation bias [16]. Consequently, this research contributes not only improving the accuracy of early detection but also expanding the scope of diagnostics towards more accurate, interpretable, and clinically meaningful framework that support precision medicine and personalized treatment strategies for lung cancer patients.

II. RESEARCH METHODS

This research aims to develop and evaluate a deep learning-based prediction model using TabNet to classify the severity of lung cancer into three classes, namely low, medium, and high. The research procedure, as illustrated in Figure 1, consists of several stages: data preprocessing, data splitting, model training and validation, and final evaluation. In the training and validation phase, the TabNet model is combined with stratified 5-fold cross validation to enhance model generalization, minimize the risk of overfitting, and ensure a more robust performance evaluation. This methodological approach is expected to not only improve the accuracy and reliability of multi-level lung cancer severity prediction but also provide interpretability in feature selection, thereby contributing to a more comprehensive early detection framework.

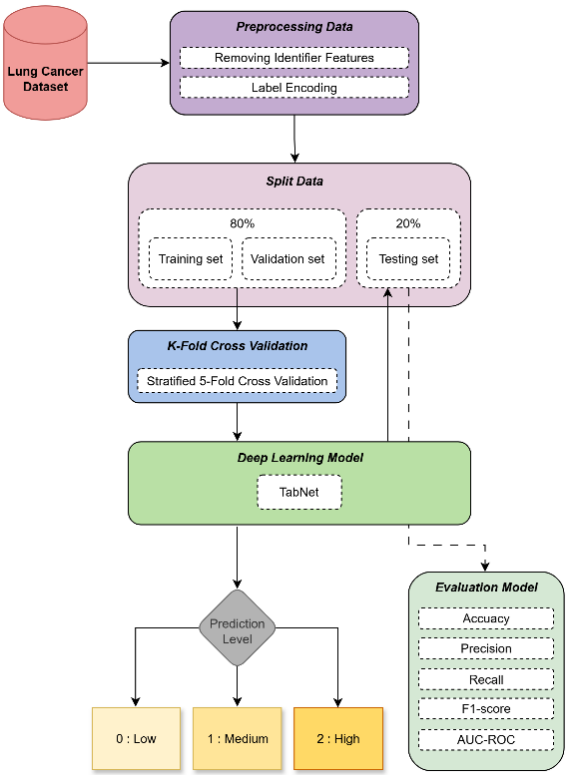


Figure 1. Research Flow Diagram

A. Dataset

The dataset used in this research was sourced from Kaggle via the following link: https://www.kaggle.com/datasets/thedevastator/cancer-patients-and-air-pollution-a-new-link/data. This dataset consists of 1000 data entries of lung cancer patients with their respective potential linkages. As illustrated in Figure 2, the distribution of the target variable "Level" is well-balanced, with the 'high', 'medium', and 'low' classes accounting for 36.5%, 33.2%, and 30.3% of the data, respectively. This balanced composition is ideal for classification tasks as it

minimizes the risk of model bias towards a majority class and enables a more robust and equitable learning process.

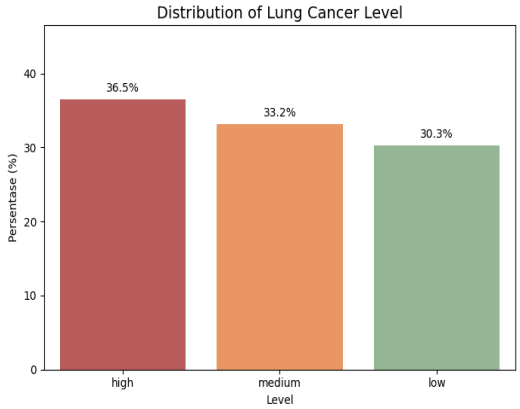


Figure 1. Distribution of Lung Cancer Level

The dataset structure has 25 predictor features with 1 target feature. The predictor features include various clinical and demographic attributes of patients relevant to the diagnosis of lung cancer. Meanwhile, the target feature is the categorical variable "Level", which categorizes the severity of lung cancer into three classes, namely low, medium, and high. The 25 predictor features encompass a range of patient data as detailed in Table 1. The use of this dataset aims to test the ability of the TabNet model to perform multiclass classification on complex medical tabular data by leveraging these diverse attributes.

TABEL I
PREDICTOR FEATURES IN THE LUNG CANCER DATASET

Category	Features Names
Identifiers & Demographics	Patient Id, index, Age, Gender
Lifestyle & Environment	Air Pollution, Alcohol use, Dust Allergy, Occupational Hazards, Genetic Risk, chronic Lung Disease, Balanced Diet, Obesity, Smoking, Passive Smoker
Clinical Symptoms	Chest Pain, Coughing of Blood, Fatigue, Weight Loss, Shortness of Breath, Wheezing, Swallowing Difficulty, Clubbing of finger nails, Frequent Cold, Dry Cough, Snoring

B. *Preprocessing Data*

The data preprocessing stage is a fundamental step in deep learning modeling to improve data quality, increase the performance and computational efficiency of predictive models [18]. In this research, there are a series of data preprocessing techniques that will be applied to the lung cancer dataset before it is used to train the TabNet model, consisting of:

1) *Identifier Feature Removal:* Identifier features are features in the dataset that are used to uniquely identify each data entry, such as Patient Id and index. The presence of such features can add complexity to the model without providing relevant or predictive information to the target variable [19]. As such, this rarity aims to prevent the model from overfitting and ensure that the model learns from features that are relevant to the target prediction.

