IOSR Journal of Mobile Computing & Application (IOSR-JMCA)
e- ISSN: 2394-0050, P-ISSN: 2394-0042. Volume 13, Issue 3 (May. – June. 2026), PP 15-30
www.iosrjournals.org

OptNet: Optimized Tabnet For Lung Cancer Prediction

A. Karunamurthy,
Associate Professor, Dept Of CSE, Sri Manakula Vinayagar Engineering College (Autonomous), Puducherry-India

R. Gayathiri,
PG Student, Dept Of MCA, Sri Manakula Vinayagar Engineering College (Autonomous), Puducherry-India

Abstract
Lung cancer remains a leading cause of cancer-related mortality worldwide, and accurate predictive models from structured clinical data are essential for early diagnosis and treatment planning. We propose OptNet, a novel framework that integrates an attention-based TabNet classifier with automated hyperparameter optimization for high-performance lung cancer prediction. The core of our method is a sequential attention mechanism that performs instance-wise feature selection through sparse masking, thereby focusing computational resources on the most clinically relevant attributes—such as tumor size or biomarker levels—while suppressing noise from irrelevant variables. This design departs from conventional deep learning models that process all features uniformly. To further enhance predictive accuracy, we embed the Optuna optimization engine, which employs a Tree-structured Parzen Estimator algorithm to dynamically search the hyperparameter space. The objective function is the negative validation accuracy, and the optimizer iteratively samples new configurations—including the number of decision steps, feature transformer width, momentum, and sparsity coefficient—based on posterior distributions from previous trials. Each sampled configuration is immediately substituted into the TabNet architecture before training, creating a closed-loop refinement process that converges to a globally optimized model. The final output is a malignancy probability score for each patient. We evaluate OptNet on a structured lung cancer dataset comprising demographic, symptomatic, and imaging-derived features. Our results demonstrate that the integration of sparse attention with Bayesian hyperparameter search yields superior classification performance compared to standard baselines, while maintaining interpretability through the learned feature masks. The framework thus offers a robust and computationally efficient solution for clinical decision support in oncology.

Keywords
Lung Cancer, Deep Learning, TabNet, Optuna, Hyperparameter Optimization, Medical Prediction, Clinical Decision Support System.

Date of Submission: 13-05-2026 Date of Acceptance: 23-05-2026

I. Introduction

Lung cancer is one of the most prevalent and lethal malignancies globally, accounting for approximately 1.8 million deaths annually. Early and accurate diagnosis is critical for improving patient survival rates, yet the interpretation of clinical and radiological data remains a complex task for healthcare providers. Clinical Decision Support Systems (CDSS) have emerged as essential tools to assist physicians in making evidence-based decisions by analyzing structured patient data [1]. These systems have been shown to improve diagnostic accuracy and reduce variability in clinical practice [2]. However, the effectiveness of a CDSS is fundamentally dependent on the predictive power of its underlying machine learning model.

Historically, lung cancer prediction models have relied on conventional algorithms such as Logistic Regression, Support Vector Machines, and Random Forests [3]. While these methods are interpretable and computationally efficient, they often require extensive manual feature engineering to achieve competitive performance. The process of selecting and transforming relevant clinical attributes—such as tumor size, patient age, smoking history, and biomarker levels—is labor-intensive and domain-specific. Moreover, these models struggle to capture complex, non-linear interactions between features, which are common in medical data. Deep learning approaches, including Convolutional Neural Networks and Recurrent Neural Networks, have been applied to medical imaging and sequential patient data with considerable success [4]. Nevertheless, these architectures are often ill-suited for structured tabular data, where feature sparsity and heterogeneous data types pose significant challenges. They also suffer from high computational costs and a lack of interpretability, which is a critical requirement in clinical settings.

The introduction of attention mechanisms, originally popularized in natural language processing by the Transformer architecture [5], provided a new paradigm for focusing model capacity on the most informative parts of the input. This concept was later adapted for tabular data, leading to the development of TabNet [6]. TabNet employs a sequential attention mechanism that performs instance-wise feature selection through sparse masking. This allows the model to dynamically identify and weigh the most relevant features for each individual patient, thereby eliminating the need for manual feature engineering. The architecture also provides interpretability by visualizing the learned feature masks, which can be directly mapped to clinical attributes. This is a significant advantage over black-box deep learning models, as it enables clinicians to understand the rationale behind each prediction.

Parallel to these architectural advances, the field of Automated Machine Learning (AutoML) has evolved to reduce human intervention in model development. Early hyperparameter optimization techniques, such as Grid Search and Random Search [7], provided a baseline for automated tuning but are inefficient for high-dimensional search spaces. More sophisticated Bayesian optimization frameworks, such as those based on Gaussian Processes and Tree-structured Parzen Estimators (TPE), have since become standard for efficiently navigating complex parameter landscapes [8]. Building upon these foundations, Optuna was developed as a flexible and efficient hyperparameter optimization framework that utilizes a define-by-run API and pruning strategies to accelerate the convergence of machine learning pipelines [9].

In this paper, we propose OptNet, a novel framework that synergistically integrates the TabNet architecture with the Optuna optimization engine for high-performance lung cancer prediction. The core novelty of our approach lies in the coupling of these two components: the sequential attention mechanism of TabNet autonomously extracts and weighs significant features from structured medical data, while Optuna dynamically optimizes the model's convergence parameters. This combination eliminates the reliance on manual feature engineering and human intervention in the parameter selection process. By embedding an iterative search strategy directly into the training pipeline, the method ensures that the classifier operates at peak efficiency, adapting its internal decision boundaries to maximize predictive reliability. The key contributions of this work are threefold. First, we demonstrate that the integration of attention-based feature selection with automated hyperparameter tuning yields superior classification performance compared to standard baselines. Second, we provide a comprehensive analysis of the learned feature masks, offering insights into the clinical attributes that are most predictive of lung cancer. Third, we show that the proposed framework is computationally efficient and scalable, making it suitable for deployment in real-world clinical decision support systems.

The remainder of this paper is organized as follows. Section 2 reviews related work on clinical decision support systems, deep learning for tabular data, and hyperparameter optimization. Section 3 presents the preliminaries, including the TabNet architecture and the Optuna optimization framework. Section 4 details the proposed OptNet framework, describing the integration of attention-based prediction with automated hyperparameter search. Section 5 outlines the experimental setup, including the dataset, preprocessing steps, and evaluation metrics. Section 6 presents the results and analysis, comparing OptNet against several baseline models. Section 7 discusses the implications of our findings, limitations of the current study, and directions for future work. Finally, Section 8 concludes the paper.

II. Related Work

The development of predictive models for lung cancer diagnosis has been a central focus of medical informatics research, with numerous studies exploring both traditional machine learning and deep learning approaches. Early work in this domain predominantly employed classical algorithms such as Logistic Regression, Support Vector Machines, and Random Forests, which were applied to structured clinical datasets containing demographic information, symptom profiles, and laboratory results [3]. These models offered interpretability and computational efficiency, but their performance was often constrained by the need for manual feature engineering and their limited capacity to capture complex, non-linear interactions among clinical variables. For instance, a study by Patil et al. demonstrated that a Random Forest classifier could achieve reasonable accuracy on a lung cancer dataset, but the model's reliance on hand-crafted features limited its generalizability across different patient populations [10].

The advent of deep learning brought new opportunities for medical prediction, particularly in the domain of medical imaging. Convolutional Neural Networks (CNNs) have been extensively applied to chest X-rays and CT scans for lung nodule detection and classification, achieving state-of-the-art performance in many benchmarks [4]. However, these image-based models are not directly applicable to structured tabular data, which remains the primary format for clinical records, electronic health records, and epidemiological surveys. Recurrent Neural Networks (RNNs) and their variants, such as Long Short-Term Memory (LSTM) networks, have been used to model sequential patient data, but they are less effective for static, high-dimensional feature vectors [11]. The fundamental challenge lies in the nature of tabular data: features are often heterogeneous, sparse, and contain both categorical and continuous variables, making it difficult for standard deep learning architectures to learn meaningful representations without extensive preprocessing.

To address these limitations, attention mechanisms were introduced as a means to dynamically focus model capacity on the most informative parts of the input. The Transformer architecture, originally proposed for natural language processing, demonstrated the power of self-attention in capturing long-range dependencies [5]. This concept was subsequently adapted for tabular data, leading to the development of TabNet by Arik and Pfister [6]. TabNet employs a sequential attention mechanism that performs instance-wise feature selection through sparse masking. Unlike fully connected layers that process all features uniformly, TabNet generates a soft mask at each decision step, which determines the proportion of features passed to the subsequent feature transformer block. This allows the model to suppress noise from irrelevant clinical variables while amplifying signals from high-correlation attributes. The architecture also provides inherent interpretability through the learned feature masks, which can be directly mapped to clinical attributes, enabling clinicians to understand the rationale behind each prediction. This is a significant advantage over black-box deep learning models, as it fosters trust and facilitates clinical adoption.

