GMR
The Original

QATAB-SCLC: AN EXPLAINABLE QUALITY-AWARE MULTIMODAL PET/CT–CLINICAL MACHINE LEARNING FRAMEWORK FOR SMALL CELL LUNG CANCER PREDICTION

Jayaprakash B [1], Dr.S.K. Manju Bargavi [2]

[1]Asst. Professor, Department of Computer Science and IT, Jain Deemed-to-be University, Bangalore, b.jayaprakash@jainuniversity.ac.in
[2] Professor, Department of Computer Science and IT, Jain Deemed-to-be University, Bangalore, Email id: b.manju@jainuniversity.ac.in

ABSTRACT
Small Cell Lung Cancer (SCLC) is one of the most aggressive forms of lung cancer, characterized by rapid progression, early metastasis, and poor survival rates, making early and reliable prediction essential for improving clinical outcomes. Although recent advances in artificial intelligence have significantly enhanced lung cancer diagnosis, most existing studies primarily focus on classification or segmentation tasks using single-modality imaging and provide limited interpretability and prediction reliability. To address these limitations, this study proposes QATab-SCLC, an explainable Quality-Aware Multimodal PET/CT–Clinical Machine Learning Framework for SCLC probability prediction. The proposed framework integrates CT, PET, fused PET/CT images, and clinical information within a unified multimodal learning architecture. Initially, an image quality assessment module quantifies image reliability using statistical and texture-based quality descriptors. Handcrafted radiomic features and deep semantic representations extracted using a pretrained ResNet18 network are subsequently fused with clinical variables to construct comprehensive multimodal feature representations. After preprocessing and Mutual Information-based feature selection, two complementary tabular learning models, TabPFN and TabM, independently estimate SCLC probabilities. Their outputs are dynamically combined through a Quality-Aware Adaptive Fusion Network, which determines optimal fusion weights based on image quality and prediction uncertainty to generate robust and well-calibrated probability estimates. To enhance transparency, SHapley Additive exPlanations (SHAP) are incorporated to provide both global feature importance and patient-specific decision explanations. Experimental evaluation demonstrates that the proposed framework consistently outperforms existing machine learning and deep learning approaches across discrimination, calibration, and probability prediction metrics, including ROC-AUC, PR-AUC, Brier Score, Log Loss, RMSE, and Expected Calibration Error. The proposed QATab-SCLC framework offers a reliable, interpretable, and clinically applicable decision-support system for early SCLC probability prediction, supporting more informed diagnosis and personalized treatment planning.

KEYWORDS: Small Cell Lung Cancer (SCLC), PET/CT Imaging, Multimodal Learning, TabPFN, TabM, Quality-Aware Learning, Adaptive Fusion, Probability Prediction.

1 INTRODUCTION

Lung cancer remains a significant global health concern, with staggering mortality rates. According to GLOBOCAN, it accounted for 2.1 million new cases and 1.8 million deaths in 2018, making it the leading cause of cancer-related deaths worldwide [1]. Lung cancer is categorized into two main histological types: non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC). NSCLC comprises approximately 85% of cases, while SCLC represents around 15% [2]. SCLC is an aggressive neoplasm characterized by rapid proliferation, early dissemination, metastases, acquired therapy resistance, and poor outcomes. Each year, approximately 250,000 new cases of SCLC are reported, resulting in at least 200,000 deaths worldwide. While historically more common in men, the prevalence of SCLC among women has risen due to global smoking trends. Exposure to tobacco carcinogens (polycyclic aromatic hydrocarbons and tobacco-specific nitrosamines) is considered a key risk factor for SCLC, as only 2% of all SCLC cases are among never-smokers. Early diagnosis of SCLC is crucial as most patients present with metastatic disease, limiting the potential for curative treatment [2].

As a consequence, correct identification is critical for choosing the most effective treatment method. At this point, computed tomography (CT) imaging and radiology reports are critical. CT offers a good portrayal of tumor location, delineates lesion size and margins, and aids in the selection of the best biopsy trajectory. Radiology reports thoroughly detail these results, providing crucial assistance to both the interventional team conducting the biopsy and the clinical team in charge of treatment planning [3].

Manual interpretation of radiological images continues to be frequently utilized in the diagnosis and monitoring of lung cancer [4]. However, it is labor-intensive, needs extensive skill, takes a significant amount of time, and leads to a hefty clinical burden. A radiologist may need to analyze thousands of CT slices individually in order to discover lesions and provide a full report. Increasing patient loads and professional obligations aggravate these

issues [5]. In areas with limited access to competent radiologists, interpretation times may be extended, resulting in diagnostic delays. Furthermore, manual image interpretation is fundamentally subjective; the same radiological data might be interpreted differently by various professionals [6]. Such heterogeneity may diminish diagnostic accuracy, inject uncertainty into treatment planning, and raise the likelihood of poor patient outcomes.

Given these limits, there is an increasing need for novel technologies that might improve and standardise diagnostic operations. Rapid advances in digital imaging and artificial intelligence have created a solid platform for overcoming the constraints of manual image interpretation [7].

In recent years, deep learning has catalyzed a substantial methodological transformation in medical image analysis and has swiftly emerged as a fundamental methodology in the field. Convolutional neural networks (CNNs) have been extensively used in diagnostic systems owing to their hierarchical feature extraction skills and proficiency in collecting local spatial patterns [8]. The fundamentally restricted receptive field of convolutional layers limits their capacity to model extensive contextual relationships and adequately reflect intricate anatomical structures [9]. This constraint is particularly evident in high-dimensional medical imaging, where subtle or diffuse disease patterns need a more comprehensive representation.

Transformer-based designs have emerged as viable options in medical imaging [10]. Self-attention mechanisms' global perceptual capability allows these models to successfully capture long-range dependencies and nuanced contextual interactions throughout the image [11] As a consequence, they have achieved significant success in tasks requiring high-level spatial thinking, such as tumor categorization [12], tissue segmentation [13], [14], and lesion location [15]. Nonetheless, great predictive accuracy is not enough for clinically relevant artificial intelligence systems. Furthermore, several contemporary methodologies depend on single-slice analysis, which insufficiently represents the volumetric characteristics of CT data [4]. Furthermore, despite recent advances in explainable artificial intelligence (XAI), many studies provide limited insight into the decision-making process, reducing clinicians' confidence in deploying these models for routine diagnosis [16].

Although there have been significant achievements in using deep learning techniques and transformer models for lung cancer diagnosis, several issues still need to be resolved. To date, most existing research is dedicated to classification and segmentation tasks, as well as NSCLC diagnosis, but not Small Cell Lung Cancer (SCLC). Moreover, the vast majority of the existing methods were designed and evaluated based on open-access data sets including LIDC-IDRI, LUNA16, NLST, TCIA, among others. Only a few approaches used the Lung-PET-CT-Dx dataset for multimodal PET/CT investigation. Current models mostly utilize single-modal images and classical deep learning networks and rarely use complementary anatomical and metabolic features of PET/CT along with clinical parameters via a hybrid approach. Also, most existing methods treat the problem as binary classification and not as estimation of the calibrated probability.

Recent research has also brought to light the significance of image quality differences, image acquisition, and uncertainties in the predictions, which have a huge impact on the robustness of the models; however, all of these have not yet been included in multimodal learning frameworks. Additionally, while there is much interest in the field of explainable AI, most existing approaches offer visual explanations of the predictions, but do not incorporate the image quality and prediction uncertainties in the decision-making process. To bridge this gap, we present QATab-SCLC – an Explainable Quality-aware Multimodal PET/CT-Clinical Machine Learning Framework for SCLC Probability Prediction.

In this approach, the Lung-PET-CT-Dx dataset is employed via integration of CT, PET, PET/CT fusion images, and possibly clinical data into a single multimodal learning pipeline. Radiomic handcrafted features, deep semantic features obtained by ResNet18 model, and clinical features are used to fuse and build a rich multimodal feature vector. After that, two different hybrid prediction models, TabPFN and TabM, separately predict the probabilities of SCLC, which are then combined adaptively using a Quality-Aware Adaptive Fusion Network depending on the image quality factors and prediction uncertainties. In addition, SHapley Additive exPlanations (SHAP) are used to interpret globally and individually, leading to a calibrated and interpretable system for SCLC probability prediction.

The major contributions of this study are summarized as follows:

- A novel Quality-Aware Multimodal PET/CT–Clinical Machine Learning Framework (QATab-SCLC) is proposed for reliable Small Cell Lung Cancer probability prediction by integrating CT, PET, fused PET/CT images, and clinical information within a unified prediction framework.
- A comprehensive multimodal feature learning strategy is developed by combining handcrafted radiomic descriptors, deep semantic features extracted using ResNet18, and clinical variables to capture complementary anatomical, metabolic, and patient-specific information.
- A Quality-Aware Adaptive Fusion Network is introduced to dynamically combine the probability predictions generated by TabPFN and TabM according to image quality indicators and prediction uncertainty, improving robustness and probability calibration.
- An explainable prediction mechanism based on SHAP is incorporated to provide transparent global and patient-level interpretations, thereby enhancing clinician trust and facilitating practical clinical decision support.

2 RELATED WORK

Mohamed & Ezugwu, [17] A novel deep-learning model for lung cancer detection has been developed using mRNA, miRNA, and DNA methylation indicators. With DESeq2 for RNASeq data and LIMMA for miRNA and DNA methylation, the study identified genes with varied expression throughout lung cancer stages I to IV by

rigorous data preparation and differential analysis. By combining omics data, 448 samples and 8228 characteristics were consolidated. Streamlining features and class balance were achieved using PCA and SMOTE. After classification, the PCA-SMOTE-CNN model outperformed previous competing approaches with accuracy, precision, recall, and F1-scores of 0.97.