2) *Label Encoding:* One of the transformation techniques used to convert the categorical value of a target feature into a numerical representation in order to be processed by models that can generally only accept numerical inputs [20]. In this research, the target variable used is Level, as a representation of the severity of lung cancer which consists of three categorical classes, namely low, medium, and high. The label encoding process is performed by converting the categorical labels in text form into numeric, namely 0 for the Low class, 1 for the medium class, and 2 for the High class, as listed in Table 2. This technique is necessary to enable TabNet modeling to perform multiclass classification of target variables with numeric inputs [21].

TABEL II
RESULT FROM LABEL ENCODING TARGET FEATURE

Class label before encoding	Class label after encoding
Low	0
Medium	1
High	2

C. *Split Data*

After preprocessing, the dataset was divided while maintaining a consistent data size and retaining 23 features. An 80:20 ratio was applied to the splitting process, resulting in three primary subsets: training, validation, and testing. Specifically, 64% of the data was allocated for training, 16% for validation, and 20% for testing. The training subset was used to fit the model and learn feature target relationship. The validation subset was used to monitor performance during training and prevent overfitting. The testing subset provided an objective measure of the final model's performance [22]. Table 3 presents the distribution of data after splitting.

TABEL III
RESULT FROM LABEL ENCODING TARGET FEATURE

Subset	Count
Training	640
Validation	160
Testing	200

D. *TabNet Model*

The TabNet model was chosen as the main approach in this research due to its superior ability to accurately handle tabular data end-to-end, as well as its good interpretability. TabNet is a deep learning architecture that processes tabular data by integrating an attention-based sparse feature selection mechanism. This mechanism allows the model to sequentially select the most relevant subset of features at each decision step [21], [23].

TabNet architecture consists of three main components: feature transformer, attentive transformer, and decision step. The feature transformer functions to process features into an

initial latent representation, which is then passed to the Attentive Transformer to determine relevant features that will be focused on in the next decision step. Equation (1) is the formula of TabNet's working mechanism.

$$a[i-1] : M[i] = sparsemax(P[i-1] \, . \, h_i([a-1])) \quad (1)$$

Where $a\,[\,i\,-\,1]$ is the process from the previous step, $P\,[\,i\,-\,1]$ is the scale prior of feature usage, and $h_i\,(\,[\,a\,-\,1\,]\,)$ is the result of the attentive transformer, which takes the representation from the previous step and transforms it. Thus, TabNet can provide good performance by performing sparse feature selection to support model interpretability [24].

This research implemented a classification model for lung cancer severity was implemented using the PyTorch-based TabNetClassifier with the default parameter configuration, without any modifications of the network structure. The model was trained from scratch, meaning all network parameters were randomly initialized without utilizing pre-trained weights. This approach was chosen to purely evaluate the baseline performance of the TabNet architecture on tabular medical data, without applying extensive hyperparameter tuning. However, some parameters in the training process (model.fit) were adjusted to accommodate data characteristics and improve training efficiency. Table 4 presents details of the default parameters used in the TabNet model architecture. Training parameters such as the maximum number of epochs, batch size, and the early stopping mechanism were configured according to experiment requirements. It should be emphasized that explicit parameter tuning was not performed in this research, so the obtained results reflect the model's performance with the default settings from the PyTorch-TabNet library.

TABEL III
PARAMETER SETTINGS OF TABNET MODEL

	Parameter	Value
Model Parameters	n_d	8
	n_a	8
	n_steps	3
	gamma	1.3
	optimizer_fn	Adam
	optimizer_params	lr=2e-2
	scheduler_params	step_size=50, gamma=0.9
	mask_type	'sparsemax'
Training Parameters (model.fit)	patience	10
	max_epoch	50
	batch_size	512
	virtual_batch_size	256
	num_workers	0
	eval_metric	'accuracy'

E. K-fold Cross Validation

In machine learning, model evaluation is often conducted through cross-validation method to measure how well a model can generalize to unseen data. In K-fold cross validation, the dataset is partitioned into k subsets or folds of equal size. In each iteration, one fold is used as validation data, while the remaining k-1 folds are used for training set [25]. Altough effective, this method has a drawback since the random partitioning of data may fail to preserve the original distribution of target classes. To address this issue, Stratified K-fold Cross Validation is employed, ensuring that each fold maintains the sam class proportion as the entire dataset, thereby providing a more reliable representation of the overall distribution [26].

In this research, stratified 5-fold Cross Validation is applied which involves dividing the training data into five balanced folds based on the target class distribution. At each iteration, the TabNet model is trained on four folds and validated on the fifth fold, until the combination is completed. Thus, stratified 5-fold CV helps to thoroughly evaluate the model's performance on a proportional representation of classes, as well as prevent potential overfitting due to class imbalance during training.

