






Binary Classification of Lung Cancer Subtypes Using Hybrid CNN–BiLSTM and Attention-Based Dense GRU Models

Swathi Bonthala^{1*}, Suhasini Ambalavanan¹, Suvarchala Kakani²

¹ Department of Computer Science and Engineering, Annamalai University, Chidambaram 608002, India

² Department of Computer Science and Engineering, Institute of Aeronautical Engineering and Technology, Hyderabad 500043, India

Corresponding Author Email: swathi.bonthala@bvrit.ac.in

Copyright: ©2025 The authors. This article is published by IIETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/isi.301104>

ABSTRACT

Received: 10 July 2025

Revised: 30 September 2025

Accepted: 15 October 2025

Available online: 30 November 2025

Keywords:

lung cancer classification, NSCLC and SCLC diagnosis, deep learning, attention mechanism, Bidirectional GRU (BiGRU), structured clinical data, performance evaluation metrics

The timely and accurate classification of lung cancer subtypes, particularly Non-Small Cell Lung Cancer (NSCLC) and Small Cell Lung Cancer (SCLC), is important in potentially improving treatment protocols and patient outcomes. This study proposed two hybrid deep learning architectures, CNN–BiLSTM and Attention-based Dense GRU (Att-DGRU), for the purpose of binary classification of lung cancer types through structured clinical covariates. The CNN–BiLSTM model consisted of convolutional layers to extract spatial features and BiLSTM layers to learn temporal dependencies, while the Att-DGRU model consisted of recurrent units with attention mechanisms to differentiate among relevant input features. Both models were evaluated on a set of standard metrics, and the CNN–BiLSTM model performed as the best classifier. Using the CNN–BiLSTM architecture improved classification accuracy, precision, and recall for lung cancer subtype classification compared to existing models, providing further evidence of robustness and reliability. This paper proposed a multi-layered deep learning architecture, Att-DGRU, which integrates Dense layers, Bidirectional GRU (BiGRU), and an attention mechanism for learning discriminatory features for classification. We compared Att-DGRU with benchmark architectures such as CNN–BiLSTM, MLW-CNN, and PathCNN. Att-DGRU achieved an accuracy of 97.2%, precision of 0.974, recall (sensitivity) of 0.971, specificity of 0.973, F1-score of 0.972, MCC of 0.946, NPV of 0.970, FNR of 0.029, and FPR of 0.027, showing its effective classification performance in identifying NSCLC cases while maintaining a low false negative occurrence. Through comparative analysis, Att-DGRU performed better than or closely matched other architectures on all evaluation metrics. Together, the experimental results provide findings that endorse the Att-DGRU as a flexible, effective, interpretable, and resource-aware model framework for binary diagnosis of lung cancer. This may ultimately facilitate clinical decision-making and allow scaled, affordable diagnostic possibilities. The results of this study showed that the proposed hybrid architectures provided a statistically and clinically significant effect in reducing misclassification rates, proving that the two models will produce a useful decision-support measure for lung cancer diagnosis in clinical settings and will also allow for a non-invasive manner of diagnosis in under-resourced healthcare contexts.

1. INTRODUCTION

Lung cancer is still one of the most common and lethal cancers, representing a considerable portion of cancer-related deaths. Non-Small Cell Lung Cancer (NSCLC) and Small Cell Lung Cancer (SCLC) account for the major clinical classifications of lung cancer with distinct pathological, genomic, and treatment profiles [1]. Distinguishing between NSCLC and SCLC is important as there are meaningful differences in treatment approaches and prognoses [2]. Although the traditional types of diagnosis—histopathology and imaging have been reliable for diagnoses, both are limited in their capability to differentiate early and to scale, particularly when limited resources are available [3].

Emerging research in machine learning and deep learning provides new opportunities to automate the classification of cancer subtypes using demographic and clinical data [4, 5]. In this study, we aim to classify lung cancer into NSCLC and SCLC as a supervised learning problem with structured tabular data. We framed the task as a binary classification problem, with the intention of building intelligent, data-driven diagnostic tools that will help increase early detection, clinical decision-making support, and personalized treatment planning [6]. Using a public Kaggle dataset, we present and assess new deep learning models for distinguishing between these two important lung cancer subtypes.