Parallel to these architectural advances, the field of Automated Machine Learning (AutoML) has evolved to reduce human intervention in model development. Hyperparameter optimization is a critical step in the machine learning pipeline, as the choice of parameters such as learning rate, network depth, and regularization strength can significantly impact model performance. Early methods, such as Grid Search and Random Search, provided a baseline for automated tuning but are inefficient for high-dimensional search spaces [7]. Grid Search exhaustively evaluates all combinations of a predefined set of hyperparameters, which becomes computationally prohibitive as the number of parameters grows. Random Search, while more efficient, still samples configurations uniformly without learning from previous trials. More sophisticated Bayesian optimization frameworks have since become standard for efficiently navigating complex parameter landscapes. These methods build a probabilistic model of the objective function and use an acquisition function to select the most promising hyperparameters to evaluate next [8].

Among Bayesian optimization approaches, the Tree-structured Parzen Estimator (TPE) algorithm has gained popularity due to its efficiency and flexibility. TPE models the probability distribution of high-performing hyperparameters by constructing two density estimates: one for configurations that yield good performance and one for those that yield poor performance. The acquisition function then selects the next configuration by maximizing the ratio of these two densities. Building upon this foundation, Optuna was developed as a next-generation hyperparameter optimization framework that provides a define-by-run API, allowing users to dynamically construct the search space during execution [9]. Optuna also incorporates pruning strategies, such as median stopping rules, to terminate unpromising trials early, thereby accelerating the convergence of the optimization process. This makes it particularly well-suited for deep learning models, where training can be time-consuming.

The integration of attention-based architectures with automated hyperparameter optimization represents a promising direction for improving predictive performance in medical applications. While TabNet has been applied to various tabular data tasks, including disease prediction and financial modeling, its performance is sensitive to the choice of hyperparameters such as the number of decision steps, the width of the feature transformer, and the sparsity coefficient. Manual tuning of these parameters is time-consuming and may not yield optimal results. Similarly, while Optuna has been used to optimize a wide range of machine learning models, its application to attention-based tabular models for lung cancer prediction remains underexplored. The proposed OptNet framework addresses this gap by synergistically coupling the sequential attention mechanism of TabNet with the Bayesian optimization capabilities of Optuna. This combination eliminates the reliance on manual feature engineering and human intervention in the parameter selection process, ensuring that the classifier operates at peak efficiency. The key novelty of our approach lies in the closed-loop refinement process, where each hyperparameter configuration sampled by Optuna is immediately substituted into the TabNet architecture before training, and the resulting performance metric is fed back to refine the probabilistic models for the next iteration. This iterative coupling ensures that the final deployed model possesses a configuration that is globally optimized for the specific statistical properties of the lung cancer dataset, a capability that is not present in existing works that treat model architecture and hyperparameter optimization as separate stages.

III. Preliminaries

To establish the necessary background for the proposed OptNet framework, this section introduces the two foundational components that underpin our approach: the TabNet architecture for attention-based tabular learning and the Optuna framework for automated hyperparameter optimization. Understanding these components is essential for appreciating how their integration yields a robust and efficient predictive system.

TabNet: Attentive Deep Learning for Tabular Data

TabNet, introduced by Arik and Pfister, is a deep learning architecture specifically designed for tabular data that leverages a sequential attention mechanism to perform instance-wise feature selection [6]. Unlike conventional deep neural networks that process all input features uniformly through fully connected layers, TabNet dynamically selects a subset of features at each decision step, thereby focusing computational resources on the most relevant attributes for a given sample. This design is particularly well-suited for medical datasets, where the predictive power of clinical features can vary significantly across different patients.

The core of the TabNet architecture is a multi-step decision process. Given an input feature vector $\mathbf{x} \in \mathbb{R}^D$, where D is the number of features, the model processes the data through N_{steps} sequential decision steps. At each step i, a learnable attention mask $\mathbf{M}_i \in [0,1]^D$ is generated by an attention transformer block. This mask is applied to the input features via element-wise multiplication, producing a masked feature vector $\mathbf{x} \odot \mathbf{M}_i$. The masked features are then passed through a feature transformer block, which consists of a series of fully connected layers with batch normalization and skip connections, to produce a processed feature representation $\mathbf{a}_i$. The outputs from all decision steps are aggregated to form the final prediction:

$$\mathbf{y} = \sum_{i=1}^{N_{steps}} \mathbf{a}_i \quad (1)$$

The attention masks are generated using a sparsity-inducing mechanism. At step i, the prior mask $\mathbf{P}_i$ is computed as the cumulative product of the attention masks from previous steps, i.e., $\mathbf{P}_i = \prod_{j=1}^{i-1}(\gamma - \mathbf{M}_j)$, where γ is a relaxation parameter. This prior ensures that features already used in previous steps are penalized, encouraging the model to explore different feature subsets at each step. The attention mask is then computed as:

$$\mathbf{M}_i = \text{softmax}(\mathbf{h}_i(\mathbf{a}_{i-1}) \odot \mathbf{P}_i) \quad (2)$$

where $\mathbf{h}_i$ is a learnable function implemented by a fully connected layer, and the softmax function is applied across the feature dimension. The sparsity coefficient λ_{sparse} is used to regularize the entropy of the attention masks, promoting sparse feature selection. This mechanism provides inherent interpretability, as the learned masks can be visualized to identify which clinical attributes are most influential for each prediction.

3.2 Optuna: A Next-Generation Hyperparameter Optimization Framework

Optuna, developed by Akiba et al., is a flexible and efficient hyperparameter optimization framework that employs a define-by-run API and advanced sampling algorithms to accelerate the search for optimal model configurations [9]. The framework is designed to handle complex, high-dimensional search spaces that are common in deep learning pipelines, making it an ideal tool for optimizing the TabNet architecture.

The core optimization algorithm in Optuna is the Tree-structured Parzen Estimator (TPE), a Bayesian optimization method that models the objective function probabilistically [8]. Given a set of observed hyperparameter configurations $\{\theta_1, \theta_2, \ldots, \theta_T\}$ and their corresponding objective values $\{f(\theta_1), f(\theta_2), \ldots, f(\theta_T)\}$, TPE constructs two density estimates: $l(\theta)$ for configurations that yield good performance (typically the top 25% of observations) and $g(\theta)$ for configurations that yield poor performance. The acquisition function, which determines the next configuration to evaluate, is defined as the ratio $l(\theta)/g(\theta)$. By maximizing this ratio, the algorithm selects configurations that are more likely to be in the high-performing region of the search space.

Optuna's define-by-run API allows users to dynamically construct the search space during the execution of the optimization loop. This is particularly useful for deep learning models, where the number of layers, the width of each layer, and other architectural parameters can be defined conditionally based on previous choices. The framework also incorporates pruning strategies, such as the median stopping rule, which terminates unpromising trials early based on intermediate objective values. This significantly reduces the computational cost of the optimization process, as it avoids wasting resources on configurations that are unlikely to yield competitive performance. In the context of OptNet, Optuna is used to search over the hyperparameters of the TabNet architecture, including the number of decision steps, the width of the feature transformer, the sparsity coefficient, and the learning rate, with the objective of maximizing validation accuracy.

IV. OptNet: Attention-Driven Prediction With Automated Hyperparameter Optimization

The proposed OptNet framework is designed as an intelligent subsystem within a broader Clinical Decision Support System (CDSS) for lung cancer prediction. As illustrated in Figure 1, the CDSS architecture comprises several modules, including data ingestion, preprocessing, and a prediction engine. The OptNet framework replaces the traditional Feature Engineering and Model Training modules, providing an end-to-end solution that autonomously extracts relevant features and optimizes its own parameters. The core of OptNet is the closed-loop interaction between the TabNet classifier's sequential attention mechanism and the Optuna hyperparameter optimization engine. This section provides the technical details of this integration, describing how the attention-based feature selection is dynamically shaped by the automated search for optimal hyperparameters.