F. Model Evaluation

In this research, the performance of the TabNet model in predicting the severity of lung cancer was assessed using evaluation metrics including accuracy, precision, recall, F1-score and AUC-ROC [26], [27].

Accuracy reflects the ratio of correctly classififed samples to the total predictions. The formula for accuracy is presented in Equation (2).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

Precision indicates the proportion of true positive predictions among all instances predictes as positive as shown in Equation (3). Recall measures the model's ability to correctly detect positive cases, as shown in Equation (4).

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

The F1-score, presented in Equation (5), represents the harmonic mean of precision and recall, providing a balanced evaluation of the two metrics.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

To complement these metrics, the Receiver Operating Characteristic (ROC) was also used, which illustrates the trade-off between the True Positive Rate (TPR) and the False Positive Rate (FPR), as shown in Equation (6).

$$TPR = \frac{TP}{TP + FN}, FPR = \frac{FP}{FP + TN} \quad (6)$$

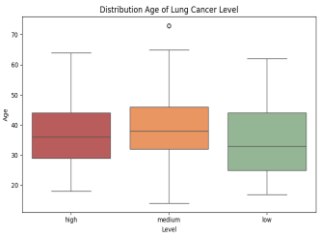
For multiclass classification, the One-vs-Rest (OvR) scheme was applied, where each class is alternately treated as the positive category while the remaining classes are considered as negative. This approach enables a more comprehensive evaluation of the model's performance in predicting lung cancer severity.

III. Results and Discussion

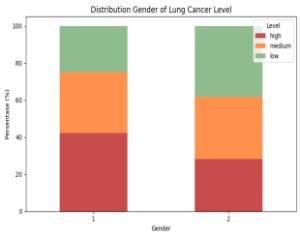
A. Exploratory Data Analysis

An Exploratory Data Analysis (EDA) visualized in Figure 3, was conducted as a preliminary step to understand the dataset's fundamental characteristics and identify initial patterns. Analysis of demographic features, such as age (Figure 3(a)) and gender (Figure 3(b)), indicates relatively uniform distributions across the severity levels, suggesting these variables may have limited predictive power when considered independently. In contrast, more significant insights emerged from examining the relationship between clinical risk factors and symptoms with the target variable.

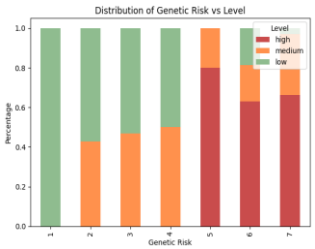
Features such as 'Genetic Risk' (Figure 3(c)), 'Shortness of Breath' (Figure 3(d)), 'Fatigue' (Figure 3(e)) demonstrated a strong and progressive visual correlation with the severity levels. As the values of these features increase, there is a notable and consistent rise in the proportion of patients classified under the medium and high severity categories, highlighting their strong discriminative capacity. Collectively, this EDA validates the dataset's suitability for the severity prediction task and establishes a solid framework for interpreting the feature importance results that will be subsequently derived from the trained model.



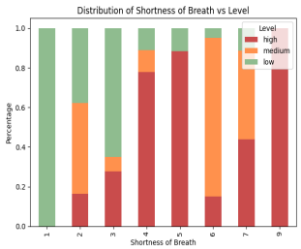
(a) Distribution Age of Lung Cancer Level



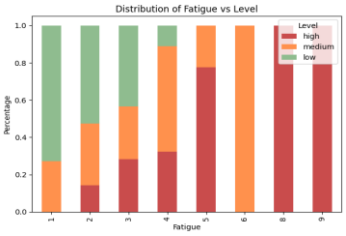
(b) Distribution Gender of Lung Cancer Level



(c) Distribution of Genetic Risk vs Level



(d) Distribution of Shortness of Breath vs Level



(e) Distribution of Fatigue vs Level

Figure 3. Exploratory Data Analysis

The correlation analysis, visualized in Figure 4, was conducted to evaluate the linear relationship between the variables and support the modeling strategy. The analysis revealed several features that had strong positive correlations with the target variable 'Level', including Coughing of Blood (0.78), Alcohol Use (0.72), and Genetic Risk (0.70). These strong correlations provide initial validation that learnable patterns exist in the data, thus reinforcing the model's potential success in classifying lung cancer severity. The correlation matrix also revealed high correlations between certain predictor features, such as 'OccuPational Hazards' and 'Genetic Risk' (0.89).

However, the matrix also shows a wide variation in the degree of correlation between features, with many features showing weak or moderate correlations. This indicates that not all features have an equal predictive contribution. This insight further justifies selecting the TabNet model for this research. With the attention mechanism, TabNet can perform automatic feature selection during training. This capability allows the model to dynamically assign greater importance to the most informative features while reducing the influence of less relevant ones. Therefore, manual feature selection is unnecessary to handle the complexity and variation in feature relevance present in medical tabular data.