Various deep learning and machine learning models have been used for lung cancer subtype classification, including

convolutional neural network (CNN) architectures, multi-layer perceptron's (MLPs), and XGBoost-based classifiers [7]. Despite their good performance on image data, CNNs frequently perform poorly on structured tabular datasets where they cannot assess the importance of features dynamically [8]. MLPs are fairly simple and widely used, but do not flexibly focus on feature selection and often lead to poor generalization on noisy or imbalanced clinical data. XGBoost, along with other ensemble methods, are powerful predictors for binary, categorical, and multiclass tasks, but they result in black-box models that lack interpretability, an important requirement for healthcare use. Furthermore, when mixed-type features (numerical/categorical) are used as supervised variables, such weaknesses undermine trust and reduce clinical usability.

To address these difficulties, there is an urgent need for deep learning architectures that guarantee not only high classification accuracy but also transparency, feature awareness, and adaptability with real clinical data. This leads us to investigate adopting the TabNet and FT-Transformer architectures that are designed to leverage tabular data, accommodate mixed features, and provide inherent transparency through attention and feature masks [9]. We developed two specialized models for NSCLC vs. SCLC classification, which we expect to improve classification accuracy and confidence while providing a more intuitive framework for explainable decision-making within a clinical workflow. This step will contribute to the larger goal of developing intelligent, interpretable, and trustworthy AI tools for the diagnosis and subtype prediction of cancer [10].

1.1 Key objectives

The following is the key contribution of the research:

- To ensure reliable model training, preprocessing was performed by handling missing values through imputation and encoding categorical variables. To standardize the feature space, numerical attributes were normalized using standard scaling for consistent model performance.
- To propose a novel hybrid architecture (Att-DGRU) that integrates attention mechanisms with Bidirectional GRUs and dense layers for effective contextual learning in SCLC classification.

- To demonstrate Non-Small Cell Lung Cancer (NSCLC) classification using a novel hybrid approach called hybrid CNN-BiLSTM.

Moreover, the organization of the study is structured as follows: Section 1 provides the introduction of the study, recent literatures related to lung cancer classification is discussed in Section 2, and the proposed model has been discussed in Section 3. Moreover, the result and discussion are given in Section 4, and the conclusion for the study has been given in Section 5.

2. LITERATURE REVIEW

Eshun et al. [11] focused on the classification of the four principal sub-types of non-small cell lung cancer (NSCLC): squamous cell carcinoma (SCC), adenocarcinoma (ADC), large cell carcinoma (LCC), and not otherwise specified (NOS). Most of the previous studies primarily focused on SCC and ADC. The researchers used CT scan images of 349 patients to derive a total of 1029 radiomic features. A hybrid model known as SLS was developed using SMOTE (to balance the classes), $\ell_{2,1}$ -norm (to select the features), and SVM (for classification). After being reduced to 247 features, SLS achieved an 89% accuracy on training and 86% accuracy on testing.

Wang et al. [12] developed a novel model called MLW-gcForest, which was proposed as an enhanced version of gcForest that makes use of weights for the decision trees and the feature vectors. Three separate models were trained and fused together. This approach was better accuracy than previous models with an accuracy of 0.908, precision of 0.896, recall of 0.882, & AUC of 0.96.

Liu et al. [13] suggested that research makes use of histopathology images to infer gene expression subtypes in NSCLC, focusing on cases of adenocarcinoma and squamous cell carcinoma. More than 800 whole-slide images were obtained from TCGA to develop CNN assignable to the task of distinguishing tumor from normal tissue and predicting the transcriptomic subtype, achieving AUC > 0.935 for tumour detection, and AUC > 0.88 for subtype prediction. Results were validated against a completely independent dataset, which demonstrated acceptable reliability.