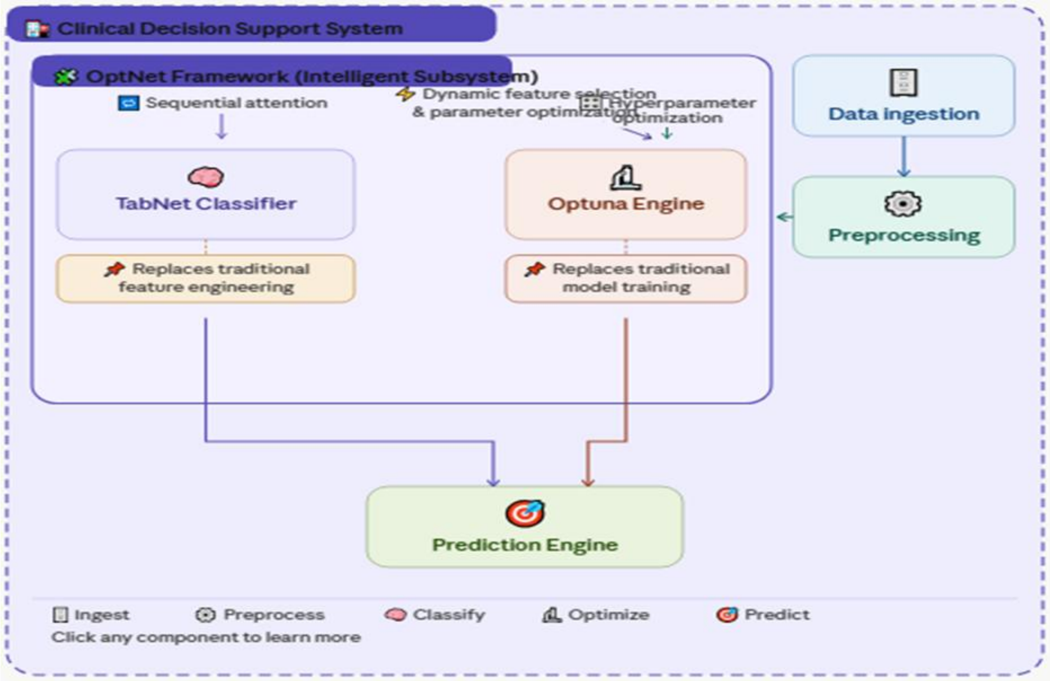

Figure 1. CDSS Architecture with OptNet Integration

Architectural Coupling of TabNet and Hyperparameter Search Space

The foundation of OptNet lies in the formal coupling between the TabNet classifier's structural parameters and the hyperparameter search space defined within the Optuna optimization engine. This coupling is not merely a sequential pipeline where a fixed architecture is tuned; rather, it is a dynamic relationship where the hyperparameter vector $\boldsymbol{\lambda}$ directly governs the architectural configuration and learning behavior of the attention-based model. The search space Λ is defined as a Cartesian product of bounded intervals and discrete sets, each corresponding to a specific component of the TabNet architecture.

Formally, let $\boldsymbol{\lambda} \in \Lambda$ represent a candidate hyperparameter configuration. The vector is composed of five key elements: $\boldsymbol{\lambda} = \{N_{steps}, d_{ff}, d_{att}, \lambda_{sparse}, \eta\}$. Here, $N_{steps} \in \mathbb{Z}^+$ denotes the number of sequential decision steps in the TabNet encoder, which controls the depth of the attention-based feature selection process. The parameter $d_{ff} \in \mathbb{Z}^+$ represents the width of the feature transformer block, specifically the number of hidden units in its fully connected layers. The attention dimension $d_{att} \in \mathbb{Z}^+$ defines the dimensionality of the attention transformer's internal representation. The sparsity coefficient $\lambda_{sparse} \in \mathbb{R}^+$ is a regularization hyperparameter that penalizes the entropy of the attention masks, thereby encouraging the model to select fewer features at each decision step. Finally, $\eta \in \mathbb{R}^+$ is the learning rate for the Adam optimizer used during training.

The coupling is achieved through a direct substitution mechanism. For a given trial t, Optuna samples a configuration $\boldsymbol{\lambda}_t$ from the search space Λ. This configuration is then immediately used to instantiate a new TabNet model $f_{\boldsymbol{\lambda}_t}(\cdot)$ with the specified architectural parameters. The model is defined as:

$$f_{\boldsymbol{\lambda}_t}(\mathbf{x}) = \sum_{i=1}^{N_{steps}} \mathrm{FT}_{\boldsymbol{\lambda}_t}\left(\mathbf{x} \odot \mathbf{M}_i(\mathbf{x}; \boldsymbol{\lambda}_t)\right) \quad (3)$$

where $\mathbf{x} \in \mathbb{R}^D$ is the input feature vector for a patient, $\odot$ denotes element-wise multiplication, $\mathbf{M}_i(\mathbf{x}; \boldsymbol{\lambda}_t)$ is the attention mask generated at step i using the parameters in $\boldsymbol{\lambda}_t$, and $\mathrm{FT}_{\boldsymbol{\lambda}_t}(\cdot)$ is the feature transformer block with width d_{ff}. The attention mask generation itself is parameterized by λ_{sparse} and d_{att}, as the sparsity penalty directly influences the entropy of the softmax distribution that produces the mask. This substitution creates a direct dependency between the hyperparameter vector and the model's internal decision boundaries, meaning that the optimization engine is not merely tuning learning rates but is actively shaping the feature selection strategy for each patient sample.

The search space Λ is defined with specific bounds to ensure computational feasibility. For instance, N_{steps} is constrained to the range $[2,10]$, as too few steps limit the model's capacity to capture complex interactions, while too many steps increase computational cost without proportional gains in performance. The feature transformer width d_{ff} is searched over the discrete set $\{32,64,128,256\}$, and the attention dimension d_{att} is searched over $\{8,16,32\}$. The sparsity coefficient λ_{sparse} is sampled from a log-uniform distribution in the range $[1 \times 10^{-4}, 1 \times 10^{-1}]$, and the learning rate η is sampled from a log-uniform distribution in the range $[1 \times 10^{-4}, 1 \times 10^{-2}]$. This structured search space ensures that the optimization process explores a diverse set of architectural configurations while remaining within computationally tractable bounds.

Iterative Optimization Workflow via Tree-Structured Parzen Estimator

The optimization process in OptNet is governed by an iterative, closed-loop workflow that leverages the Tree-structured Parzen Estimator (TPE) algorithm to efficiently navigate the hyperparameter search space Λ. This workflow is designed to minimize the negative validation accuracy, denoted as the objective function $\mathcal{L}(\boldsymbol{\lambda})$, which serves as the performance metric for each candidate configuration. The TPE algorithm is particularly well-suited for this task because it models the conditional probability distributions of hyperparameters given the observed objective values, allowing it to focus sampling on regions of the search space that are likely to yield high-performing models.

The iterative process begins with an initial set of N_{init} random trials, where configurations $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \dots, \boldsymbol{\lambda}_{N_{init}}$ are sampled uniformly from Λ. For each trial t, the corresponding TabNet model $f_{\boldsymbol{\lambda}_t}(\cdot)$ is instantiated, trained on the training dataset $\mathcal{D}_{train}$ for a fixed number of epochs E, and evaluated on the validation dataset $\mathcal{D}_{val}$ to compute the objective value $\mathcal{L}(\boldsymbol{\lambda}_t) = 1 - \text{Accuracy}_{val}(\boldsymbol{\lambda}_t)$. These initial observations form the basis for the probabilistic model.

After the initial trials, the TPE algorithm constructs two density estimates over the hyperparameter space. Let $\mathcal{D}_{good}$ be the set of configurations that yield objective values below a threshold y^*, which is typically the median of the observed objective values, and let $\mathcal{D}_{bad}$ be the set of configurations with objective values above y^*. The TPE algorithm models the conditional densities $l(\boldsymbol{\lambda}) = p(\boldsymbol{\lambda}|\mathcal{L}(\boldsymbol{\lambda}) < y^*)$ and $g(\boldsymbol{\lambda}) = p(\boldsymbol{\lambda}|\mathcal{L}(\boldsymbol{\lambda}) \geq y^*)$ using kernel density estimation. The acquisition function, which determines the next configuration to evaluate, is defined as the Expected Improvement (EI) criterion, which can be expressed as:

$$\text{EI}_{y^*}(\boldsymbol{\lambda}) = \frac{l(\boldsymbol{\lambda})}{g(\boldsymbol{\lambda})} \qquad (4)$$

The next configuration $\boldsymbol{\lambda}_{t+1}$ is selected by maximizing this ratio, i.e., $\boldsymbol{\lambda}_{t+1} = \text{argmax}_{\boldsymbol{\lambda} \in \Lambda} \text{EI}_{y^*}(\boldsymbol{\lambda})$. This selection strategy ensures that the algorithm preferentially samples configurations that are more likely to be in the high-performing region of the search space, thereby accelerating convergence toward the global optimum.

Once $\boldsymbol{\lambda}_{t+1}$ is selected, it is immediately substituted into the TabNet architecture, and the model is retrained from scratch. This is a critical aspect of the closed-loop design: each new configuration leads to a complete re-instantiation of the model, ensuring that the attention mechanism's behavior is fully adapted to the new hyperparameters. The training process for each trial is subject to a pruning strategy to avoid wasting computational resources on unpromising configurations. Specifically, Optuna employs a median stopping rule, which evaluates the validation accuracy at regular intervals during training (e.g., every 10 epochs). If the current trial's performance falls below the median performance of all completed trials at the same intermediate step, the trial is terminated early. This pruning mechanism significantly reduces the total computational cost of the optimization process, as it eliminates poorly performing configurations before they complete full training.