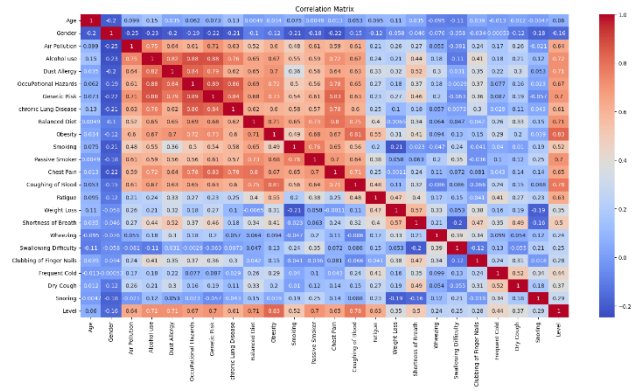


Figure 4. Correlation Matrix

B. *Model Evaluation on Lung Cancer*

Evaluation of the stability and consistency of the TabNet model performance was conducted by analyzing the accuracy of the validation set for each of the five folds in the stratified 5-fold cross validation scheme. Figure 5 presents a visualization of the inter-fold validation accuracy comparison.

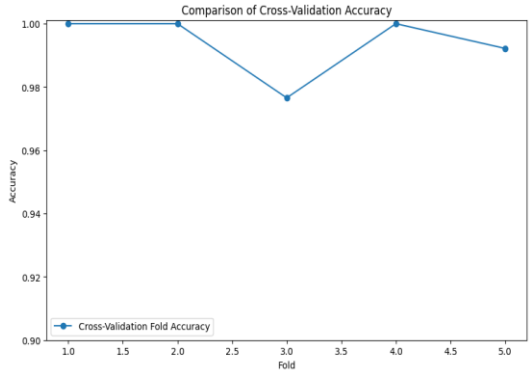


Figure 5. Comparison of The Accuracy Validation of each Fold Cross-Validation

As illustrated in Figure 5, the TabNet model shows a very high level of accuracy and is relatively consistent in most of the validation folds. Specifically, on the 1st and 2nd folds, the model achieved a perfect accuracy of 100%. The performance slightly decreased on the 3rd fold, with 97.66% accuracy. However, it was again optimized in the 4th fold with 100% accuracy, followed by the 5th fold with 99.22% accuracy. The average cross validation accuracy of these five folds is 99.38% with a standard deviation of 0.0097, showing the stability of the model's performance and generalization before the final evaluation on the testing data.

The performance of the TabNet model was comprehensively evaluated using a testing set of 200 independent entries. The evaluation was conducted using standard evaluation metrics for multiclass classification, which are listed in Table 5. They include accuracy, precision, recall, F1-score, and AUC.

TABEL V
TABNET MODEL EVALUATION RESULT

	Precision	Recall	F1-score	Support
0	1.00	0.97	0.98	61
1	0.96	1.00	0.98	66
2	1.00	0.99	0.98	73
Accuracy			0.98	200
Macro avg	0.99	0.98	0.98	200
Weighted avg	0.99	0.98	0.99	200
AUC-ROC : 0.9996				
Accuracy : 0.9850				

Based on Table 5, the TabNet model performed excellently on the testing set, achieving an accuracy of 98.50%. This result was obtained from the final evaluation of the test data, rather than from the average of the cross-validation folds. The macro average scores for precision, recall and the F1-score were 0.99, 0.98, and 0.98, respectively. Similar results were observed in the weighted averages of these metrics, confirming consistent performance of the model across all three lung cancer severity classes.

C. *Analysis of the Dynamic Training Curve and Validation of the Model's*

The average training loss curve, as illustrated in Figure 6. shows a significant and consistent decrease as the number of training epochs increases. Starting from an initial value of 1.11, the training loss reaches a minimum value of 0.01 at the 28th epoch, indicating that the TabNet model has learned the data representation effectively.

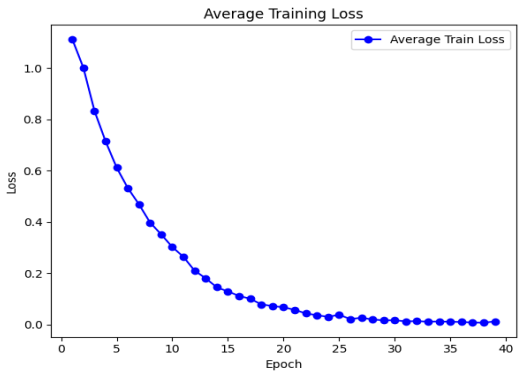


Figure 6. Average Training Loss Curve

Furthermore, Figure 7 shows a comparison between the training average accuracy curve and the validation average accuracy curve. Both curves show substantial performance improvement. The training accuracy increases progressively, reaching a value above 0.95 towards the end of the training period after the 34th epoch. The validation accuracy curve also shows a similar upward trend, reaching a value of 0.96 at the same epoch.