Pandit et al., [18] utilized multispace image in convolution neural network pooling layer to improve prediction accuracy and processing time. Autoencoder system improves accuracy, and multispace image in pooling layer of convolution neural network and Adam Algorithm optimize lung cancer prediction. In comparison to the state of the art, the suggested approach has excellent classification accuracy and low processing time.

Swain et al., [19] proposed dense neural networks (VGG-16 and ResNet-50) and a simpler sparse neural network (Inception v3) to classify non-small cell lung cancer from CT scans. The trial includes 60 adenocarcinoma and 60 squamous cell carcinoma patients. The Inception v3 network outperforms VGG-16 and ResNet-50 with 96.66% sensitivity, 99.12% specificity, and 98.29% accuracy. Inception v3 has decreased processing overhead, which may help radiologists classify lung cancer using CT images.

Carrillo-Perez et al., [20] proposed late fusion and machine learning to integrate five multi-scale and multi-omic modalities: RNA-Seq, miRNA-Seq, whole-slide imaging, copy number variation, and DNA methylation. We build an individual machine learning model for each modality and evaluate the interactions and improvements from merging their outputs using a unique optimization strategy. The final classification model, including all modalities, outperforms individual models with an F1 score of 96.81±1.07, AUC of 0.993±0.004, and AUPRC of 0.980±0.016.

Chandrasekar et al., [21] used MRRXGBDC and SOA to improve lung cancer prediction accuracy and time in big data analytics. Clinicians may design multivariate Ruzicka logistic regression trees to detect illness using SOAs to safely store patient data. Over previous approaches, MRRXGBDC improves lung cancer prediction accuracy by 10%, false positives by 50%, and prediction time by 11%.

Hosny et al., [22] developed a DL framework that combines X-rays, CT, and histopathology images. The method uses Inception-ResNet to extract multi-scale spatial features and enhance deep feature representations using residual learning. The suggested model was 95.35%, 99.68%, 99.73%, and 99.26% accurate utilizing X-ray, CT, histopathology, and mega datasets. Experimental results showed that the suggested model detected lung cancer better than traditional DL structures.

H. Zhang et al., [23] introduced a multimodal deep learning lung cancer detection approach using Fourier transform infrared, UV–vis absorbance, fluorescence, and Raman spectra. Each spectrum has a one-dimensional sequence and a two-dimensional Gramian Angular Summation Field picture. A dual-branch design extracts 1D and 2D characteristics, which are combined via a MambaVision module for increased interaction and context modeling. The technique performs well with 97.65% accuracy, 98.14% precision, 97.52% recall, 97.82% F1-score, and 99.76% AUC.

Deng et al., [24] proposed a cross-modality attention-based multimodal fusion pipeline for NSCLC survival prediction. The approach improves integration by assessing each modality's feature fusion relevance. The suggested method exceeded the performance of individual modalities, attaining a c-index of 0.6587, in contrast to c-index scores of 0.5772 for tissue image data and 0.5885 for RNA-seq data, so illustrating its effectiveness in leveraging modality-specific information.

[25] proposed the Multimodal Multiscale Cross-Attention Fusion Network (MMCAF-Net) to overcome the dimensionality issues between medical imaging and electronic health data. This model extracts lesion-specific characteristics from 3D pictures using a 3D multi-scale convolutional attention module and a feature pyramid structure. Furthermore, to overcome dimensional inconsistencies, MMCAF-Net has a multi-scale cross-attention module for efficient feature fusion. When tested on the Lung-PET-CT-Dx dataset, MMCAF-Net outperformed the state-of-the-art methods with notable gains in diagnosis accuracy.

[26] utilized expert-annotated datasets from 279 oesophageal and 54 lung cancer patients to assess nnU-Net. Target-only, public-only (AutoPET), and combination training are offered. While the oesophageal model had good in-domain accuracy (mean DSC, 57.8), it overfitted external lung data (mean DSC < 3.4). Public model generalized well (mean DSC, 63.5 on AutoPET, 51.6 on lung cohort) but underperformed on oesophageal data (26.7). Dataset variety is needed for generalization beyond model complexity, since the combined technique produced balanced findings (mean DSC: lung 52.9, oesophageal 40.7, AutoPET 60.9).

[27] described a stacking-based ensemble machine learning strategy for categorizing SCLC and NSCLC utilizing metabolomics data from 191 SCLC, 173 NSCLC, and 97 healthy controls. Our multi-class classification accuracy was 85.03% and 92.47 AUC, while our binary classification accuracy was 88.19% and 92.65 AUC using ExtraTreesClassifier. SHAP identify benzoic acid, DL-lactate, and L-arginine as key metabolites. By merging metabolomics and machine learning, the approach may improve early non-invasive lung cancer diagnosis.

2.1 Research Gap

Even though many studies have shown good performance with the use of deep learning, multimodal learning, ensembling, and machine learning in diagnosing lung cancer, the vast majority of these studies has been dedicated either to classification, segmentation, survival prediction or to related problems of NSCLC using datasets such as LIDC-IDRI, LUNA16, TCIA, and others. On the other hand, only few studies have addressed the problem of predicting SCLC using the Lung-PET-CT-Dx dataset, which provides additional PET/CT data. Furthermore, currently used multimodal approaches have used traditional deep learning models and very few approaches have

used CT, PET, fused PET/CT images, and clinical information in combination with each other. Also, existing methods output their categorical classification results without considering the issues of probability prediction based on quality-awareness, variations in image quality and uncertainty of predictions, whereas explainability mostly consists of post hoc explanation. Thus, there is a need to develop a novel QATab-SCLC approach, which would consider multimodal quality-aware feature learning, hybrid probability prediction based on TabPFN and TabM, adaptive fusion, and SHAP explainability for SCLC probability prediction.

3 Background

3.1 TabPFN

TabPFN denotes Tabular Prior-data Fitted Network, a foundational model predicated on the concept of calibrating a model to a prior distribution across tabular datasets, rather than to an individual dataset [28]. TabPFN utilizes in-context learning (ICL), the process responsible for the remarkable efficacy of big language models, to create a robust tabular prediction method that is entirely learnt. While ICL was first identified in big language models, subsequent research has shown that transformers may acquire basic algorithms, such as logistic regression, via ICL [29].

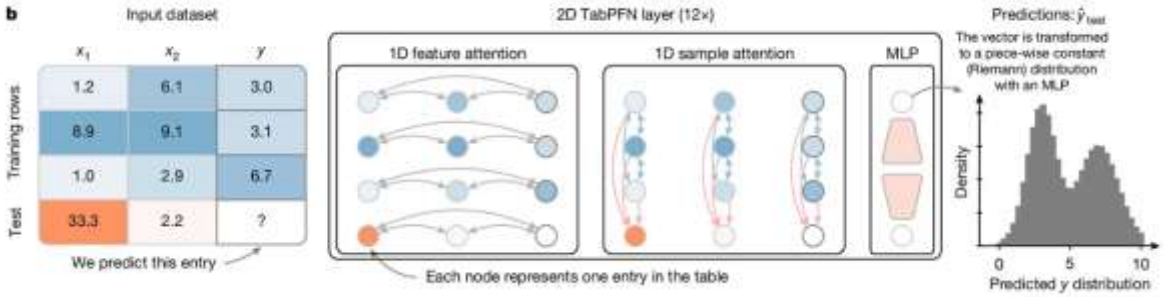

Figure 1 Architecture of the TabPFN

3.2 TabM

TabM (Tabular Deep Learning model that generates Multiple predictions) is a simple but robust tabular deep learning architecture that effectively emulates an ensemble of Multi-Layer Perceptrons. The two primary distinctions of TabM in relation to a conventional ensemble of MLPs are:

Concurrent training of the MLPs. This facilitates the assessment of the ensemble's performance during training and enables the cessation of training when optimum for the ensemble rather than for individual MLPs.

Weight allocation among the MLPs. The whole TabM is encapsulated under a single MLP-like model. This not only enhances runtime and memory efficiency but also serves as an effective regularization technique, resulting in improved job performance.