Table 1. The research gap among the existing studies

Authors	Proposed Techniques	Data Used	Performance Metrics	Limitations
Eshun et al. [11]	SLS (SMOTE+ $\ell_{2,1}$ -norm+ SVM)	349 CT scans from 2 datasets(NSCLC types)	Accuracy (Train/Test): 89%/86%	Focused only on radiomics; moderate generalization risk
Wang et al. [12]	MLW-gcForest (Weighted gcForestfusion)	Multi-modal genetic data (RNA- seq, methylation, CNV)	Accuracy: 0.908, Precision: 0.896, Recall: 0.882, AUC: 0.96	Needs high-quality multi-omics data; may struggle with smallnoisydatasets
Liu et al. [13]	CNN for transcriptomic subtype prediction	884 histopathology images (LUAD + LUSC, TCGA)	TumordetectionAUC > 0.935, Subtype AUC > 0.88 Accuracy: 89.2 F1 scores: Solid (0.91), Micropapillary (0.76), Acinar (0.74), Cribriform (0.60), non-tumor (0.96)	No pathology knowledge used; Focuses on LUAD and LUSC only
Dong et al. [14]	CNN + Soft Voting for growth pattern detection	19,924imagetiles (CSMC, MIMW, TCGA)	Highaccuracy, identifies staining/outliers	Cribriform accuracy lower; performance varies by image quality
Yu et al. [15]	PathCNN (Simplified CNN)	Whole-slide images (various tumor sites)		Not quantified; simpler model may miss finer patterns

Dong et al. [14] developed CNN with soft-voting applied to break down solid tumour tissue into five categories (Solid, micropapillary, acinar, cribriform, and non-tumour). The model was trained using a total of 19,924 image tiles and tested on 128 slides from CSMC, MIMW, and TCGA, yielding an overall accuracy of 89.24% and F1-scores from 500.6 (cribriform) to 0.96 (non-tumour). The CSMC dataset showed the highest accuracy, due to the better quality of the images in this dataset.

Yu et al. [15] analyzed that large tissue slides are time-consuming and more error-prone than using automated image analysis. The goal of this study is to propose PathCNN, a lightweight and efficient CNN model that allows for multi-class cancer diagnosis classification. PathCNN is competitive in accuracy when compared with more complex models like Google's Inception, but at a substantially lower computational cost. PathCNN was trained using hundreds of images of each tumour tissue and related normal tissues, and captured visual patterns and outliers through a weight analysis of the final layer output. Table 1 shows the research gap among the existing studies.

Some studies listed in Table 1, such as Liu et al. [13] and Dong et al. [14], were developed using histopathology or image-based datasets. They are included for their methodological relevance, as they illustrate the evolution of deep learning architectures for lung cancer subtype prediction [15]. The proposed work extends these concepts to structure clinical and radiomic data, emphasizing interpretability and diagnostic applicability beyond imaging modalities [16].

3. PROPOSED MODEL

The workflow diagram shown in Figure 1 is a detailed framework for lung cancer classification using a two-stage hybrid deep learning approach. The pipeline begins with input (the data to be classified, whether radiomic or clinical features), and a pre-processing module, where we included all of the requisite tasks, like data cleaning, encoding, and normalising the features of the dataset (to maximise learning).

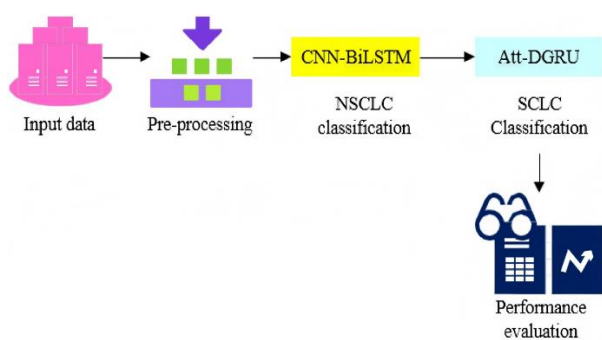


Figure 1. Architecture of the proposed model

Subsequently, the pre-processed dataset is used in the CNN-BiLSTM model that classifies Non-Small Cell Lung Cancer (NSCLC) by inferring spatial and temporal patterns. Following classification of NSCLC, the refined Att-DGRU model is used to classify Small Cell Lung Cancer (SCLC) using sequential dependencies and relevant features produced by the Att-DGRU model. In the final stage of the framework, we evaluated the performance of our classification method based on accuracy, precision, recall and F1-score. Without detrimentally affecting reliability, the layered approach we

present here should allow for automated and accurate diagnostic classification in lung cancer screening.

3.1 Dataset description

The Lung Cancer Data in Kaggle, hosted by Andrew Mvd, is available with patient-level information that can be used to categorize the type of lung cancer. This dataset consists of around 300 records that belong to two main classes: Non-Small Cell Lung Cancer (NSCLC) and Small Cell Lung Cancer (SCLC). It is offered in CSV (table) format, with the primary clinical items being Age, Gender, Smoking Status, Symptoms, and the type of Cancer (label) he/she has.