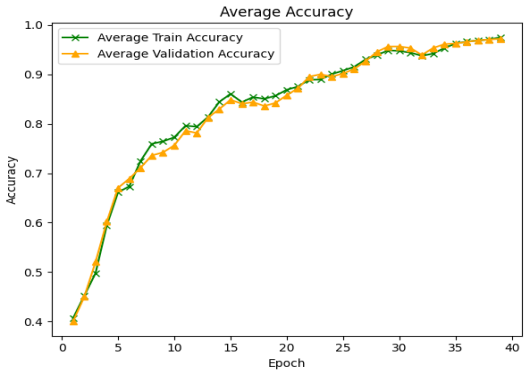


Figure 7. Comparison of Average Accuracy Training and Validation Curve

The close alignment and absence of divergence between the traning and validation accuracy curves indicates that the model generalizes well to unseen data. Futhermore, the gradual reduction in training loss without a sudden increase in validation loss suggests that the model is neither overfitting nor underfitting. These results imply that the TabNet model achieves a balanced learning state, maintaining high accuracy and stable generalization throughout the training process.

D. Evaluation of the Model's with AUC-ROC Analysis

The discriminative ability of the TabNet model in distinguishing between lung cancer severity classes was further evaluated through AUC-ROC values. The ROC curve visualizes the trade-off between True Positive Rate (TPR) or sensitivity and False Positive Rate (FPR) at various classification thresholds. Figure 8. presents the macro averaged ROC curve for the TabNet model on the testing set.

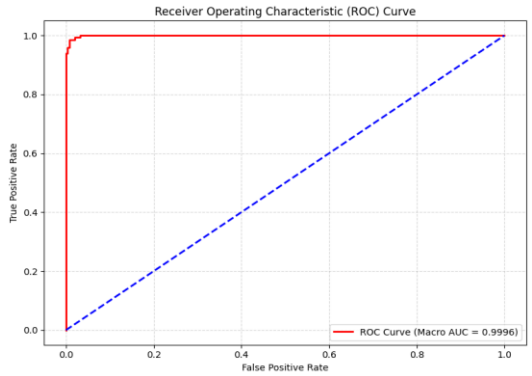


Figure 8. Average of Macro Receiver Operating Characteristic (ROC) Curve

As illustrated in Figure 8, the macro averaged ROC curve shows a very close to perfect performance, with an AUC value recorded at 0.9996. This confirms that the TabNet model has an excellent capacity to accurately discriminate between lung cancer severities. This strong AUC performance is in line with the results of other evaluation metrics, further strengthening the potential of the TabNet model as a reliable and effective clinical decision support tool.

E. Confusion Matric

The visualization of TabNet modeling in classifying each sample in the testing set is presented using the confusion matrix in Figure 9. This matrix compares the actual lung cancer severity class on the vertical axis (actual) with the class predicted by the model on the horizontal axis (predicted) for all three classes.

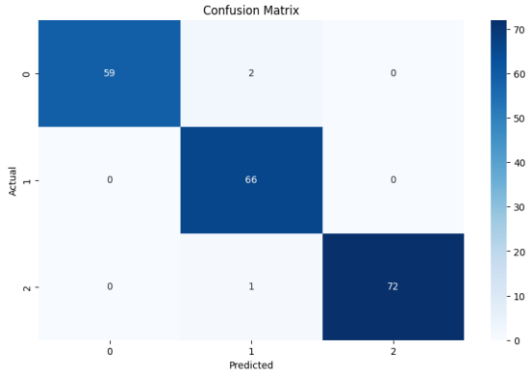


Figure 9. Confusion Matrix

Based on Figure 9, the main diagonal of the matrix shows the dominant number of correct predictions for each class, i.e. for class 0 (actual 'low') out of a total of 61 samples, 59 were correctly classified, for class 1 (actual 'medium') it was perfectly classified, where 66 'medium' samples were correctly identified and for class 2 (actual 'high') out of 73 samples, 72 were correctly classified. The majority of samples (197 out of 200) showed that misclassification tended to occur between ordinal adjacent classes, between 'low' and 'medium', and between 'high' and 'medium'. Overall, the confusion matrix shows that the TabNet model has a very high accuracy in classifying the severity of lung cancer in the dataset used.

F. Error Analysis

To gain a deeper understanding of the limitations and specific error patterns of the TabNet model, an error analysis was performed based on the confusion matrix generated from the evaluation on the testing set. Out of a total of 200 samples in the test data, the TabNet model made three misclassifications, equivalent to an overall error rate of 1.5%. Further analysis of the off-diagonal elements in the confusion matrix are listed in Table 6, revealed the following error patterns:

1) Misclassification of Class 'Low' to 'Medium': There were 2 instances where samples actually belonging to the 'Low' severity category (Class 0) were incorrectly predicted by the model as 'Medium' (Class 1). These errors indicate a tendency for the model to slightly overestimate the severity of few 'Low' cases or misinterpret features similarities between some 'Low' cases and the characteristics of 'Medium' cases, thereby confusing the model.