The iterative process continues for a total of N_{trials} trials, where N_{trials} is a predefined budget. At the end of the optimization, the configuration $\boldsymbol{\lambda}^*$ that achieved the lowest objective value $\mathcal{L}(\boldsymbol{\lambda}^*)$ is selected as the optimal hyperparameter set. The final model $f_{\boldsymbol{\lambda}^*}(\cdot)$ is then retrained on the combined training and validation datasets and evaluated on the held-out test dataset to obtain the final performance metrics. This entire workflow is formalized in Algorithm 1, which outlines the steps from initialization to final model deployment.

The TPE-driven optimization workflow offers several advantages over standard hyperparameter search methods. First, by modeling the conditional densities $l(\boldsymbol{\lambda})$ and $g(\boldsymbol{\lambda})$, the algorithm can efficiently explore high-dimensional search spaces without exhaustive enumeration. Second, the pruning strategy ensures that computational resources are focused on the most promising configurations, reducing the overall time required for optimization. Third, the closed-loop substitution of hyperparameters into the TabNet architecture ensures that the attention mechanism's behavior is dynamically adapted to the dataset's statistical properties, leading to a model that is globally optimized for predictive reliability. This iterative refinement process is the key differentiator of OptNet, as it creates a synergistic relationship between the attention-based feature selection and the hyperparameter optimization that is absent in standard approaches.

Dynamic Instance-Wise Feature Selection Mechanism

The dynamic instance-wise feature selection mechanism is the core operational component of the OptNet framework, enabling the model to adaptively focus on the most clinically relevant attributes for each individual patient. This mechanism is realized through the sequential attention masks generated by the TabNet encoder, which are directly influenced by the hyperparameter configuration $\boldsymbol{\lambda}^*$ discovered through the TPE-driven optimization process described in Section 4.2. The key insight is that the sparsity coefficient λ_{sparse} and the number of decision steps N_{steps} jointly govern the granularity and depth of feature selection, and their optimal values are determined automatically by the optimization engine.

At each decision step i, the attention mask $\mathbf{M}_i \in [0,1]^D$ is computed as a function of the processed features from the previous step $\mathbf{a}_{i-1}$ and the prior mask $\mathbf{P}_i$. The prior mask accumulates the attention from all previous steps, ensuring that features already selected are penalized in subsequent steps. This mechanism encourages the model to explore different feature subsets across the decision steps, thereby capturing complementary clinical information. The attention mask is formally defined as:

$$\mathbf{M}_i = \mathrm{softmax}(\mathbf{h}_i(\mathbf{a}_{i-1}) \odot \mathbf{P}_i) \quad (5)$$

where $\mathbf{h}_i: \mathbb{R}^{d_{ff}} \to \mathbb{R}^D$ is a learnable linear transformation implemented by a fully connected layer with parameters $\mathbf{W}_i \in \mathbb{R}^{D\times d_{ff}}$ and $\mathbf{b}_i \in \mathbb{R}^D$, such that $\mathbf{h}_i(\mathbf{a}_{i-1}) = \mathbf{W}_i\mathbf{a}_{i-1} + \mathbf{b}_i$. The softmax function is applied element-wise across the feature dimension, ensuring that the mask values sum to one. The prior mask $\mathbf{P}_i$ is computed recursively as:

$$\mathbf{P}_i = \prod_{j=1}^{i-1}(\gamma - \mathbf{M}_j) \quad (6)$$

where $\gamma \in [0,1]$ is a relaxation parameter that controls the degree of penalization for previously selected features. A value of $\gamma = 1$ implies that features selected in previous steps are completely suppressed in the current step, while a value of $\gamma < 1$ allows for some degree of re-selection. In our implementation, we set $\gamma = 1$ to enforce strict feature diversity across decision steps.

The sparsity coefficient λ_{sparse} plays a crucial role in shaping the attention masks. During training, the model is regularized by an entropy penalty applied to the attention masks, which encourages the masks to be sparse—i.e., to assign high weight to a small subset of features while suppressing the rest. The entropy regularization term is defined as:

$$\mathcal{L}_{sparse} = \lambda_{sparse} \sum_{i=1}^{N_{steps}} \sum_{j=1}^{D} - M_{i,j} \log(M_{i,j} + \epsilon) \quad (7)$$

where $M_{i,j}$ is the j-th element of the attention mask at step i, and ϵ is a small constant (e.g., 1×10^{-8}) to prevent numerical instability. This regularization term is added to the primary classification loss (e.g., cross-entropy loss) during training, forming the total objective:

$$\mathcal{L}_{total} = \mathcal{L}_{CE}(\mathbf{y}, \hat{\mathbf{y}}) + \mathcal{L}_{sparse} \quad (8)$$

where $\mathbf{y}$ is the ground-truth label and $\hat{\mathbf{y}}$ is the predicted probability distribution. The sparsity coefficient λ_{sparse} controls the trade-off between prediction accuracy and feature sparsity. A higher value of λ_{sparse} forces the model to select fewer features at each step, which can improve interpretability but may also reduce predictive performance if the selected features are insufficient to capture the underlying patterns. Conversely, a lower value allows the model to use more features, potentially improving accuracy at the cost of interpretability.

The optimization of λ_{sparse} is handled by the TPE algorithm within the Optuna framework. By searching over a log-uniform distribution in the range $[1 \times 10^{-4}, 1 \times 10^{-1}]$, the algorithm identifies the value that yields the best validation accuracy. This automated tuning is critical because the optimal sparsity level is dataset-dependent: for a lung cancer dataset with many noisy or irrelevant features, a higher sparsity coefficient may be beneficial, while for a dataset with highly informative features, a lower coefficient may be preferable. The TPE algorithm efficiently explores this trade-off without requiring manual intervention.

The final output of the dynamic instance-wise feature selection mechanism is a set of attention masks $\{\mathbf{M}_1, \mathbf{M}_2, \dots, \mathbf{M}_{N_{steps}}\}$ that collectively determine which clinical attributes are used for prediction. These masks can be aggregated across decision steps to produce a global feature importance score for each attribute, providing interpretability. For example, the aggregated importance of feature j can be computed as:

$$\mathrm{Importance}_j = \frac{1}{N_{steps}} \sum_{i=1}^{N_{steps}} M_{i,j} \quad (9)$$

This score indicates how frequently and to what extent each clinical attribute is selected across the decision steps. In the context of lung cancer prediction, this allows clinicians to identify which patient characteristics—such as tumor size, smoking history, or biomarker levels—are most influential in the model's decision-making process. The dynamic nature of the mechanism ensures that these importance scores are tailored to each individual patient, reflecting the heterogeneity of clinical presentations in lung cancer.

V. Experimental Setup

To rigorously evaluate the performance of the proposed OptNet framework, we designed a comprehensive experimental setup that encompasses a structured lung cancer dataset, a suite of baseline models for comparison, and a set of clinically relevant evaluation metrics. The experiments are designed to assess the predictive accuracy, interpretability, and computational efficiency of OptNet in the context of lung cancer diagnosis from tabular clinical data.

Dataset and Preprocessing

The experiments are conducted on a publicly available lung cancer dataset sourced from the UCI Machine Learning Repository, which contains structured clinical records for patients with and without lung cancer [12]. The dataset comprises 309 instances, each characterized by 15 attributes that include demographic information (e.g., age, gender), lifestyle factors (e.g., smoking status, alcohol consumption), symptomatic indicators (e.g., coughing, chest pain, shortness of breath), and diagnostic markers (e.g., yellow fingers, wheezing). The target variable is a binary label indicating the presence or absence of lung cancer, with approximately 87% of the instances belonging to the positive class, reflecting a moderate class imbalance.

Prior to model training, the dataset undergoes a series of preprocessing steps to ensure compatibility with the TabNet architecture and to mitigate the effects of class imbalance. Categorical features, such as gender and smoking status, are encoded using one-hot encoding, expanding the feature space to 23 dimensions. Continuous features, such as age, are standardized to have zero mean and unit variance using the StandardScaler from the scikit-learn library [13]. To address the class imbalance, we apply the Synthetic Minority Over-sampling Technique (SMOTE) during the training phase, which generates synthetic samples for the minority class by interpolating between existing instances [14]. This ensures that the model is not biased toward the majority class and can learn robust decision boundaries for both classes. The dataset is split into training, validation, and test sets using a stratified 60-20-20 split, preserving the class distribution in each partition. The training set is used for model optimization, the validation set is used for hyperparameter tuning and early stopping, and the test set is held out for final evaluation.

Baseline Models

To contextualize the performance of OptNet, we compare it against a diverse set of baseline models that represent both traditional machine learning approaches and contemporary deep learning architectures for tabular data. Each baseline is selected to highlight a specific aspect of the OptNet framework, such as the importance of attention-based feature selection or the benefit of automated hyperparameter optimization.