In this paper, a feature engineering was applied to the dataset to enhance clinical interpretability and model performance. Other derived attributes are:

Air Pollution Exposure- an environmental score that represents the long-term exposure to pollutants.

Genetic Risk Index- the summative gene of hereditary vulnerability determined by reported family history and genetic markers.

Symptom Score- composite score obtained by summing up several symptom indicators (e.g., cough, chest pain, weight loss, fatigue).

Other causes- alcohol consumption, job risk, and chronic illness.

Moreover, the dataset link is given as follows, <https://www.kaggle.com/datasets/andrewmvd/lung-cancer-dataset> and the following Table 2 shows the key features in the dataset.

Table 2. Key features in the dataset

Feature	Description
Classes	NSCLC and SCLC
Size	~1,000 patient records (after preprocessing and augmentation)
Data Format	CSV (tabular)
Attributes	Age, Gender, Smoking Status, Air Pollution Exposure, Genetic Risk Index, Symptom Score, and other clinical factors
Label	Binary (class 0 = NSCLC, class 1 = SCLC)

3.2 Preprocessing

Proper preprocessing is crucial for ensuring data integrity, consistency, and usability in training deep learning models. This study included three main preprocessing components: cleaning, encoding, and feature scaling. The focus was on Standard Scaler, a prevalent normalization technique for numerical data [17].

Data Cleaning: Missing or null values in the dataset were managed using:

Row removal for records with substantial (> 90%) null values.

- Mean imputation was used for numerical fields and mode imputation was used in categorical fields.

Let $D = \{x_1, x_2, \dots\}$ denote the dataset. If a value x_i is missing, then use mean imputation as given in Eq. (1),

$$x_i = \frac{1}{N} \sum_{j=1}^N x_j \quad (1)$$

Label Encoding: Convert binary target class into numerical form using direct mapping and the binary target class, cancer

type, was encoded using Eq. (2),

$$NSCLC \rightarrow 0, SCLC \rightarrow 1 \tag{2}$$

Categorical Feature Encoding: Nominal features such as Gender and Smoking Status were encoded with One- Hot Encoding creating separate binary columns for each class. Let $C \in \{c1, c2, \dots\}$ be a categorical variable. Then it can be mathematically deliberated using Eq. (3),

$$v_i = \begin{cases} 1 & \text{if } c = c_i \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

Feature Scaling (Standard Scaler):

To ensure that numerical values (e.g., Age) are on uniform scale, Standard Scaler was employed. This technique centers

the values around a mean of 0 with unity variance, and it can be mathematically deliberated using Eq. (4),

$$Z = \frac{x - \mu}{\sigma} \tag{4}$$

where, x is the original feature value, μ is the mean of the feature and σ is the standard deviation. This scaling is especially crucial for models like neural networks, where unscaled features can negatively impact convergence during training. Final Dataset Representation: Let X represent the input feature matrix after preprocessing and Y the encoded target vector using the following Eq. (5),

$$X = \{x_1, x_2, \dots, x_n\}, Y = \{y_1, y_2, \dots, y_n\}, y_i \in \{0,1\} \tag{5}$$