2) Misclassification of 'High' Class to 'Medium': There was 1 instance where a sample actually belonging to the 'High' severity category (Class 2) was misclassified by

the model as 'Medium' (Class 1). This error suggests a potential for the model to underestimate severity in certain 'High' cases, or the overlap of features between that 'High' case and the 'Medium' case.

TABEL VI
MISCLASSIFICATION OF ERROR ANALYSIS

		Predicted		
		Low (Class 0)	**Medium (Class 1)**	**High (Class 2)**
Actual	**Low (Class 0)**	59	2	0
	Medium (Class 1)	0	66	0
	High (Class 2)	0	1	72

The concentration of misclassifications in the 'Medium' class suggests that the decision boundaries between ordinally adjacent severity levels (Low-Medium and Medium-High) have overlapping feature characteristics compared to the distinction between "Low" and "High" classes. Nevertheless, the very low total number of errors generally reinforces the robustness and high accuracy of the TabNet models in classifying lung cancer severity on the tested dataset.

G. Feature Importance

A significant advantage of TabNet modeling is its ability to provide insight into the feature importance that influences the model's decision-making process. TabNet's attention mechanism enables the extraction of feature importance scores, which indicate the relative contribution of each predictor feature to the lung cancer severity classification result. As shown in Figure 10, the top ten importance features include Genetic Risk, Shortness of Breath, Dust Allergy, Wheezing, Chest Pain, Fatigue, Air Pollution, Smoking, Obesity, and Coughing of Blood, with Genetic Risk and Shortness of Breath stand out as the most dominant predictors. These findings align with the medical literature, which emphasizes that genetic predisposition significantly increases tsusceptibility to lung cancer through inherited mutations affecting tumor suppression and DNA repair pathways. Shortness of breath is also a major clinical indicator of disease progression due to reduced pulmonary capacity and airway obstruction. These results highlight the combined impact of genetic, clinical, and lifestyle-related factors. They also demonstrate TabNet's interpretability in supporting medical experts in better understanding the relative contribution of each feature to lung cancer severity classification.

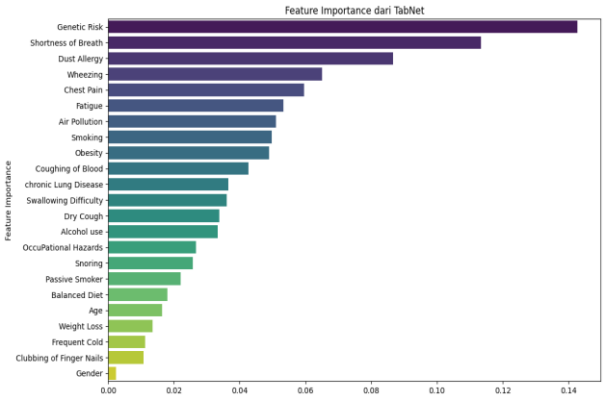


Figure 10. Feature Importance of TabNet Model

The identification of feature importance not only provides interpretability to the model, but also impacts the potential for wider clinical application. These features can be used as a focus of attention in the process of early diagnosis and assessment of the severity of lung cancer patients, as well as a basis for consideration in the development of data-based medical decision support systems. The test results of this research demonstrate the successful development and evaluation of the TabNet deep learning model, which can accurately classify lung cancer severity. The model achieved outstanding performance on the testing set: 98.50% accuracy, a macro F1 score of 0.98, and a macro-averaged AUC-ROC of 0.9996. These results were obtained using the model's default settings and without hyperparameter tuning. These metrics underscore TabNet's robust predictive and discriminative capabilities.

The model's stability was further confirmed through stratified 5-fold cross-validation, which showed consistent validation accuracy averaging 99.38% with a very low standard deviation. The model also demonstrated good generalization and exhibited no signs of overfitting, as evidenced by stable convergence in the loss and accuracy curves. Confusion matrix analysis revealed that most misclassifications occurred between adjacent severity levels (e.g., low vs. medium, medium vs. high), as would be expected given the overlap in clinical symptoms. Nevertheless, only three misclassifications were recorded out of 200 samples, which reinforces the model's reliability in clinical classification tasks.

H. Comparison with Previous Researches

The TabNet model proposed in this research outperforms baseline models by a significant margin. Previous research have evaluated lung cancer severity classification using conventional machine learning algorithms and ensemble learning methods such as logostic regrssion, decision tree combined with support vector machine (SVM), random forest, and adaboost-random forest ensemble. As shown in Table 7, the TabNet model achieved an accuracy of 98.50%, which is significantly higher than the performance of Logistic Regression (87%), Decision Tree + SVM (63.2%), and

Random Forest (88.9%). Although the ensemble approach proposed by Roy et al. using Adaboost and Random Forest achieved relatively high accuracy (95.40%), the TabNet model developed in this research still outperformed it. These results higlights the advantege of TabNet's deep learning architecture's in capturing complex nonlinear interactions within tabular medical data, resulting in a more accurate and reliable classification of lung cancer severity.