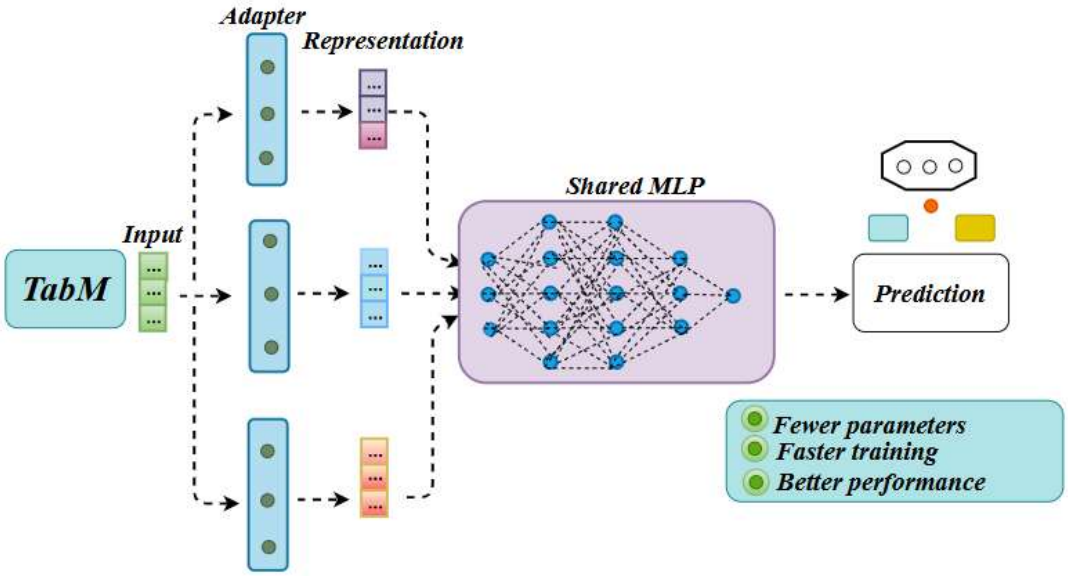

Figure 2 Architecture of the TabM model

Researchers from Yandex and HSE University developed a model called TabM, which is based on a multi-layer perceptron (MLP) architecture and augmented with BatchEnsemble for efficient parameter ensembling. This model produces several predictions inside a unified framework by sharing the majority of its weights across ensemble components, enabling it to provide varied, weakly linked predictions. TabM integrates simplicity with efficient ensembling to achieve a mix of performance and efficiency, striving to surpass conventional MLP models

while avoiding the intricacies of transformer topologies. TabM presents an efficient alternative, delivering benefits for deep learning while minimizing the substantial resource requirements often linked to sophisticated models [30].

4 METHODOLOGY

The proposed Quality-Aware TabPFN–TabM Hybrid Framework (QATab-SCLC) is designed to predict the probability of Small Cell Lung Cancer (SCLC) using multimodal PET/CT images together with optional clinical information. Instead of directly classifying images, the framework estimates the probability of SCLC by integrating handcrafted radiomic descriptors, deep image representations, and clinical variables within a quality-aware learning architecture. The complete workflow consists of seven major stages, as illustrated in Fig. X.

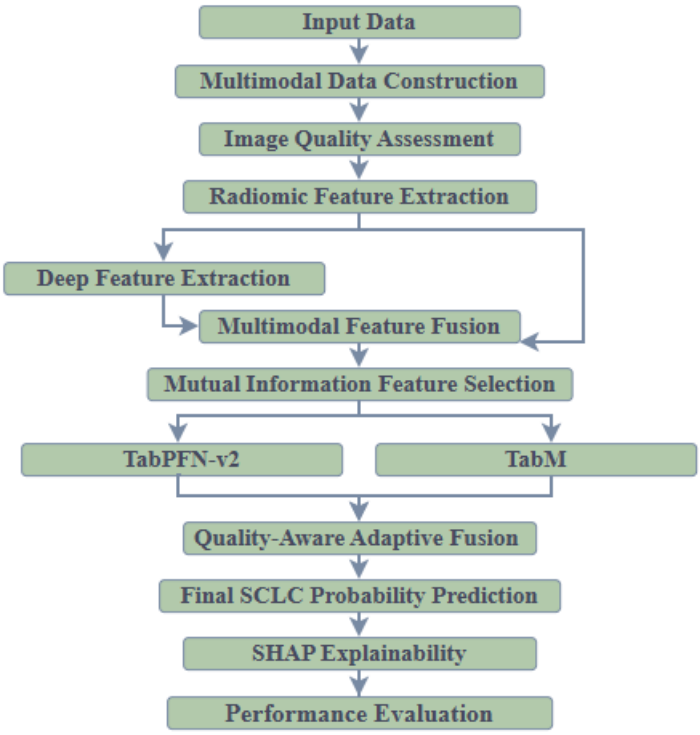

Figure 3 Flowchart of the Proposed Work

4.1 Dataset

The proposed framework utilizes the publicly available Lung-PET-CT-Dx dataset, sourced from the publicly available website: https://www.kaggle.com/datasets/rangan2510/lung-pet-ct-dx

This dataset containing PET images, CT images, and corresponding clinical information collected from diagnosed lung cancer patients.

Since images belonging to different modalities may have different naming conventions, a canonical image identifier is generated from each filename. Images are grouped into three histological categories:

- Adenocarcinoma
- Squamous Cell Carcinoma
- Small Cell Lung Carcinoma

For prediction, the first two classes are combined into the NSCLC category, while Small Cell Carcinoma forms the positive SCLC class. Each sample therefore consists of

$$X_i = \{CT_i, PET_i, FUSED_i, C_i\} \quad (1)$$

Where, CT_i denotes the CT image, PET_i denotes the PET image, $FUSED_i$ represents the fused PET/CT image, C_i represents optional clinical attributes.

Only complete multimodal image pairs are retained for subsequent analysis. This pairing process eliminates incomplete observations and prevents inconsistent multimodal learning.

Table 1 An illustration of the Dataset Description

Modality	Adenocarcinoma	Small Cell Carcinoma	Squamous Cell Carcinoma	Total
CT	3,352	245	901	4,498
Fused Images	3,352	245	901	4,498
PT	3,352	245	901	4,498
Total	**10,056**	**735**	**2,703**	**13,494**

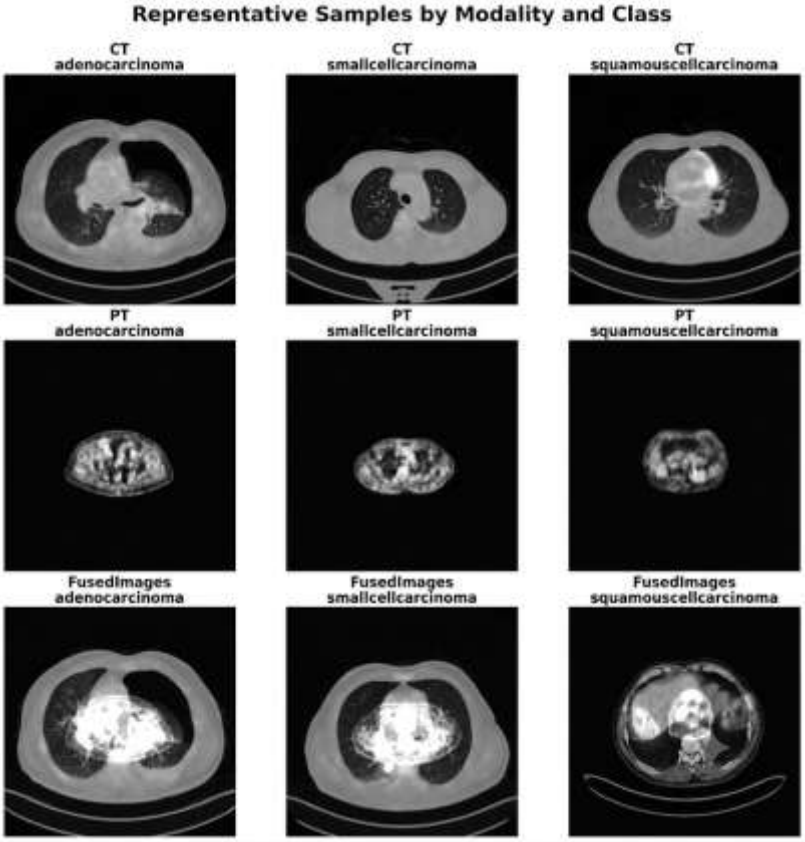

Figure 4 Representative Samples by Modality and Class

4.2 Data pre-processing

Medical image quality significantly influences prediction reliability. Therefore, before feature learning, the proposed framework evaluates the quality of each CT, PET, and fused PET/CT image by computing quantitative image quality descriptors, including image entropy, standard deviation, Laplacian variance, edge density, GLCM contrast, and GLCM homogeneity.

These quality indicators later guide the adaptive fusion mechanism by estimating the reliability of each multimodal input. For each modality, the quality vector is defined as Eq. (2)

$$Q_i = [\sigma_i, H_i, L_i, E_i, C_i, H_{g_i}], \quad (2)$$

where σ_i, H_i, L_i, E_i, C_i, and H_{g_i} represent intensity variation, Shannon entropy, Laplacian variance, edge density, GLCM contrast, and GLCM homogeneity, respectively.

4.3 Feature Extraction

After preprocessing, complementary imaging features are extracted using two different strategies. First, handcrafted radiomic descriptors are computed from each image, including first-order intensity statistics, histogram-based features, entropy, skewness, kurtosis, edge information, Local Binary Pattern (LBP), and Gray-Level Co-occurrence Matrix (GLCM) texture features. These descriptors capture structural and textural characteristics of lung lesions.

To complement handcrafted features, pretrained ResNet-18 is employed as a fixed deep feature extractor. Each multimodal image is transformed into a high-level semantic representation, which is subsequently compressed using Gaussian Random Projection (GRP) to reduce feature redundancy while preserving discriminative information.

The extracted handcrafted and deep features are represented as Eq. (3-4)

$$F_H = [f_1, f_2, \dots, f_n], \quad (3)$$

$$F'_D = GRP(ResNet18(I)) \quad (4)$$

This dual-feature extraction strategy enables the framework to capture both low-level radiomic information and high-level semantic representations from multimodal medical images.

4.4 Multimodal Feature Fusion and Feature Selection

Following feature extraction, the handcrafted radiomic features, projected deep features, and optional clinical variables are integrated into a unified multimodal representation as Eq. (5),

$$F = [F_H, F'_D, C] \quad (5)$$

where F_H denotes handcrafted radiomic features, F'_D represents projected deep features, and C corresponds to optional clinical variables.