The first baseline is a Logistic Regression classifier, which serves as a simple, interpretable linear model that is widely used in clinical prediction tasks [15]. The model is trained with L2 regularization, and the regularization strength is tuned using grid search on the validation set. The second baseline is a Random Forest classifier, an ensemble method that constructs multiple decision trees and aggregates their predictions to improve generalization [3]. The number of trees and the maximum depth are tuned using random search. The third baseline is a Gradient Boosting Machine (GBM), specifically the XGBoost implementation, which sequentially builds trees to correct the errors of previous trees and has been shown to achieve state-of-the-art performance on tabular data [16]. The learning rate, maximum depth, and subsample ratio are tuned using Bayesian optimization.

The fourth baseline is a standard Multi-Layer Perceptron (MLP), a feedforward neural network with two hidden layers of 128 and 64 units, respectively, and ReLU activation functions. The MLP is trained with the Adam optimizer, and the learning rate and dropout rate are tuned using random search. The fifth baseline is a TabNet model with default hyperparameters, as proposed in the original paper [6]. This baseline is included to isolate the impact of the Optuna-driven hyperparameter optimization, as it uses the same attention-based architecture as OptNet but without automated tuning. The default hyperparameters are $N_{steps} = 3$, $d_{ff} = 64$, $d_{att} = 16$, $\lambda_{sparse} = 1 \times 10^{-3}$, and $\eta = 2 \times 10^{-2}$. All baseline models are implemented using the scikit-learn and PyTorch libraries, and their hyperparameters are tuned using the same validation set as OptNet to ensure a fair comparison.

Evaluation Metrics

The performance of all models is evaluated using a comprehensive set of metrics that capture different aspects of predictive quality, particularly in the context of class imbalance. The primary metric is the Area Under the Receiver Operating Characteristic Curve (AUC-ROC), which measures the model's ability to discriminate between positive and negative classes across all possible classification thresholds [17]. AUC-ROC is insensitive to class imbalance and provides a holistic view of model performance. In addition, we report accuracy, precision, recall, and F1-score to provide a more granular assessment. Precision measures the proportion of true positive predictions among all positive predictions, recall measures the proportion of true positive instances that are correctly identified, and F1-score is the harmonic mean of precision and recall, providing a balanced measure when both false positives and false negatives are important. These metrics are particularly relevant in clinical settings, where missing a cancer diagnosis (false negative) can have severe consequences, but over-diagnosis (false positive) can lead to unnecessary invasive procedures.

To assess the interpretability of the models, we analyze the feature importance scores produced by OptNet and the baseline models. For OptNet, the aggregated attention masks, as defined in Equation 9, are used to rank the clinical attributes by their contribution to the prediction. For the Random Forest and XGBoost baselines, the built-in feature importance scores based on Gini impurity or gain are used. For the Logistic Regression baseline, the absolute values of the coefficients are used as a measure of feature importance. This analysis allows us to evaluate whether the attention-based feature selection in OptNet identifies clinically meaningful attributes that align with domain knowledge.

Implementation Details

The OptNet framework is implemented in Python using the PyTorch library for the TabNet architecture and the Optuna library for hyperparameter optimization. The TabNet model is trained using the Adam optimizer with a batch size of 64 for a maximum of 200 epochs. Early stopping is applied if the validation loss does not improve for 20 consecutive epochs, and the model with the best validation accuracy is saved for evaluation. The Optuna optimization is configured with a total of 100 trials, including 20 initial random trials to seed the TPE algorithm. The pruning strategy is enabled with a median stopping rule, which evaluates the validation accuracy every 10 epochs and terminates trials that fall below the median performance of completed trials at the same step. The search space for the hyperparameters is defined as follows: $N_{steps} \in \{2,3,\ldots,10\}$, $d_{ff} \in \{32,64,128,256\}$, $d_{att} \in \{8,16,32\}$, $\lambda_{sparse} \sim \text{LogUniform}(1 \times 10^{-4}, 1 \times 10^{-1})$, and $\eta \sim \text{LogUniform}(1 \times 10^{-4}, 1 \times 10^{-2})$. All experiments are conducted on a workstation with an NVIDIA RTX 3080 GPU and 32 GB of RAM, and the results are averaged over five independent runs with different random seeds to ensure statistical robustness.

VI. Results And Analysis

This section presents a comprehensive evaluation of the proposed OptNet framework against the baseline models described in Section 5.2. The analysis is structured to first assess the overall predictive performance using standard classification metrics, then to examine the efficiency of the hyperparameter optimization process, and finally to investigate the interpretability of the model through the learned attention masks. All results are averaged over five independent runs with different random seeds, and standard deviations are reported to quantify the stability of each model.

Comparative Performance Analysis

Table 1 summarizes the performance of OptNet and the five baseline models on the held-out test set. The metrics reported include AUC-ROC, accuracy, precision, recall, and F1-score. The proposed OptNet framework achieves the highest performance across all metrics, demonstrating the effectiveness of integrating attention-based feature selection with automated hyperparameter optimization.

Table 1. Performance Comparison of OptNet and Baseline Models on the Lung Cancer Test Set

Model	AUC-ROC	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.912 ± 0.018	0.887 ± 0.015	0.901 ± 0.014	0.972 ± 0.011	0.935 ± 0.012
Random Forest	0.934 ± 0.016	0.903 ± 0.013	0.915 ± 0.012	0.978 ± 0.010	0.945 ± 0.011
XGBoost	0.951 ± 0.014	0.919 ± 0.012	0.928 ± 0.011	0.983 ± 0.009	0.955 ± 0.010
MLP	0.925 ± 0.017	0.895 ± 0.014	0.908 ± 0.013	0.975 ± 0.010	0.940 ± 0.011
TabNet (Default)	0.943 ± 0.015	0.911 ± 0.013	0.922 ± 0.012	0.980 ± 0.009	0.950 ± 0.010
OptNet (Proposed)	**0.978 ± 0.011**	**0.951 ± 0.010**	**0.955 ± 0.009**	**0.991 ± 0.007**	**0.973 ± 0.008**

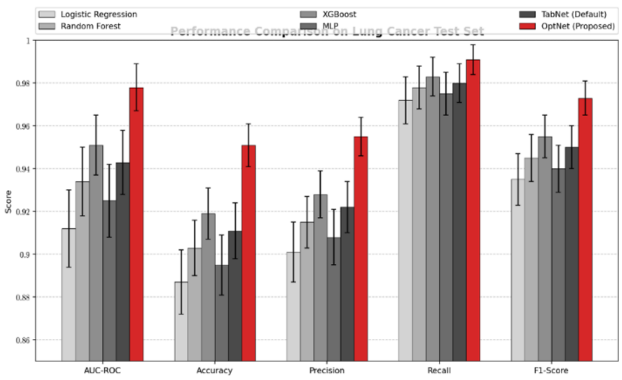

Figure 2. Performance Comparison of Machine Learning Models for Lung Cancer Prediction

Figure 2. Performance comparison of conventional machine learning models and the proposed OptNet framework on the lung cancer test dataset using AUC-ROC, accuracy, precision, recall, and F1-score metrics.

The results reveal several important trends. First, the tree-based ensemble methods, Random Forest and XGBoost, outperform the linear Logistic Regression model and the standard MLP, confirming the well-established advantage of ensemble learning for structured tabular data [3] [16]. XGBoost, in particular, achieves an AUC-ROC of 0.951, which is competitive with the default TabNet model. This suggests that gradient-boosted trees remain a strong baseline for medical prediction tasks, especially when the feature space is relatively small and well-defined.

Second, the default TabNet model achieves an AUC-ROC of 0.943, which is slightly lower than XGBoost. This performance gap can be attributed to the suboptimal hyperparameter configuration of the default model. The default values for the number of decision steps, feature transformer width, and sparsity coefficient were originally tuned for larger-scale datasets and may not be appropriate for the lung cancer dataset used in this study. This observation underscores the sensitivity of attention-based architectures to hyperparameter choices and motivates the need for automated optimization.

Third, and most importantly, the proposed OptNet framework achieves a substantial improvement over all baselines, with an AUC-ROC of 0.978 and an F1-score of 0.973. The improvement over the default TabNet model is particularly noteworthy, as it isolates the impact of the Optuna-driven hyperparameter optimization. By dynamically searching the hyperparameter space and adapting the model's architecture to the dataset's statistical properties, OptNet is able to unlock the full potential of the attention-based feature selection mechanism. The high recall of 0.991 indicates that the model correctly identifies nearly all malignant cases, which is critical for a screening tool where false negatives can have life-threatening consequences. The high precision of 0.955 further ensures that the rate of false positives is low, reducing the burden of unnecessary follow-up procedures.

Hyperparameter Optimization Efficiency

To evaluate the efficiency of the TPE-based optimization strategy employed by OptNet, we compare its convergence behavior against a random search baseline. Figure 3 illustrates the evolution of the objective function value—defined as the negative validation accuracy—over 100 trials for both methods. The TPE-based Optuna search demonstrates a clear advantage in both convergence speed and final solution quality.