TABEL VII
COMPARISON WITH PREVIOUS RESEARCH

Author	Model Algoritma	Result
Deepa Yadav	Logistic Regression	87%
Roy et al.	Adaboost + Random Forest	95.40%
Septhya et al.	Decision Tree + SVM	63.2%
Marzuq et al.	Random Forest Decision Tree	88.9%
Ours	**TabNet**	**98.50%**

IV. CONCLUSION

This research successfully demonstrated the effectiveness of the TabNet deep learning model for multiclass classification of lung cancer severity. Utilizing default parameter configuration and a stratified 5-fold cross validation scheme, the model showed high performance stability. The average cross validation accuracy reached 99.38% (+- 0.0097), with no indication of significant overfitting. Comprehensive evaluation on the testing set yielded excellent classification performance, with an accuracy of 98.50%, a macro average F1-score of 0.98, and a macro average AUC-ROC value of 0.9996. These results confirm the potential of TabNet as a reliable approach for predictions on medical tabular data.

Error analysis of the 3 misclassifications (1.5% error rate) revealed that errors occurred towards the ordinally adjacent 'Medium' class, with no extreme misclassifications between the 'Low' and 'High' classes, and perfect identification of the 'Medium' class. This indicates the model's strong discriminative ability, although it suggests a potential for overlapping or borderline features between adjacent severity levels. Future development should focus on extensive TabNet hyperparameters optimization to maximize the model's potential performance and direct clinical integration by medical experts to assess the relevance and potential implementation in real world diagnostic practice. Such efforts will further solidify its considerable potential for clinical decision support systems, particularly in enabling more precise, data driven early detection and risk management of lung cancer.

REFERENCES

[1] M. A. Rahman Wahid, A. Nugroho, and A. Halim Anshor, "Prediksi Penyakit Kanker Paru-Paru Dengan Algoritma Regresi Linier," Bull. Inf. Technol. BIT, vol. 4, no. 1, pp. 63–74, Mar. 2023, doi: 10.47065/bit.v4i1.501.

[2] Z. G. Syafira, C. A. Sari, I. U. W. Mulyono, F. Agustina, S. Suprayogi, and M. Doheir, "Regionprops Segmentation in Convolutional Neural Network for Identification of Lung Cancer Disease and Position," J. Masy. Inform., vol. 16, no. 2, pp. 199–215, Jul. 2025, doi: 10.14710/jmasif.16.2.73967.

[3] I. Buana and D. A. Harahap, "Asbestos, Radon dan Polusi Udara sebagai Faktor Resiko Kanker Paru pada Perempuan Bukan Perokok," AVERROUS J. Kedokt. Dan Kesehat. Malikussaleh, vol. 8, no. 1, pp. 1–16, Jul. 2022, doi: 10.29103/averrous.v8i1.7088.

[4] D. Benaya, "Implementasi Random Forest dalam Klasifikasi Kanker Paru-Paru," JOINTER J. Inform. Eng., vol. 5, no. 01, pp. 27–31, Jun. 2024, doi: 10.53682/jointer.v5i01.331.

[5] World Health Organization, "Cancer," 2025. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cancer

[6] R. Patra, "Prediction of Lung Cancer Using Machine Learning Classifier," in Computing Science, Communication and Security, vol. 1235, N. Chaubey, S. Parikh, and K. Amin, Eds., in Communications in Computer and Information Science, vol. 1235. , Singapore: Springer Singapore, 2020, pp. 132–142. doi: 10.1007/978-981-15-6648-6_11.

[7] Ferlay, M. Colombet, I. Soerjomataram, D. M. Parkin, M. Piñeros, A. Znaor, and F. Bray, "900 World fact sheet," International Agency for Research on Cancer, Feb. 8, 2024. [Online]. Available: https://gco.iarc.who.int/media/globocan/factsheets/populations/900-world-fact-sheet.pdf

[8] Ferlay, M. Colombet, I. Soerjomataram, D. M. Parkin, M. Piñeros, A. Znaor, and F. Bray, "360 Indonesia fact sheet," International Agency for Research on Cancer, Feb. 8, 2024. [Online]. Available: https://gco.iarc.who.int/media/globocan/factsheets/populations/360-indonesia-fact-sheet.pdf

[9] World Health Organization, "Lung Cancer," 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/lung-cancer

[10] D. Yadav, "Lung Cancer Prediction using Supervised ML Algorithms," Int. Res. J. Mod. Eng. Technol. Sci., Oct. 2022, doi: 10.56726/IRJMETS30472.

[11] R. B. Sinaga, D. Widiyanto, and B. T. Wahyono, "Deteksi Dini Penyakit Kanker Paru dengan Gabungan Algoritma Adaboost dan Random Forest," SENAMIKA (Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya), vol. 3, no. 2, pp. 855-864, 2022 [Online]. Available: https://ejournal.undip.ac.id/index.php/senamika

[12] D. Septhya et al., "Implementasi Algoritma Decision Tree dan Support Vector Machine untuk Klasifikasi Penyakit Kanker Paru: Implementation of Decision Tree Algorithm and Support Vector Machine for Lung Cancer Classification," MALCOM Indones. J. Mach. Learn. Comput. Sci., vol. 3, no. 1, pp. 15–19, May 2023, doi: 10.57152/malcom.v3i1.591.