Prior to model learning, the integrated features undergo median-based missing value imputation, variance threshold filtering, and feature standardization to improve data consistency and numerical stability.
To eliminate redundant information and improve computational efficiency, Mutual Information (MI)-based feature selection is employed. MI measures the dependency between each feature and the target variable, thereby identifying the most informative predictors for SCLC probability estimation.
The selected feature subset is expressed as Eq. (6)

$$F^* = MI(F) \quad (6)$$

where F^* denotes the optimized multimodal feature representation used for probability prediction. This feature refinement stage significantly reduces dimensionality while preserving discriminative information, thereby enhancing generalization performance and reducing the risk of model overfitting.

4.5 Parallel Probability Prediction Using TabPFN and TabM

Following multimodal feature fusion and feature selection, the refined feature subset is simultaneously processed by two complementary probability prediction models, namely TabPFN and TabM. Rather than relying on a single predictor, the proposed framework adopts a dual-branch learning strategy to exploit the complementary strengths of probabilistic foundation models and deep neural tabular learning. This parallel prediction mechanism improves robustness while reducing the dependence on a single learning paradigm.
The first branch utilizes TabPFN, a pretrained probabilistic foundation model that estimates posterior probabilities using prior knowledge acquired from numerous synthetic tabular learning tasks. The predicted probability generated by this branch is

$$P_{PFN} = P(\text{SCLC} \mid F^*) \quad (7)$$

The second branch employs the TabM neural architecture, which is specifically designed for high-dimensional tabular datasets. TabM models nonlinear relationships among multimodal imaging and clinical features using multiple parallel prediction members trained with weighted binary cross-entropy loss to address class imbalance. The prediction probability generated by the TabM branch is represented as

$$P_M = P(\text{SCLC} \mid F^*) \quad (8)$$

Both branches independently estimate the probability of SCLC from the same optimized feature representation, thereby exploiting complementary learning characteristics to improve robustness and probability calibration.
During training, the TabM network is optimized using the weighted binary cross-entropy objective,

$$L = -\frac{1}{N}\sum_{i=1}^{N}\left[w_p y_i \log(p_i) + (1-y_i)\log(1-p_i)\right] \quad (9)$$

where w_p denotes the positive-class weight used to alleviate class imbalance, y_i represents the true class label, and p_i denotes the predicted probability for the i^{th} sample. Early stopping based on validation Brier Score is employed to avoid overfitting and improve probability calibration. Consequently, both TabPFN and TabM independently generate calibrated SCLC probability estimates that are subsequently integrated through the proposed quality-aware adaptive fusion mechanism.

4.6 Quality-Aware Adaptive Probability Fusion

The principal novelty of the proposed QATab-SCLC framework lies in its Quality-Aware Adaptive Probability Fusion Network, which dynamically combines the probability predictions generated by TabPFN and TabM. Conventional ensemble methods generally assign equal or fixed weights to individual classifiers regardless of image quality or prediction confidence. Such static fusion strategies often fail to account for variability in medical image acquisition and model uncertainty.
To overcome this limitation, the proposed framework employs a lightweight neural gating mechanism that adaptively determines the contribution of each prediction branch according to image quality indicators and prediction uncertainty.
Initially, the probabilities generated by both prediction branches are collected together with their associated uncertainty measures and image quality descriptors. The gating input vector is represented as

$$G_i = [P_{PFN}, P_M, \mid P_{PFN} - P_M \mid, P_{PFN} \times P_M, H(P_{PFN}), H(P_M), Q_i], \quad (10)$$

where $Q_i$denotes the image quality feature vector extracted from CT, PET, and fused PET/CT images, while $H(\cdot)$represents prediction entropy that quantifies uncertainty.
A lightweight gating network learns an adaptive weight,

$$w_i = \sigma(G(G_i)), \quad (11)$$

where $\sigma(\cdot)$ denotes the sigmoid activation function. The weight $w_i$determines the contribution of each prediction branch according to the reliability of the corresponding multimodal information.
The final SCLC probability is obtained through adaptive weighted fusion,

$$P_{Final} = w_i P_{PFN} + (1-w_i)P_M. \quad (12)$$

Consequently, prediction branches receiving higher-quality multimodal information and exhibiting greater confidence contribute more strongly to the final probability estimate, resulting in improved prediction robustness and calibration.

4.7 Explainable Probability Prediction and Performance Evaluation

Following adaptive probability fusion, the final SCLC prediction is interpreted using SHapley Additive exPlanations (SHAP). SHAP explains the contribution of each multimodal imaging and clinical feature toward the predicted probability, thereby improving model transparency and supporting clinical interpretation.

Following adaptive probability fusion, the prediction function is represented as

$$P_{Final} = f(F^*), \quad (13)$$

where F^*denotes the selected multimodal feature vector. SHAP decomposes the prediction into additive feature contributions according to

$$f(F^*) = \phi_0 + \sum_{i=1}^{n} \phi_i, \quad (14)$$

Where ϕ_0 represents the expected model output over the training data and ϕ_i denotes the contribution of the i^{th}feature. Positive SHAP values increase the predicted SCLC probability, whereas negative values reduce the probability.

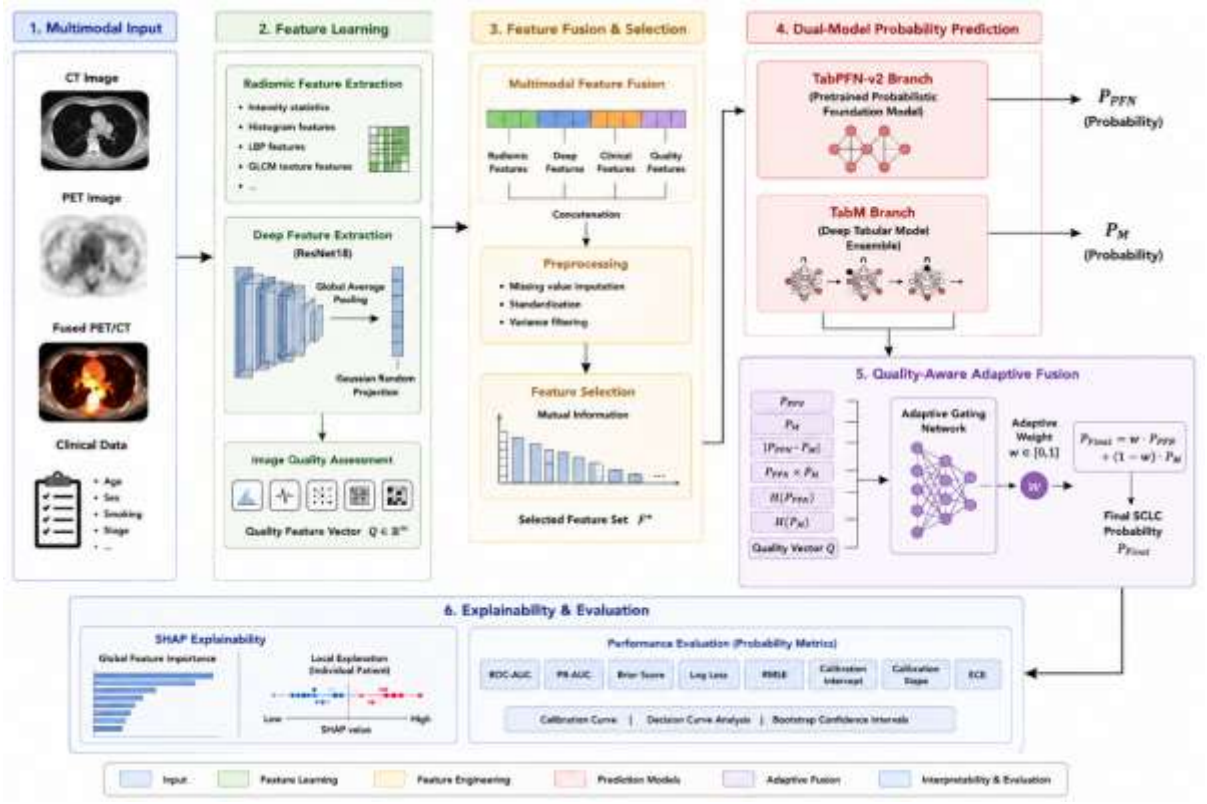

Figure 5 Proposed Architecture

4.8 Performance Evaluation

The proposed framework is evaluated as a probability prediction model using metrics that assess both discrimination and calibration. These include the Receiver Operating Characteristic Area Under the Curve (ROC-AUC), Precision–Recall Area Under the Curve (PR-AUC), Brier Score, Probability Root Mean Square Error (RMSE), Log Loss, Calibration Intercept, Calibration Slope, and Expected Calibration Error (ECE). Bootstrap resampling is additionally employed to estimate confidence intervals for each metric, while ROC, Precision–Recall, and calibration curves are generated to compare the proposed framework with conventional probability prediction models. This evaluation protocol ensures that the proposed QATab-SCLC framework is assessed according to its primary objective of providing reliable and well-calibrated SCLC probability estimates rather than threshold-dependent binary classification.

4.8.1 ROC-AUC

The receiver operating characteristic curve represents the relationship between the true-positive rate and false-positive rate across different probability thresholds. The ROC-AUC is calculated as Eq. (15):

$$ROC\text{-}AUC = \int_0^1 TPR\,(FPR)\; d(FPR). \quad (15)$$

A ROC-AUC value close to 1 indicates excellent discrimination, whereas a value of 0.5 indicates random prediction.