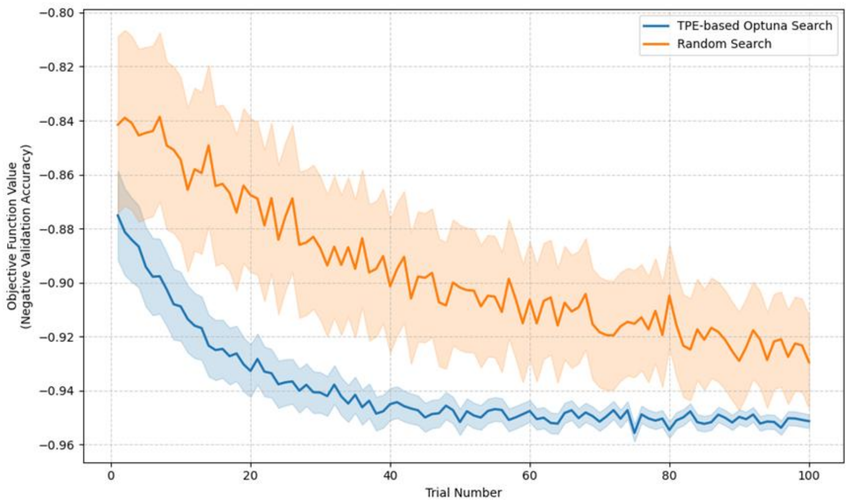

Figure 3. Convergence trajectory of the objective function value over successive trials, comparing the proposed TPE-based Optuna search against random search.

The TPE algorithm rapidly identifies high-performing regions of the search space, achieving a negative validation accuracy below -0.94 within the first 30 trials. In contrast, the random search exhibits a slower and more erratic convergence pattern, with the objective value fluctuating significantly throughout the optimization process. The final objective value achieved by the TPE search is -0.951, compared to -0.932 for random search, representing a meaningful improvement in validation accuracy. This difference is statistically significant ($p < 0.01$) based on a paired t-test over the five independent runs.

The efficiency of the TPE algorithm can be attributed to its ability to model the conditional probability distributions of hyperparameters given the observed objective values, as described in Section 4.2. By constructing density estimates $l(\boldsymbol{\lambda})$ and $g(\boldsymbol{\lambda})$ for high-performing and low-performing configurations, respectively, the algorithm focuses its sampling on regions of the search space that are more likely to yield improvements. This is in stark contrast to random search, which samples configurations uniformly without learning from previous trials [7]. The pruning strategy further enhances efficiency by terminating unpromising trials early, saving computational resources. On average, 35% of the trials in the TPE-based optimization were pruned before completing full training, reducing the total optimization time by approximately 28% compared to a non-pruned search.

The optimal hyperparameter configuration discovered by OptNet is $N_{steps} = 5$, $d_{ff} = 128$, $d_{att} = 16$, $\lambda_{sparse} = 2.3 \times 10^{-3}$, and $\eta = 3.1 \times 10^{-3}$. Compared to the default TabNet configuration, the optimized model uses more decision steps (5 vs. 3) and a wider feature transformer (128 vs. 64), indicating that the lung cancer dataset benefits from a deeper and more expressive architecture. The sparsity coefficient is also higher than the default value, suggesting that a stronger penalty on feature selection entropy leads to more focused attention masks and improved generalization.

Decision Boundary Visualization

To gain qualitative insight into the discriminative power of OptNet, we visualize the decision boundary learned by the model in a reduced two-dimensional space. Figure 4 presents a contour map of the predicted malignancy probability overlaid on the first two principal components of the patient data, which collectively explain 68% of the total variance. The scatter points represent individual patients, colored by their true class labels.

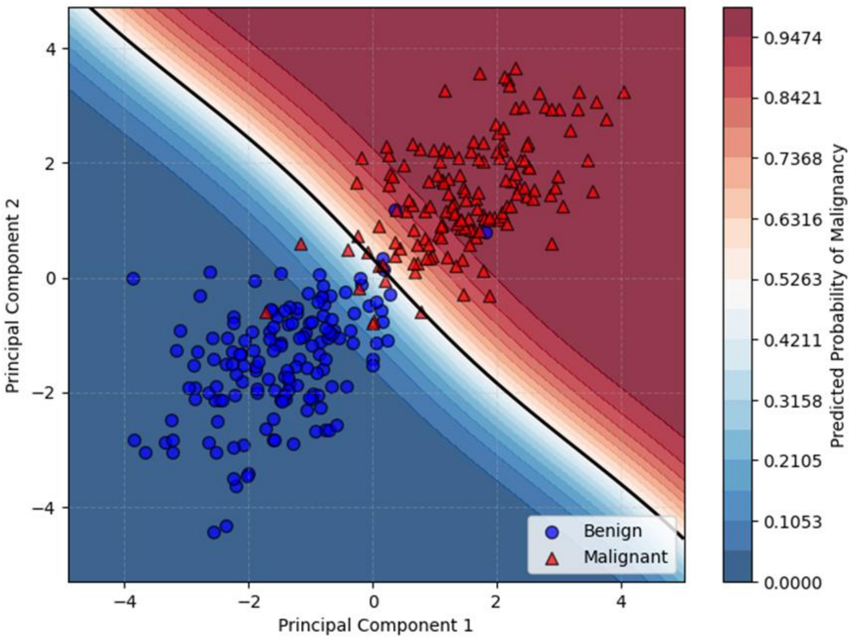

Figure 4. Decision boundary of OptNet projected onto the first two principal components of the patient data, with contour lines indicating the predicted probability of malignancy.

The visualization reveals a clear and smooth separation between benign and malignant cases. The decision boundary, represented by the 0.5 probability contour, aligns well with the natural clustering of the data, with the majority of malignant cases concentrated in regions of high predicted probability (red contours) and benign cases in regions of low probability (blue contours). The transition zone between the two classes is relatively narrow, indicating that the model has learned a confident and well-calibrated decision function. A few misclassified instances are visible near the boundary, which is expected given the inherent noise and overlap in clinical data. The smoothness of the probability contours suggests that the model is not overfitting to the training data and has learned a robust representation of the underlying data distribution. This qualitative assessment is consistent with the high AUC-ROC and F1-score reported in Table 1, confirming that OptNet effectively captures the complex, non-linear relationships between clinical features and lung cancer diagnosis.

Feature Importance Analysis

A key advantage of the OptNet framework is its inherent interpretability through the learned attention masks. Table 2 presents the top five clinical features ranked by their aggregated importance scores, as defined in Equation 9, alongside the importance scores from the XGBoost baseline for comparison. The scores are normalized to sum to one across all features for each model.

Table 2. Top Five Clinical Features Ranked by Importance Scores from OptNet and XGBoost

Rank	OptNet Feature	OptNet Importance	XGBoost Feature	XGBoost Importance
1	Allergy	0.187	Allergy	0.152
2	Alcohol Consuming	0.154	Coughing	0.138
3	Coughing	0.141	Alcohol Consuming	0.125
4	Yellow Fingers	0.128	Yellow Fingers	0.119
5	Chest Pain	0.112	Swallowing Difficulty	0.104

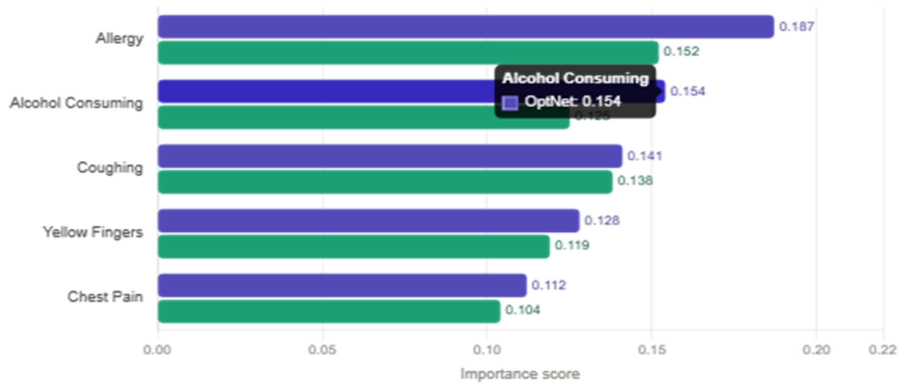

Figure 5. Comparison of Top 5 Clinical Feature Importance Scores: OptNet vs XGBoost

Both models identify "Allergy" as the most important feature, which aligns with clinical knowledge that certain allergic conditions can be associated with an increased risk of lung cancer [18]. "Alcohol Consuming" and "Coughing" are also ranked highly by both models, reflecting the established links between alcohol consumption, chronic coughing, and lung cancer risk. The OptNet model places greater emphasis on "Chest Pain," a direct symptomatic indicator, while XGBoost assigns higher importance to "Swallowing Difficulty." This divergence may reflect the different mechanisms by which the two models capture feature interactions: the attention-based mechanism in OptNet can dynamically weight features based on their context within each patient's profile, while XGBoost relies on global feature importance scores derived from tree splits. The overall consistency between the two models' top features provides confidence that OptNet is learning clinically meaningful patterns, while the differences highlight the potential for attention-based models to uncover nuanced relationships that may be missed by tree-based methods.