[13] R. D. Marzuq, S. A. Wicaksono, and N. Y. Setiawan, "Prediksi Kanker Paru-Paru menggunakan Algoritme Random Forest Decision Tree," Jurnala Pengembangan Teknolologi Informasi Ilmu Komputer, vol. 7, no. 7, pp.3448-3456 [Online]. Available: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/12964

[14] M. N. H. Chowdhury et al., "Deep learning for early detection of chronic kidney disease stages in diabetes patients: A TabNet approach," Artif. Intell. Med., vol. 166, p. 103153, Aug. 2025, doi: 10.1016/j.artmed.2025.103153.

[15] B. Regmi and C. Shah, "Classification Methods Based on Machine Learning for the Analysis of Fetal Health Data," Nov. 18, 2023, arXiv: arXiv:2311.10962. doi: 10.48550/arXiv.2311.10962.

[16] A. U. Mazlan et al., "A Review on Recent Progress in Machine Learning and Deep Learning Methods for Cancer Classification on Gene Expression Data," Processes, vol. 9, no. 8, p. 1466, Aug. 2021, doi: 10.3390/pr9081466.

[17] A. Oktaviana, D. P. Wijaya, A. Pramuntadi, and D. Heksaputra, "Prediksi Penyakit Diabetes Melitus Tipe 2 Menggunakan Algoritma K-Nearest Neighbor (K-NN): Prediction of Type 2 Diabetes Mellitus Using The K-Nearest Neighbor (K-NN) Algorithm," MALCOM Indones. J. Mach. Learn. Comput. Sci., vol. 4, no. 3, pp. 812–818, May 2024, doi: 10.57152/malcom.v4i3.1268.

[18] F. Akbar, H. W. Saputra, A. K. Maulaya, M. F. Hidayat, and R. Rahmaddeni, "Implementasi Algoritma Decision Tree C4.5 dan Support Vector Regression untuk Prediksi Penyakit Stroke: Implementation of Decision Tree Algorithm C4.5 and Support Vector

Regression for Stroke Disease Prediction," MALCOM Indones. J. Mach. Learn. Comput. Sci., vol. 2, no. 2, pp. 61–67, Sep. 2022, doi: 10.57152/malcom.v2i2.426.

[19] T. Guo, S. Guo, J. Zhang, W. Xu, and J. Wang, "Efficient Attribute Unlearning: Towards Selective Removal of Input Attributes from Feature Representations," Apr. 16, 2022, arXiv: arXiv:2202.13295. doi: 10.48550/arXiv.2202.13295.

[20] S. Borrohou, R. Fissoune, and H. Badir, "Critical Role of Data Transformation in Preprocessing: Methods, Algorithms, and Challenges," in Model and Data Engineering, vol. 15590, C. Ordonez, G. Sperlì, E. Masciari, and L. Bellatreche, Eds., in Lecture Notes in Computer Science, vol. 15590. , Cham: Springer Nature Switzerland, 2025, pp. 108–122. doi: 10.1007/978-3-031-87719-3_9.

[21] S. Ö. Arik and T. Pfister, "TabNet: Attentive Interpretable Tabular Learning," Proc. AAAI Conf. Artif. Intell., vol. 35, no. 8, pp. 6679–6687, May 2021, doi: 10.1609/aaai.v35i8.16826.

[22] S. Aftab and S. Akhtar, "Diabetic Retinopathy Severity Classification Using Data Fusion and Ensemble Transfer Learning," J. Softw. Eng. Appl., vol. 18, no. 01, pp. 1–23, 2025, doi: 10.4236/jsea.2025.181001.

[23] E. Ismanto, A. Fadlil, and A. Yudhana, "Analisis Perbandingan Model Fully Connected Neural Networks (FCNN) dan TabNet Untuk Klasifikasi Perawatan Pasien Pada Data Tabular," vol. 5, no. 3, pp. 526-532, 2024, doi: 10.37859/coscitech.v5i3.8256.

[24] K. McDonnell, F. Murphy, B. Sheehan, L. Masello, and G. Castignani, "Deep learning in insurance: Accuracy and model interpretability using TabNet," Expert Syst. Appl., vol. 217, p. 119543, May 2023, doi: 10.1016/j.eswa.2023.119543.

[25] D. Berrar, "Cross-Validation," in Encyclopedia of Bioinformatics and Computational Biology, Elsevier, 2021, pp. 542–545. doi: 10.1016/B978-0-12-809633-8.20349-X.

[26] S. Prusty, S. Patnaik, and S. K. Dash, "SKCV: Stratified K-fold cross-validation on ML classifiers for predicting cervical cancer," Front. Nanotechnol., vol. 4, p. 972421, Aug. 2022, doi: 10.3389/fnano.2022.972421.

[27] Ž. Đ. Vujovic, "Classification Model Evaluation Metrics," Int. J. Adv. Comput. Sci. Appl., vol. 12, no. 6, 2021, doi: 10.14569/IJACSA.2021.0120670