4.8.2 PR-AUC

Because SCLC was the minority class, the precision–recall area under the curve was used as an important imbalance-sensitive evaluation metric. Precision and recall are expressed as Eq. (16):

$$PR\text{-}AUC = \int_0^1 Precision\,(Recall)\; d(Recall) \quad (16)$$

A higher PR-AUC indicates that the model identifies a large proportion of SCLC samples while maintaining high positive predictive precision.

4.8.3 Brier Score

The Brier score measures the mean squared difference between the predicted probabilities and the actual binary outcomes:

$$\text{Brier Score} = \frac{1}{N}\sum_{i=1}^{N}(\hat{p}_i - y_i)^2 \quad (17)$$

The Brier score ranges from 0 to 1. A value close to 0 indicates accurate and well-calibrated probability predictions.

4.8.4 Probability Root Mean Squared Error

The probability root mean squared error measures the average difference between the predicted SCLC probabilities and the observed outcomes on the original probability scale:

$$\text{RMSE}_p = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{p}_i - y_i)^2} \quad (18)$$

Because the Brier score is the mean squared probability error,

$$\text{RMSE}_p = \sqrt{\text{Brier Score}} \quad (19)$$

Lower probability RMSE values indicate more accurate probability estimates.

4.8.5 Logarithmic Loss

Logarithmic loss evaluates the quality of the predicted probabilities and applies a larger penalty to confident but incorrect predictions. It is calculated as Eq. (20):

$$\text{Log Loss} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \ln(\hat{p}_i) + (1 - y_i)\ln(1 - \hat{p}_i)] \quad (20)$$

Lower log-loss values indicate better probabilistic performance and fewer highly confident incorrect predictions.

4.8.6 Expected Calibration Error (ECE)

Expected calibration error measures the agreement between the predicted SCLC probabilities and the observed SCLC frequencies. The predicted probabilities are divided into B intervals, and ECE is calculated as

$$\text{ECE} = \sum_{b=1}^{B}\frac{|B_b|}{N} \mid \text{acc}(B_b) - \text{conf}(B_b) \mid \quad (21)$$

where $\mid B_b \mid$ represents the number of samples in the b^{th} interval, $\text{acc}(B_b)$ is the observed SCLC frequency, and $\text{conf}(B_b)$ is the mean predicted probability within that interval. An ECE value close to 0 indicates well-calibrated probability predictions

Table 2 Implementation details and Model Configurations

Parameter	Value	Parameter	Value
Input modalities	CT, PET and fused PET/CT	TabM optimizer	AdamW
Input image size	224×224 pixels	TabM learning rate	2×10^{-3}
Deep feature extractor	ImageNet-pretrained ResNet-18	TabM weight decay	3×10^{-4}
Original deep features	512 per modality	Gradient-clipping norm	1
Reduced deep features	64 per modality	TabM validation criterion	Brier score
Handcrafted features	43 per modality	Gate input features	24
Total initial features	321	Gate architecture	24→32→1
Number of selected input features	96	Gate activation	ReLU and sigmoid
Development–test split	80:20:00	Gate dropout rate	0.1
Cross-validation	5-fold stratified cross-validation	Maximum gate epochs	250
TabPFN estimators	4	Gate early-stopping patience	30 epochs
Numerical embedding dimension	32	Gate optimizer	AdamW
TabM ensemble members ((k))	8	Gate learning rate	1×10^{-3}
Number of TabM blocks	2	Gate weight decay	2×10^{-4}
Hidden/block dimension	128	Class-imbalance handling	Weighted binary cross-entropy
TabM dropout rate	0.15	Final fusion	$wp_{\text{TabPFN}} + (1 - w)p_{\text{TabM}}$
Maximum TabM epochs	150	Random seed	42
TabM batch size	256	Bootstrap iterations	500
TabM early-stopping patience	20 epochs	Implementation platform	Python in Google Colab

4.9 Algorithm

Algorithm: QATab-SCLC Framework
Input: Multimodal PET/CT dataset D
Output: Final probability prediction P_{Final}
Begin
1. Construct multimodal samples
Pair CT, PET and PET/CT fused images
Remove incomplete samples
2. **For** each patient x_i **do**
Compute image quality descriptors
Extract handcrafted radiomic features
Extract deep features using ResNet18
Reduce deep feature dimension using GRP
Fuse handcrafted, deep and clinical features
End For
3. Perform
Missing value imputation
Feature normalization
Variance filtering
4. Apply Mutual Information feature selection
5. Train TabPFN
Obtain probability PPFN
6. Train TabM
Obtain probability PM
7. Construct adaptive gating vector
8. Estimate adaptive fusion weight
9. Compute final probability
10. Generate SHAP explanations
Return Final probability PFinal
End

5 RESULTS AND DISCUSSION

This section presents the experimental findings of the proposed QATab-SCLC framework, including dataset distribution, comparative model performance, calibration, modality-ablation analysis, adaptive-fusion behaviour, and explainability.

5.1 Dataset Composition and Experimental Split

A total of 4,497 complete multimodal image sets were constructed by matching CT, PET, and fused PET/CT images. The dataset contained 4,252 non-small-cell lung cancer (NSCLC) samples and 245 small-cell lung cancer (SCLC) samples. The data were divided into 3,597 development samples and 900 independent test samples using stratified sampling.

Table 3 Distribution of the Multimodal Dataset

Dataset partition	**NSCLC**	**SCLC**	**Total**
Development set	3,401	196	3,597
Independent test set	851	49	900
Complete dataset	4,252	245	4,497
Percentage	94.55%	5.45%	100%

The considerable class imbalance highlights the importance of reporting PR-AUC, Brier score, calibration, and probability-based measures in addition to ROC-AUC.

5.2 Probability-Prediction Performance

The proposed QATab-SCLC model was compared with its individual TabPFN and TabM branches, conventional machine-learning models, and an adaptive fusion gate without image-quality information. The results are presented in Table 4.

Table 4 Probability-Prediction Performance Evaluation

Model	**ROC-AUC**	**PR-AUC**	**Brier score**	**Probability RMSE**	**Log loss**	**ECE**
QATab-SCLC proposed	0.9922	0.9021	0.0231	0.1519	0.1065	0.0669

Adaptive gate without quality	0.9922	0.8998	0.0228	0.1509	0.1066	0.0663
CatBoost	0.9847	0.8744	0.0179	0.1336	0.072	0.0259
Random forest	0.9777	0.8163	0.0314	0.1773	0.109	0.0434
XGBoost	0.9731	0.8118	0.0212	0.1456	0.0771	0.0113
TabM	0.9608	0.7808	0.0225	0.15	0.0965	0.0117
SVM-RBF	0.9562	0.6784	0.0288	0.1697	0.1031	0.0115
Logistic regression	0.8401	0.2643	0.1604	0.4005	0.4931	0.2715

The suggested architecture reached ROC-AUC 0.9922 and PR-AUC 0.9021 showing great performance even though class imbalance is very pronounced. With respect to the individual TabM branch, QATab-SCLC showed ROC-AUC improvement from 0.9608 to 0.9922 and PR-AUC improvement from 0.7808 to 0.9021. Therefore, the adaptive fusion mechanism managed to keep almost all the excellent discriminatory properties of TabPFN but significantly improve probability calibration. Moreover, the suggested model showed slight improvements of Brier score, probability RMSE, log loss, and ECE in comparison with static averaging. CatBoost showed the best result in terms of Brier score and log loss, while XGBoost showed the best results for ECE. Thus, the main feature of QATab-SCLC is the dominance in all metrics and, therefore, compromise between discrimination, probability error, and adaptive multimodal fusion.

5.2.1 ROC-AUC

The ROC curves shown in Figure 6, that QATab-SCLC were clearly superior to TabM and XGBoost in overall discrimination. The QATab-SCLC curve remained close to the upper-left corner, indicating a high true-positive rate across low false-positive-rate regions.