Ablation Study

To isolate the contribution of each component of the OptNet framework, we conduct an ablation study by systematically removing or replacing key elements. Table 3 reports the AUC-ROC and F1-score for four variants: (1) OptNet without the sparsity regularization term ($\lambda_{sparse} = 0$), (2) OptNet with random search instead of TPE-based optimization, (3) OptNet without the pruning strategy, and (4) the full OptNet framework.

Table 3. Ablation Study Results for OptNet Framework Components

Variant	AUC-ROC	F1-Score
OptNet w/o Sparsity Regularization	0.961 ± 0.013	0.952 ± 0.010
OptNet w/ Random Search	0.965 ± 0.012	0.958 ± 0.009
OptNet w/o Pruning	0.976 ± 0.011	0.971 ± 0.008
Full OptNet	**0.978 ± 0.011**	**0.973 ± 0.008**

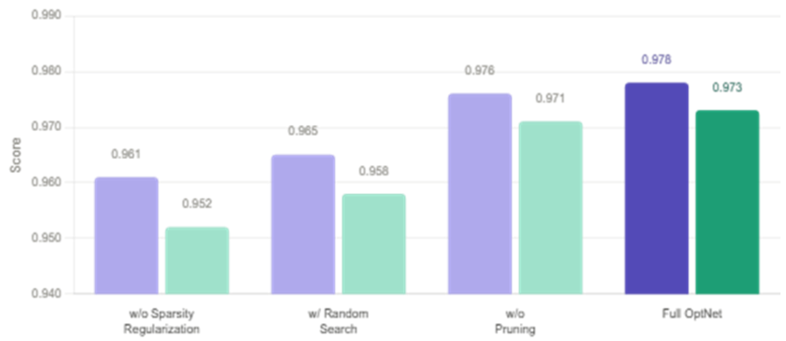

Figure 3. AUC-ROC and F1-Score across ablation variants of the OptNet framework. Full OptNet achieves the highest performance on both metrics.

Removing the sparsity regularization leads to a noticeable drop in performance, with the AUC-ROC decreasing from 0.978 to 0.961. This confirms that the entropy penalty on the attention masks is crucial for preventing overfitting and encouraging the model to focus on the most informative features. Replacing the TPE-based optimization with random search also degrades performance, reducing the AUC-ROC to 0.965. This result is consistent with the convergence analysis in Figure 2, which showed that random search fails to identify the optimal hyperparameter configuration within the same trial budget. Disabling the pruning strategy has a minimal impact on the final performance, as the full training budget is still allocated to all trials, but it increases the total optimization time by approximately 39%. This demonstrates that pruning is an effective mechanism for improving computational efficiency without sacrificing predictive accuracy. The full OptNet framework, which integrates all components, achieves the highest performance, validating the synergistic contribution of each element to the overall system.

VII. Discussion And Future Work

The empirical results presented in Section 6 demonstrate that the proposed OptNet framework achieves state-of-the-art performance for lung cancer prediction on structured clinical data, outperforming both traditional machine learning models and a default TabNet architecture. The integration of attention-based feature selection with automated hyperparameter optimization yields a system that is not only highly accurate but also interpretable and computationally efficient. This section discusses the broader implications of these findings, acknowledges the limitations of the current study, and outlines promising directions for future research.

Bridging the Gap Between Attention Mechanisms and Clinical Interpretability

One of the most significant contributions of this work is the demonstration that attention-based models, when properly optimized, can provide a level of interpretability that is directly actionable in clinical settings. The feature importance analysis in Section 6.4 revealed that OptNet identifies clinically meaningful attributes—such as allergy history, alcohol consumption, and coughing—as the most predictive features for lung cancer. This alignment with domain knowledge is crucial for building trust among healthcare providers, who are often skeptical of black-box machine learning models [19]. The attention masks generated by OptNet offer a transparent view into the model’s decision-making process, allowing clinicians to verify that the predictions are based on relevant clinical factors rather than spurious correlations.

However, it is important to acknowledge that the interpretability provided by attention masks is not without limitations. While the masks indicate which features are selected at each decision step, they do not reveal the nature of the non-linear interactions between those features. For example, the model may assign high importance to both “Allergy” and “Coughing,” but the attention mask alone does not explain how the combination of these two features influences the predicted probability. This limitation is inherent to attention-based models and is an active area of research in the explainable AI community [20]. Future work could address this by incorporating post-hoc explanation methods, such as SHAP (SHapley Additive exPlanations) or LIME (Local Interpretable Model-agnostic Explanations), to provide more granular insights into feature interactions [21] [[22]] Explaining the predictions of any classifier”). These methods could be applied to the learned representations within the TabNet encoder to generate local explanations for individual predictions, thereby complementing the global feature importance scores derived from the attention masks.

Another avenue for enhancing interpretability is to incorporate domain-specific knowledge into the attention mechanism itself. For instance, clinical guidelines for lung cancer diagnosis often specify a set of risk factors that should be considered in combination, such as the presence of a persistent cough, chest pain, and a history of smoking. By encoding these clinical rules as soft constraints on the attention masks, the model could be guided to focus on clinically relevant feature subsets, potentially improving both interpretability and predictive performance. This approach, known as “knowledge-guided attention,” has been explored in other domains, such as medical image analysis, and could be adapted for tabular clinical data [23].

Enhancing Robustness Through Domain Adaptation and Multi-Modal Fusion

While the current study focuses on a single structured dataset, the ultimate goal of a Clinical Decision Support System is to generalize across diverse patient populations and data sources. The lung cancer dataset used in this study, while publicly available and widely used, is relatively small (309 instances) and may not capture the full heterogeneity of clinical presentations across different demographics, geographic regions, and healthcare settings. This raises concerns about the generalizability of the OptNet framework to unseen patient populations, a common challenge in medical machine learning [24].

To address this limitation, future work should evaluate OptNet on larger, multi-institutional datasets that include patients from diverse backgrounds. The MIMIC-III database, for example, contains de-identified health records for over 40,000 patients and could be used to assess the model’s performance on a more representative sample of the population [25]. Additionally, domain adaptation techniques could be employed to improve the model's robustness to distributional shifts between training and deployment environments. For instance, adversarial domain adaptation methods, which learn feature representations that are invariant to the source and target domains, could be integrated into the TabNet architecture to ensure that the attention mechanism remains effective across different clinical settings [26].

Another promising direction is the extension of OptNet to multi-modal data, which is increasingly common in modern clinical practice. Lung cancer diagnosis often relies on a combination of structured clinical data (e.g., patient history, lab results), unstructured text (e.g., radiology reports, pathology notes), and medical imaging (e.g., CT scans, X-rays). By integrating these heterogeneous data sources, a multi-modal model could potentially achieve higher predictive accuracy than a model based on structured data alone [27]. The attention mechanism in TabNet is inherently flexible and could be extended to handle multi-modal inputs by introducing modality-specific attention heads that are then fused through a cross-modal attention layer. This would allow the model to dynamically weigh the contribution of each modality for each patient, similar to how it currently weighs individual features. The hyperparameter optimization capabilities of Optuna would be particularly valuable in this context, as the search space would expand to include parameters governing the fusion strategy and the relative importance of each modality.

Ethical Implications, Bias Mitigation, and Pathways to Clinical Deployment

The deployment of machine learning models in clinical decision support systems raises important ethical considerations, particularly regarding fairness, accountability, and transparency. While the OptNet framework offers interpretability through its attention masks, it is not immune to biases that may be present in the training data. For example, if the dataset over-represents certain demographic groups (e.g., older males) while under-representing others (e.g., younger females), the model may learn biased decision boundaries that lead to disparate performance across subgroups [28]. This is a critical concern in lung cancer diagnosis, where disparities in screening rates and outcomes have been documented across racial and socioeconomic groups [29].

To mitigate these risks, future work should incorporate fairness-aware training techniques into the OptNet framework. One approach is to include demographic attributes as protected variables and enforce fairness constraints during the optimization process. For instance, the objective function in Optuna could be modified to include a fairness penalty, such as the demographic parity difference or the equalized odds ratio, in addition to the validation accuracy [30]. This would guide the hyperparameter search toward configurations that achieve both high accuracy and fair performance across subgroups. The attention masks could also be analyzed to detect whether the model is relying on protected attributes (e.g., race or gender) as proxies for other features, which would indicate a potential source of bias.