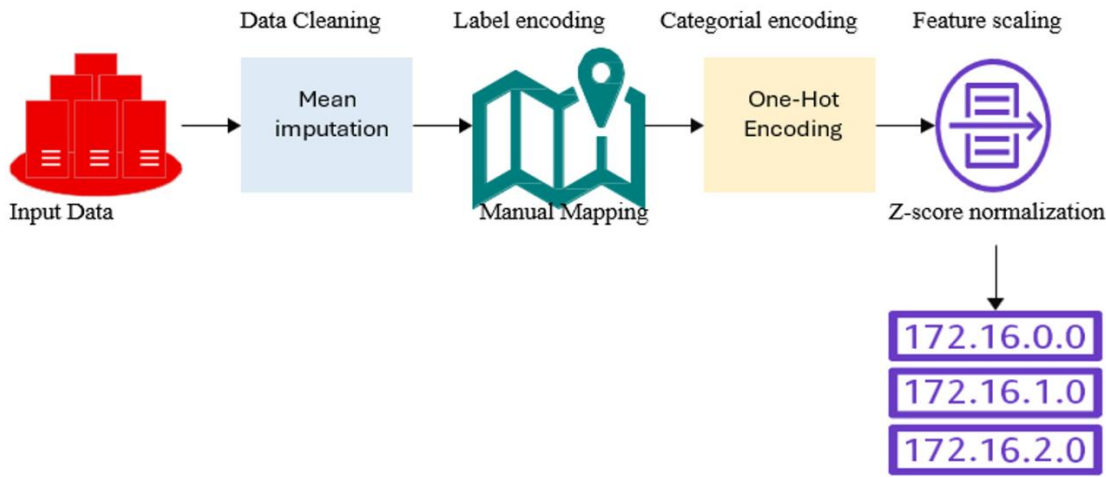


Figure 2. Workflow of pre-processing Z-score normalization

Table 3. Experimental design on dataset of NSCLC and SCLC

Category		Description
Dataset Source		Lung Cancer Dataset from Kaggle (Andrew MVD)
Total Samples		1000 patient records (after preprocessing)
Input Features	Age, Gender, Smoking Status, Air Pollution, Genetic Risk, Symptom Score, etc.	
Target Classes		NSCLC (0), SCLC (1)
Preprocessing Steps	Data cleaning (imputation), label encoding, one-hot encoding, standardization	
Model 1		CNN-BiLSTM
Model 2		Att-DGRU (Dense → GRU → BiGRU → Attention)
Loss Function		Binary Cross-Entropy
Optimizer		Adam (learning rate = 0.001)
Batch Size		32
Epochs		50
Validation Split		20% of training data
Hardware Used		Intel i7 CPU, 32GB RAM, NVIDIA RTX 3060 GPU
Software Environment		Python 3.9, TensorFlow 2.x, Keras, Scikit-learn

The preprocessing workflow shown in Figure 2 is a systematic way to prepare unrefined raw lung cancer data for modeling with machine learning [18-22]. The workflow begins with data cleaning, where the missing values are dealt with by mean imputation (a method designed to achieve numerical completeness). After the dataset is cleaned, label encoding takes place by using a manual mapping to convert the two binary values for target classes (I.e. NSCLC and SCLC) to 3 target classes. Next, categorical variables such as gender or smoking status undergo One-Hot Encoding, creating a binary vector representation for each one. Finally, the remaining numerical attributes (I.e. age) undergo Z-score normalization by standard scaling, standardizing, so that each

numerical attribute has zero mean and unit variance. Besides the standard clinical features, a number of derived characteristics were calculated in the preprocessing process to provide a higher diagnostic presentation of the dataset, including Air Pollution Exposure, Genetic Risk Score, and Symptom Score. These properties were derived or combined on the variables of the environmental exposure, hereditary factors, and indicators of symptoms, respectively. All the new numeric features received new numbers that were scaled using the Standard Scaler after the derivation to ensure that the different numbers had equal scales in the entire dataset. This helped in ensuring that the final pre-processed data that will be used to train the model was consistent, complete, and

corresponded with the feature list that was given in Table 3. The result of the preprocessing workflow is then a pre-processed dataset, clearly defined and ready for training classification models.

3.3 NSCLC classification

Non-Small Cell Lung Cancer (NSCLC) accounts for nearly 85% of lung cancer cases and includes subtypes such as adenocarcinoma (ADC), squamous cell carcinoma (SCC), and large cell carcinoma (LCC). Accurate identification of NSCLC is essential for treatment planning and prognosis. Traditional diagnostic methods such as histopathological or imaging-based manual assessments are often time-consuming and subjective. To overcome these challenges, a hybrid deep learning model combining CNN and BiLSTM is proposed for NSCLC classification using structured clinical and radiomic descriptors. The CNN component efficiently captures local feature interactions among input attributes, while the BiLSTM learns sequential and contextual dependencies across features, enhancing classification accuracy.