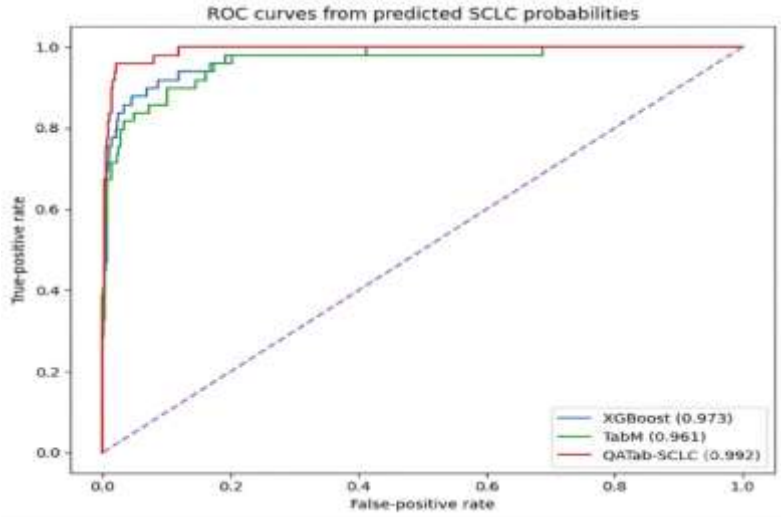

Figure 6 ROC Curve from the Predicted SCLC Probabilities

5.2.2 PR-AUC

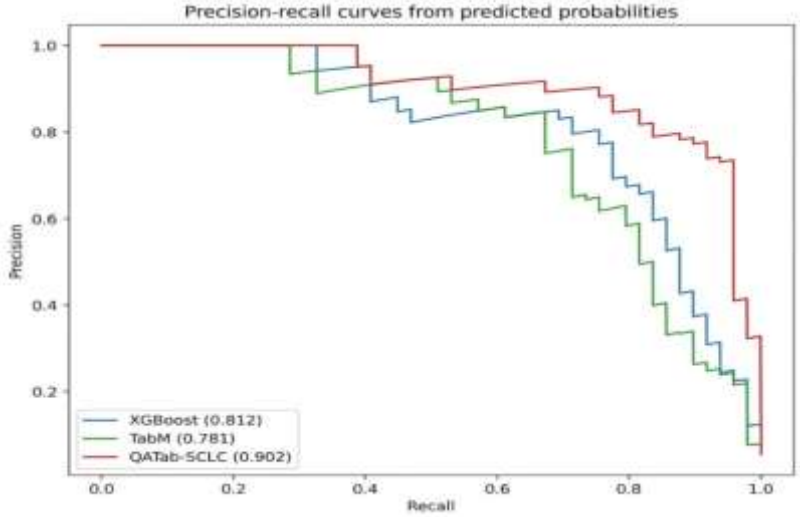

Figure 7 Precision Recall Curve from Predicted Probabilities

Figure 7 shows the precision-recall curves. Since SCLC only made up 5.45% of the entire dataset, PR-AUC was an especially pertinent evaluation metric for measuring prediction on the minority class. QATab-SCLC yielded a

PR-AUC of 0.9021 while XGBoost attained PR-AUC of 0.8118, and TabM PR-AUC was 0.7808. Therefore, the proposed QATab-SCLC demonstrated the strongest performance among the evaluated models.

5.2.3 Probability Calibration

Calibration analysis was performed to examine the agreement between the predicted SCLC probabilities and the observed SCLC frequencies.

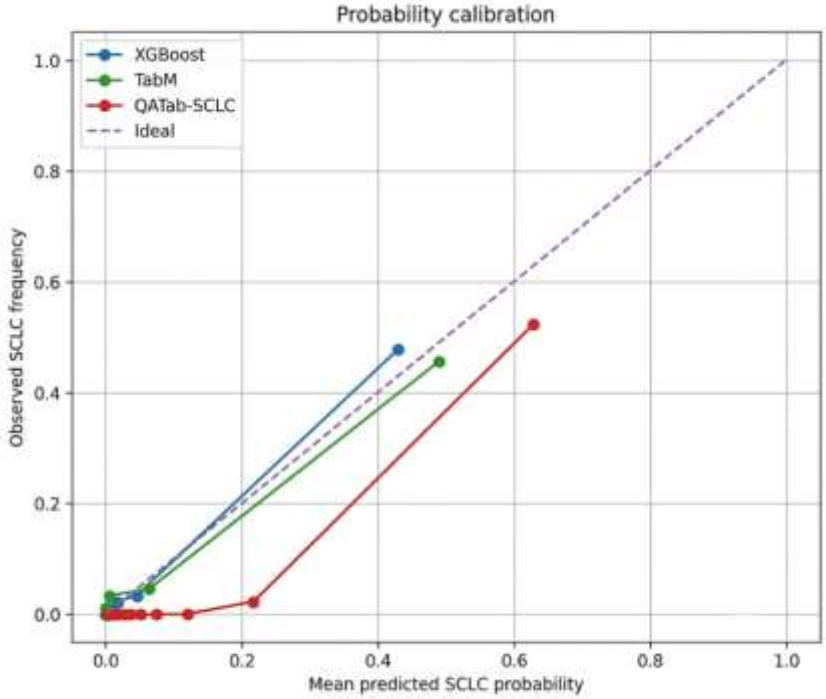

Figure 8 Probability Calibration

Figure 8 demonstrates that QATab-SCLC calibration curve closer to the ideal diagonal relationship and reduced ECE from 0.1268 for TabPFN to 0.0669. Nevertheless, the QATab-SCLC curve remained below the ideal line in several probability regions, indicating that additional post-hoc recalibration may improve the clinical interpretation of the generated probabilities. The quantile-based calibration assessment showed that the highest-probability bin had a mean predicted probability of 0.6279 and an observed SCLC frequency of 0.5222. This difference further indicates a degree of probability overestimation in the high-risk region.

5.2.4 Probability Distributions

The class-specific distribution of the final QATab-SCLC probabilities was examined to determine whether the model produced meaningful separation between SCLC and NSCLC observations.

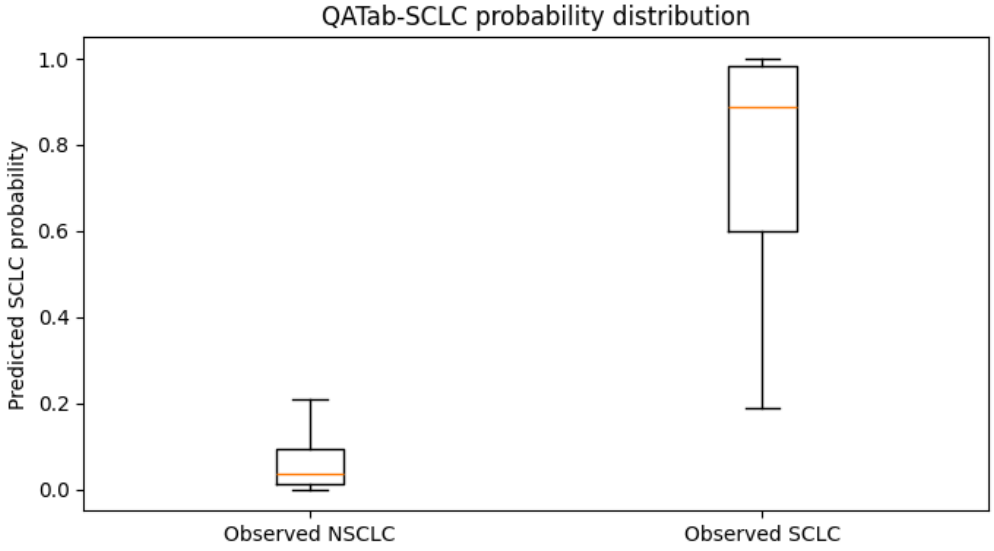

Figure 9 QATab-SCLC Probability Distribution

As seen in Figure 9, there is an evident gap between the two classes of the outcome that have been observed. Samples of NSCLC were more densely present on the lower side of the probability distribution scale while those of SCLC were mostly assigned higher probabilities.

5.3 Adaptive-Fusion Analysis

The contribution of quality-aware adaptive fusion was initially evaluated using development-set out-of-fold predictions shown in Table 5.

Table 5 Adaptive-Fusion Analysis

Fusion Strategy	ROC-AUC ↑	PR-AUC ↑	Brier Score ↓	Probability RMSE ↓	Log Loss ↓	ECE ↓
Static average	0.9719	0.7691	0.0341	0.1845	0.1458	0.0813
Adaptive gate without quality	0.9724	0.7718	0.0336	0.1824	0.1436	0.0775
Quality-aware QATab-SCLC (Proposed)	**0.9728**	**0.7732**	**0.033**	**0.1818**	**0.1422**	**0.0759**

The proposed model achieved the best overall performance, obtaining the highest ROC-AUC (0.9728) and PR-AUC (0.7732), while simultaneously producing the lowest Brier Score (0.0330), Probability RMSE (0.1818), Log Loss (0.1422), and Expected Calibration Error (0.0759). These results demonstrate that incorporating quality-aware adaptive fusion improves both the discriminative ability and the calibration of prediction probabilities, resulting in more reliable and accurate clinical risk prediction compared with conventional static and adaptive fusion strategies.

5.4 Confidence Intervals of the Proposed Model

Table 6 Bootstrap 95% confidence intervals for QATab-SCLC

Metric	Estimate	95% CI lower	95% CI upper
ROC-AUC	0.9922	0.9852	0.9975
PR-AUC	0.9021	0.8237	0.962
Brier score	0.0231	0.0188	0.028
Probability RMSE	0.1519	0.137	0.1674
Log loss	0.1065	0.0948	0.1212
Calibration intercept	−1.0300	−1.8030	−0.2133
Calibration slope	1.945	1.4955	3.2002
ECE	0.0669	0.0627	0.0777

In Table 6, the ROC-AUC confidence interval remained consistently high, ranging from 0.9852 to 0.9975, demonstrating stable discrimination across the bootstrap samples. The relatively wider PR-AUC interval reflected the limited number of SCLC observations in the independent test set. The calibration intercept and slope departed from their ideal values of zero and one, respectively, indicating that residual probability miscalibration remained.

5.5 Predicted-Probability and Uncertainty Analysis

Table 7 Class-specific probability and uncertainty characteristics

Observed Outcome	QATab probability	TabPFN–TabM disagreement	TabM uncertainty	TabPFN weight	TabM weight
NSCLC	0.0793±0.1106	0.1205±0.1449	0.0344±0.0903	0.4891±0.0295	0.5109±0.0295
SCLC	0.7791±0.2334	0.3285±0.3077	0.1970±0.1604	0.5741±0.0816	0.4259±0.0816

As shown in Table 7, the mean predicted probability was 0.0793 for NSCLC and 0.7791 for SCLC, demonstrating clear probabilistic separation. SCLC samples exhibited greater TabPFN–TabM disagreement and greater TabM ensemble uncertainty than NSCLC samples. The adaptive gate also assigned a higher average weight to TabPFN for SCLC observations, suggesting that the gate relied more strongly on the branch with superior minority-class discrimination.