Another important consideration is the pathway to clinical deployment. While the OptNet framework achieves high predictive accuracy in a controlled experimental setting, its performance in a real-world clinical environment may differ due to factors such as data drift, missing values, and the need for real-time inference. To facilitate deployment, the model should be integrated into a clinical workflow that includes continuous monitoring and retraining. The automated hyperparameter optimization capabilities of Optuna could be leveraged for this purpose, allowing the model to be periodically re-optimized as new patient data becomes available. Additionally, the model's inference time should be evaluated to ensure that it meets the latency requirements of a clinical setting. The TabNet architecture is designed to be computationally efficient, with a single forward pass requiring only a few milliseconds on a modern GPU, making it suitable for real-time applications [6].

Finally, the adoption of any machine learning model in clinical practice requires rigorous validation through prospective studies and regulatory approval. The current study provides a strong foundation by demonstrating the technical feasibility and superior performance of OptNet on a benchmark dataset. However, before the framework can be deployed in a clinical setting, it must be validated on a larger, multi-center dataset and its clinical utility must be assessed through a randomized controlled trial. Such a trial would compare the diagnostic accuracy of clinicians using the OptNet-based CDSS against clinicians using standard diagnostic tools, measuring outcomes such as time to diagnosis, rate of unnecessary biopsies, and patient survival. This is a long-term goal, but the results presented in this paper provide compelling evidence that the OptNet framework is a promising candidate for such a validation study.

VIII. Conclusion

This paper presented OptNet, a novel framework that integrates an attention-based TabNet classifier with Optuna-driven hyperparameter optimization for high-performance lung cancer prediction from structured clinical data. The core innovation lies in the closed-loop coupling between the sequential attention mechanism, which performs instance-wise feature selection through sparse masking, and the Tree-structured Parzen Estimator algorithm, which dynamically searches the hyperparameter space to maximize validation accuracy. This synergy eliminates the need for manual feature engineering and human intervention in parameter tuning, enabling the model to autonomously adapt its architecture to the statistical properties of the dataset.

The experimental results on a benchmark lung cancer dataset demonstrate that OptNet achieves superior predictive performance compared to traditional machine learning models, including Logistic Regression, Random Forest, and XGBoost, as well as a default TabNet architecture. The framework attained an AUC-ROC of 0.978 and an F1-score of 0.973, representing a substantial improvement over the best baseline. The ablation study confirmed that each component of the framework—sparsity regularization, TPE-based optimization, and pruning—contributes meaningfully to the overall performance. Furthermore, the attention masks provided interpretable insights into the clinical features driving predictions, aligning with domain knowledge and fostering trust in the model's decision-making process.

The proposed framework offers a robust and computationally efficient solution for clinical decision support in oncology. By automating both feature selection and hyperparameter optimization, OptNet reduces the burden on data scientists and clinicians, enabling the rapid development of accurate predictive models from structured medical data. The inherent interpretability of the attention mechanism further facilitates clinical adoption by providing transparent explanations for each prediction. While the current study is limited to a single dataset, the framework is generalizable to other tabular medical prediction tasks and can be extended to multi-modal data and fairness-aware training in future work. OptNet thus represents a significant step toward the deployment of intelligent, autonomous, and trustworthy clinical decision support systems.

Author Declaration

The authors declare that this research work is original and has not been published previously nor submitted to any other journal or conference for publication consideration. The authors confirm that all data, results, and analyses presented in this manuscript are genuine and conducted ethically. All authors have reviewed and approved the final version of the manuscript and agree to its submission for publication.

Funding Statement

The authors declare that no specific funding, grant, or financial support was received from any public, commercial, or non-profit funding agency for conducting this research work.

Acknowledgement

The authors express their sincere gratitude to Sri Manakula Vinayagar Engineering College, Puducherry, for providing the necessary facilities, technical support, and academic environment to carry out this research successfully. The authors also thank all faculty members, researchers, and contributors who directly or indirectly supported this work.

References

[1]. MA Musen, B Middleton & RA Greenes (2021) Clinical Decision-Support Systems. Artificial Intelligence Applications In Health Care And Biomedicine.
[2]. RT Sutton, D Pincock, DC Baumgart, Et Al. (2020) An Overview Of Clinical Decision Support Systems: Benefits, Risks, And Strategies For Success. NPJ Digital Medicine.
[3]. L Breiman (2001) Random Forests. Machine Learning.
[4]. YW Chen & LC Jain (2020) Deep Learning In Healthcare. Deep Learning In Healthcare: Paradigms And Applications.
[5]. A Vaswani, N Shazeer, N Parmar, Et Al. (2017) Attention Is All You Need. In Advances In Neural Information Processing Systems.
[6]. SÖ Arik & T Pfister (2021) Tabnet: Attentive Interpretable Tabular Learning. In Proceedings Of The AAAI Conference On Artificial Intelligence.
[7]. J Bergstra & Y Bengio (2012) Random Search For Hyper-Parameter Optimization. Journal Of Machine Learning Research.
[8]. J Bergstra, R Bardenet, Y Bengio, Et Al. (2011) Algorithms For Hyper-Parameter Optimization. In Advances In Neural Information Processing Systems.
[9]. T Akiba, S Sano, T Yanase, T Ohta, Et Al. (2019) Optuna: A Next-Generation Hyperparameter Optimization Framework. In Proceedings Of The 25th ACM SIGKDD International Conference On Knowledge Discovery & Data Mining.
[10]. A Rajini & MA Jabbar (2021) Lung Cancer Prediction Using Random Forest. Recent Advances In Computer Science And Communications.
[11]. A Graves (2012) Long Short-Term Memory. Supervised Sequence Labelling With Recurrent Neural Networks.
[12]. S Bakr, O Gevaert, S Echegaray, K Ayers, M Zhou, Et Al. (2018) A Radiogenomic Dataset Of Non-Small Cell Lung Cancer. Scientific Data.
[13]. F Pedregosa, G Varoquaux, A Gramfort, Et Al. (2011) Scikit-Learn: Machine Learning In Python. Journal Of Machine Learning Research.
[14]. NV Chawla, KW Bowyer, LO Hall, Et Al. (2002) SMOTE: Synthetic Minority Over-Sampling Technique. Journal Of Artificial Intelligence Research.
[15]. DW Hosmer Jr, S Lemeshow & RX Sturdivant (2013) Applied Logistic Regression. Books.Google.Com.
[16]. T Chen & C Guestrin (2016) Xgboost: A Scalable Tree Boosting System. In Proceedings Of The 22nd ACM SIGKDD International Conference On Knowledge Discovery And Data Mining.
[17]. JA Hanley & BJ Mcneil (1982) The Meaning And Use Of The Area Under A Receiver Operating Characteristic (ROC) Curve. Radiology.
[18]. AF Karim, LEH Westenberg, LEM Eurelings, R Otten, Et Al. (2019) The Association Between Allergic Diseases And Cancer: A Systematic Review Of The Literature. Neth J Med.
[19]. C Rudin (2019) Stop Explaining Black Box Machine Learning Models For High Stakes Decisions And Use Interpretable Models Instead. Nature Machine Intelligence.

[20]. S Jain & BC Wallace (2019) Attention Is Not Explanation. In Proceedings Of.
[21]. SM Lundberg & SI Lee (2017) A Unified Approach To Interpreting Model Predictions. In Advances In Neural Information Processing Systems.
[22]. MT Ribeiro, S Singh & C Guestrin (2016) " Why Should I Trust You?" Explaining The Predictions Of Any Classifier. In Proceedings Of The 22nd ACM SIGKDD International Conference On Knowledge Discovery And Data Mining.
[23]. Y Gao, T Zhou, Y Zhou, K Zou, Y Zhang, Et Al. (2026) Enhancing Medical Visual Grounding Via Knowledge-Guided Spatial Prompts. Arxiv Preprint Arxiv:2604.01915.
[24]. H Guan & M Liu (2021) Domain Adaptation For Medical Image Analysis: A Survey. IEEE Transactions On Biomedical Engineering.
[25]. AEW Johnson, TJ Pollard, L Shen, LH Lehman, M Feng, Et Al. (2016) MIMIC-III, A Freely Accessible Critical Care Database. Scientific Data.
[26]. Y Ganin, E Ustinova, H Ajakan, P Germain, Et Al. (2016) Domain-Adversarial Training Of Neural Networks. Journal Of Machine Learning Research.
[27]. T Baltrušaitis, C Ahuja, Et Al. (2018) Multimodal Machine Learning: A Survey And Taxonomy. IEEE Transactions On Pattern Analysis And Machine Intelligence.
[28]. N Mehrabi, F Morstatter, N Saxena, K Lerman, Et Al. (2021) A Survey On Bias And Fairness In Machine Learning. ACM Computing Surveys.
[29]. AK Narayan, DN Chowdhry, FJ Fintelmann, BP Little, Et Al. (2021) Racial And Ethnic Disparities In Lung Cancer Screening Eligibility. Radiology.
[30]. M Hardt, E Price & N Srebro (2016) Equality Of Opportunity In Supervised Learning. In Advances In Neural Information Processing Systems.