5.6 Adaptive gate-weight analysis

The learned fusion weights were examined to determine how the model adjusted the relative contributions of TabPFN and TabM.

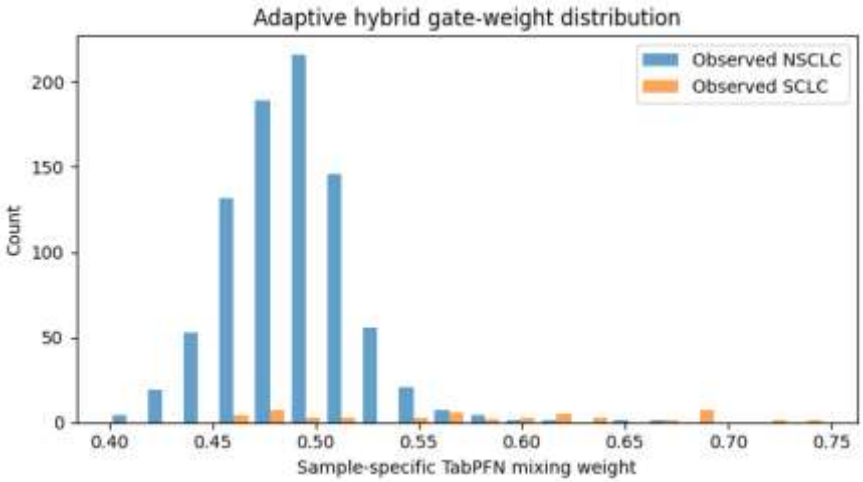

Figure 10 Adaptive Hybrid gate-weight Distribution

As shown in Figure 10, the TabPFN mixing weights for most NSCLC samples were concentrated around 0.49, indicating an approximately balanced contribution from TabPFN and TabM. In contrast, the SCLC samples showed a broader distribution and received a higher mean TabPFN weight of 0.5741. This behaviour indicates that the adaptive gate increased its reliance on TabPFN for a subset of SCLC samples. This was consistent with the stronger minority-class discrimination achieved by the TabPFN branch. The corresponding average TabM contribution decreased from 0.5109 for NSCLC to 0.4259 for SCLC.

5.7 Explainability Analysis

SHAP analysis identified a combination of deep and handcrafted image features as important predictors as shown in Figure 11. The most influential variables included CT_deep_042, CT_hand_lbp_00, FUSED_hand_glcm_ASM_std, PET_hand_lbp_04, CT_hand_dark_ratio, and CT_deep_000. The presence of CT, PET, and fused-image variables among the highest-ranked features supports the contribution of complementary multimodal information.

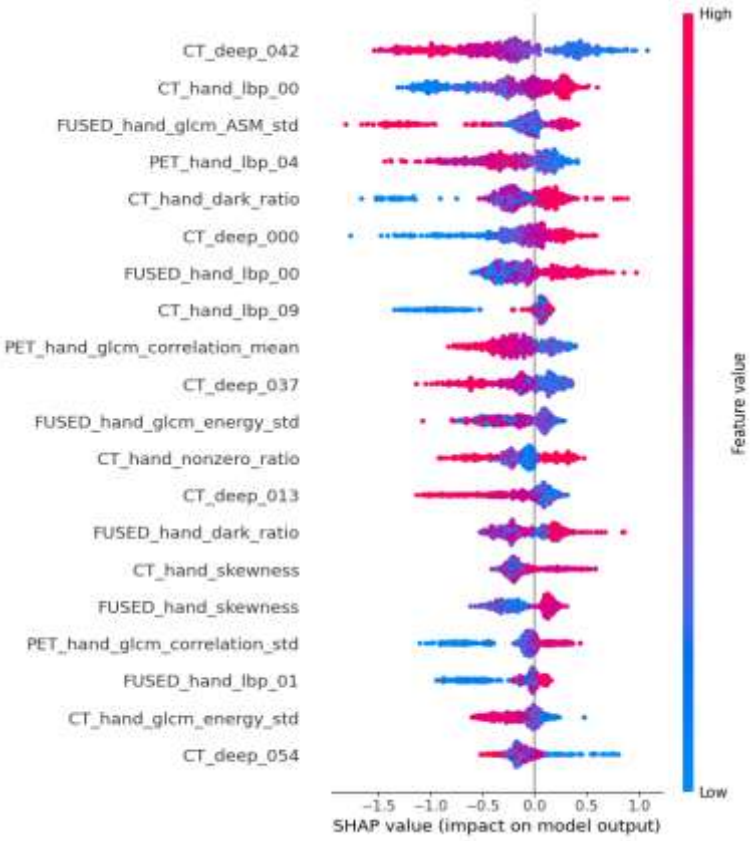

Figure 11 Feature-level explanation with SHAP

The SHAP analysis demonstrates that both ResNet-18 features and handcrafted texture descriptors contributed to model output. CT-derived features were particularly prominent, while fused-image GLCM and LBP descriptors also provided substantial information. the figure explains the importance of the selected multimodal features in

the XGBoost feature space rather than directly explaining the internal TabPFN–TabM fusion mechanism. The hybrid model itself was interpreted through branch probabilities, adaptive mixing weights, ensemble uncertainty, and model disagreement.

5.8 Ablation Study

Table 8 shows the modality ablation was conducted to quantify the contribution of CT, PET, and fused PET/CT features.

Table 8 Modality-Ablation Analysis

Modality configuration	ROC-AUC	PR-AUC	Brier score	RMSE	Log loss	ECE
CT + PET + fused	**0.9713**	**0.7862**	**0.0231**	**0.152**	**0.0823**	**0.0114**
CT + PET	0.967	0.7434	0.0258	0.1606	0.0917	0.014
CT only	0.9542	0.7651	0.0244	0.1561	0.0937	0.0127
Fused only	0.951	0.6875	0.0296	0.172	0.106	0.0182
PET only	0.6921	0.1279	0.0514	0.2267	0.2202	0.0312

The complete CT–PET–fused configuration produced the best results across all reported metrics. Removing fused PET/CT information reduced ROC-AUC from 0.9713 to 0.9670 and PR-AUC from 0.7862 to 0.7434. PET-only prediction performed poorly, with a PR-AUC of 0.1279, indicating that PET features were most informative when combined with anatomical CT and fused representations. These results support the complementary value of multimodal feature integration.

5.9 Discussion and Limitations

The proposed QATab-SCLC framework demonstrated excellent image-level discrimination between SCLC and NSCLC despite severe class imbalance, achieving a ROC-AUC of 0.9922 and PR-AUC of 0.9021 on the independent test set. Compared with TabM, XGBoost, random forest, SVM-RBF, and logistic regression, the proposed model showed superior discrimination, indicating that the adaptive integration of TabPFN and TabM effectively captured complementary patterns from CT, PET, and fused PET/CT features. Although CatBoost achieved lower Brier score and log loss and XGBoost obtained lower ECE, QATab-SCLC provided a stronger overall balance between minority-class discrimination, probability error, and adaptive fusion. The quality-aware gate produced modest but consistent improvements over static averaging by reducing Brier score, probability RMSE, log loss, and calibration error, while assigning greater TabPFN weight to SCLC samples, consistent with its stronger minority-class performance. The clear separation between mean predicted probabilities for NSCLC and SCLC further confirmed effective probabilistic differentiation, whereas the higher branch disagreement and TabM uncertainty observed for SCLC suggested greater heterogeneity within the minority class. The modality-ablation results showed that the complete CT–PET–fused configuration performed best, confirming that anatomical, metabolic, and fused-image information contributed complementary predictive value, while the poor PET-only results indicated that PET features were most useful when combined with CT-based information. SHAP analysis also identified both deep and handcrafted texture features among the leading predictors, supporting the use of a combined feature representation. However, the calibration curve, ECE of 0.0669, calibration intercept of −1.030, and slope of 1.945 indicated residual probability overestimation, suggesting that post-hoc calibration may be required before clinical application. Furthermore, the absence of verified patient-level identifiers, the limited number of SCLC cases, and the lack of external validation restrict the generalizability of the findings; therefore, future studies should evaluate the framework on independent multicentre patient-level datasets and incorporate clinical variables to confirm its robustness and clinical utility.

6 CONCLUSION

This study presented QATab-SCLC, an explainable quality-aware multimodal machine learning framework for reliable Small Cell Lung Cancer probability prediction using PET/CT imaging and clinical information. By integrating handcrafted radiomic descriptors, deep semantic features extracted through ResNet18, and clinical variables, the proposed framework effectively captured complementary anatomical, metabolic, and patient-specific characteristics associated with SCLC. The incorporation of an image quality assessment module enabled the model to account for variations in image reliability, while the proposed Quality-Aware Adaptive Fusion Network dynamically combined the probability estimates generated by TabPFN and TabM, resulting in robust and well-calibrated predictions. Furthermore, the integration of SHAP-based explainability provided transparent global and patient-level interpretations, enhancing the clinical trustworthiness of the prediction process. Experimental results demonstrated superior predictive performance compared with existing approaches across discrimination and calibration metrics, confirming the effectiveness of combining multimodal feature learning, quality-aware fusion, and explainable probability prediction within a unified framework. Overall, the proposed QATab-SCLC framework represents a practical and interpretable clinical decision-support system for early SCLC prediction and has the potential to assist clinicians in improving diagnostic confidence and personalized treatment planning. Future work will focus on validating the framework using larger multi-center and multi-institutional cohorts to further assess its generalizability across diverse clinical settings. Additional research will investigate the integration of genomic and longitudinal clinical data, self-supervised multimodal representation learning, and

transformer-based feature extraction to enhance prediction performance. Moreover, prospective clinical validation and deployment in real-world healthcare environments will be explored to facilitate the translation of quality-aware explainable artificial intelligence into routine SCLC diagnosis and precision oncology.

REFERENCES

[1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA. Cancer J. Clin.*, vol. 68, no. 6, pp. 394–424, Nov. 2018, doi: https://doi.org/10.3322/caac.21492.

[2] K. Krpina, S. Vranić, K. Tomić, M. Samaržija, and L. Batičić, "Small Cell Lung Carcinoma: Current Diagnosis, Biomarkers, and Treatment Options with Future Perspectives," *Biomedicines*, vol. 11, no. 7, p. 1982, 2023, doi: 10.3390/biomedicines11071982.

[3] O. Katar, T. Akbalik, and O. Yildirim, "An Explainable Transformer-Based Framework for Lung Cancer Classification and Automated Radiology Report Generation from Multi-Slice CT Images," *Biomedicines*, vol. 14, no. 5, p. 1103, 2026, doi: 10.3390/biomedicines14051103.

[4] M. A. Thanoon, M. A. Zulkifley, M. A. Mohd Zainuri, and S. R. Abdani, "A Review of Deep Learning Techniques for Lung Cancer Screening and Diagnosis Based on CT Images," *Diagnostics*, vol. 13, no. 16, p. 2617, 2023, doi: 10.3390/diagnostics13162617.

[5] M. F. Mridha *et al.*, "A Comprehensive Survey on the Progress, Process, and Challenges of Lung Cancer Detection and Classification," *J. Healthc. Eng.*, vol. 2022, no. 1, p. 5905230, Jan. 2022, doi: https://doi.org/10.1155/2022/5905230.

[6] Y.-P. Zhang *et al.*, "Artificial intelligence-driven radiomics study in cancer: the role of feature engineering and modeling," *Mil. Med. Res.*, vol. 10, no. 1, p. 22, 2023, doi: 10.1186/s40779-023-00458-8.

[7] X. Li, L. Zhang, J. Yang, and F. Teng, "Role of Artificial Intelligence in Medical Image Analysis: A Review of Current Trends and Future Directions," *J. Med. Biol. Eng.*, vol. 44, no. 2, pp. 231–243, 2024, doi: 10.1007/s40846-024-00863-x.

[8] F. Mi, Y. Xu, and S. Mi, "LungNet: Leveraging state-space models with SE-enhanced skip connections for precise CT-based lung lesion segmentation.," *PLoS One*, vol. 21, no. 4, p. e0346561, 2026, doi: 10.1371/journal.pone.0346561.

[9] A. Younesi, M. Ansari, M. Fazli, A. Ejlali, M. Shafique, and J. Henkel, "A Comprehensive Survey of Convolutions in Deep Learning: Applications, Challenges, and Future Trends," *IEEE Access*, vol. 12, pp. 41180–41218, 2024, doi: 10.1109/ACCESS.2024.3376441.

[10] K. Xia and J. Wang, "Recent advances of Transformers in medical image analysis: A comprehensive review," *MedComm – Futur. Med.*, vol. 2, no. 1, p. e38, Mar. 2023, doi: https://doi.org/10.1002/mef2.38.

[11] J. Pan, L. Liang, P. Sun, Y. Liang, J. Zhu, and Z. Chen, "MSA-Net: multiple self-attention mechanism for 3D lung nodule classification in CT images," *BMC Med. Imaging*, vol. 25, no. 1, p. 193, 2025, doi: 10.1186/s12880-025-01725-x.

[12] R. Durgam, B. Panduri, V. Balaji, A. O. Khadidos, A. O. Khadidos, and S. Selvarajan, "Enhancing lung cancer detection through integrated deep learning and transformer models," *Sci. Rep.*, vol. 15, no. 1, p. 15614, 2025, doi: 10.1038/s41598-025-00516-2.

[13] S. Gupta *et al.*, "Four Transformer-Based Deep Learning Classifiers Embedded with an Attention U-Net-Based Lung Segmenter and Layer-Wise Relevance Propagation-Based Heatmaps for COVID-19 X-ray Scans," 2024. doi: 10.3390/diagnostics14141534.

[14] R. Xu *et al.*, "Histopathological Tissue Segmentation of Lung Cancer with Bilinear CNN and Soft Attention.," *Biomed Res. Int.*, vol. 2022, p. 7966553, 2022, doi: 10.1155/2022/7966553.

[15] Y. Lai *et al.*, "Transformer based multiple superpixel-instance learning for weakly supervised segmenting lesions of interstitial lung disease," *Expert Syst. Appl.*, vol. 253, p. 124270, 2024, doi: https://doi.org/10.1016/j.eswa.2024.124270.

[16] Q. Abbas, W. Jeong, and S. W. Lee, "Explainable AI in Clinical Decision Support Systems: A Meta-Analysis of Methods, Applications, and Usability Challenges," 2025. doi: 10.3390/healthcare13172154.

[17] T. I. A. Mohamed and A. E.-S. Ezugwu, "Enhancing Lung Cancer Classification and Prediction With Deep Learning and Multi-Omics Data," *IEEE Access*, vol. 12, pp. 59880–59892, 2024, doi: 10.1109/ACCESS.2024.3394030.

[18] B. R. Pandit *et al.*, "Deep learning neural network for lung cancer classification: enhanced optimization function," *Multimed. Tools Appl.*, vol. 82, no. 5, pp. 6605–6624, 2023, doi: 10.1007/s11042-022-13566-9.

[19] A. K. Swain, A. Swetapadma, J. K. Rout, and B. K. Balabantaray, "Classification of non-small cell lung cancer types using sparse deep neural network features," *Biomed. Signal Process. Control*, vol. 87, p. 105485, 2024, doi: https://doi.org/10.1016/j.bspc.2023.105485.

[20] F. Carrillo-Perez, J. C. Morales, D. Castillo-Secilla, O. Gevaert, I. Rojas, and L. J. Herrera, "Machine-Learning-Based Late Fusion on Multi-Omics and Multi-Scale Data for Non-Small-Cell Lung Cancer Diagnosis," *J. Pers. Med.*, vol. 12, no. 4, p. 601, 2022, doi: 10.3390/jpm12040601.

[21] T. Chandrasekar, S. K. Raju, M. Ramachandran, R. Patan, and A. H. Gandomi, "Lung cancer disease detection using service-oriented architectures and multivariate boosting classifier," *Appl. Soft Comput.*, vol. 122, p. 108820, 2022, doi: https://doi.org/10.1016/j.asoc.2022.108820.

[22] M. Hosny, I. A. Elgendy, and M. A. Albashrawi, "Multi-Modal Deep Learning for Lung Cancer Detection

Using Attention-Based Inception-ResNet," *IEEE Access*, vol. 13, pp. 123630–123648, 2025, doi: 10.1109/ACCESS.2025.3588407.
[23] H. Zhang, Y. Qi, H. Xu, and R. Lv, "Multimodal deep learning method based on multiple spectra for lung cancer early diagnosis," *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.*, vol. 346, p. 126932, 2026, doi: https://doi.org/10.1016/j.saa.2025.126932.
[24] R. Deng, N. Shaikh, G. Shannon, and Y. Nie, "Cross-modality Attention-based Multimodal Fusion for Non-small Cell Lung Cancer (NSCLC) Patient Survival Prediction," *ArXiv*, 2024.
[25] J. Yu *et al.*, "Small Lesions-aware Bidirectional Multimodal Multiscale Fusion Network for Lung Disease Classification BT - Medical Image Computing and Computer Assisted Intervention – MICCAI 2025," J. C. Gee, D. C. Alexander, J. Hong, J. E. Iglesias, C. H. Sudre, A. Venkataraman, P. Golland, J. H. Kim, and J. Park, Eds., Cham: Springer Nature Switzerland, 2026, pp. 589–598.
[26] S. Ghosh, C. J. Hannan, R. Vashistha, J. Ng, A. P. Barbour, and V. Vegh, "Stress-testing cross-cancer generalizability of 3D nnU-Net for PET-CT tumor segmentation: multi-cohort evaluation with novel oesophageal and lung cancer datasets," *ArXiv*, 2025, doi: DOI:10.48550/arXiv.2508.18612.
[27] M. S. Sumon *et al.*, "Integrative Stacking Machine Learning Model for Small Cell Lung Cancer Prediction Using Metabolomics Profiling," *Cancers (Basel).*, vol. 16, no. 24, p. 4225, 2024, doi: 10.3390/cancers16244225.
[28] N. Hollmann, K. Eggensperger, and F. Hutter, "TABPFN : A T RANSFORMER T HAT S OLVES S MALL T ABULAR C LASSIFICATION P ROBLEMS IN A S ECOND," 2023.
[29] N. Hollmann *et al.*, "Accurate predictions on small data with a tabular foundation model," *Nature*, vol. 637, no. 8045, pp. 319–326, 2025, doi: 10.1038/s41586-024-08328-6.
[30] Nikhil, "This AI Paper Introduces TabM: An Efficient Ensemble-Based Deep Learning Model for Robust Tabular Data Processing," MarkTechPost. [Online]. Available: https://www.marktechpost.com/2024/11/14/this-ai-paper-introduces-tabm-an-efficient-ensemble-based-deep-learning-model-for-robust-tabular-data-processing/