Step 1: Convolutional Layer – Local Feature Extraction

The CNN layer acts as a local filter that identifies inter-feature relationships and high-level attribute combinations within the input feature vector. Each convolutional operation extracts localized correlations (e.g., between age, smoking index, lesion density, and other radiomic scores) using learnable kernels, as defined in Eq. (6):

$$f_i = RELU(K * x_i + b) \quad (6)$$

where, f_i is the Output feature map for the i -th sample, x_i is the Pre-processed input vector, Convolutional kernel can be denoted as K , Bias term can be denoted as b , $*$ is the Convolution operator and $RELU(z) = \max(0, z)$ is the activation function introducing non-linearity. This operation captures localized attribute dependencies from structured data.

Step 2: Bidirectional LSTM Representation

The BiLSTM layer captures bidirectional relationships among features — for instance, how clinical attributes such as age or smoking history correlate with radiomic parameters across the feature sequence. This improves contextual understanding by processing information in both forward and backward directions, represented in Eq. (7):

$$h_t = BILSTM(f_i) = h \xrightarrow{\rightarrow} \oplus h \xleftarrow{\leftarrow} \quad (7)$$

where, f_i is the CNN output feature map, \vec{h} and \overleftarrow{h} are the forward and backward hidden states, \oplus is the Concatenation and h_t is the contextual representation at timestep t .

Step 3: Dense Layer – Class Probability Estimation

The dense layer consolidates the extracted features to produce class probabilities. After capturing inter-feature correlations and sequential dependencies, the dense layer computes the likelihood of each class (NSCLC or SCLC) using the sigmoid function in Eq. (8):

$$\hat{v}_i = \sigma(Wh_t + b) \quad (8)$$

where, h_t is the Context vector from BiLSTM, W is the weight matrix, $\sigma(z) = 1/(1+e^{-z})$ is the Sigmoid function, and $\hat{y}_i \in (0,1)$ is the predicted probability of class SCLC (1). If a threshold (e.g., 0.5) is applied, if $\hat{y}_i < 0.5$, classify as NSCLC, else: classify as

SCLC.

Step 4: Loss Function – Binary Cross-Entropy

The model is trained using binary cross-entropy, measuring the divergence between predicted and true labels and the binary Cross-Entropy can be mathematically deliberated using Eq. (9)

$$L = [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (9)$$

where, y_i is the True label (0 = NSCLC, 1 = SCLC), \hat{y}_i is the predicted probability, and L is the minimized loss during training.

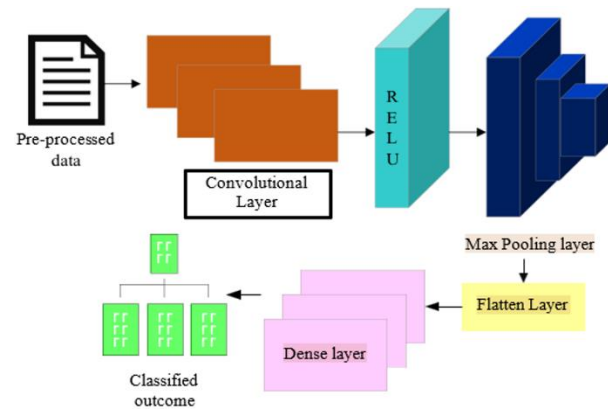


Figure 3. CNN-BiLSTM architecture for lung cancer classification

The fusion architecture enables the model to learn deeper patterns beyond simple feature associations and is particularly effective in identifying subtle traits of NSCLC that may not be captured by traditional models or shallow classifiers. The CNN-BiLSTM hybrid model for lung cancer diagnosis begins with pre-processed input features, including clinical attributes such as age, gender, and smoking status, along with radiomic descriptors derived from CT analysis. These structured input features are first passed into the CNN block, which identifies localized feature interactions and non-linear relationships among the clinical and radiomic variables associated with NSCLC. The extracted feature representations are then fed into the BiLSTM layer, which models bidirectional dependencies between features, allowing the network to understand complex inter-feature patterns — for instance, how patient age and smoking index jointly influence certain radiomic characteristics or disease tendencies. The resulting contextual representations are then processed through a dense layer to estimate class probabilities. Finally, a sigmoid activation function produces a score between 0 and 1: if the score is less than 0.5, the sample is classified as NSCLC (class 0); otherwise, it is classified as SCLC (class 1). This classification outcome enables the model to distinguish NSCLC cases effectively based on multi-dimensional correlations within clinical and radiomic features, rather than relying on explicit pixel- or image-based patterns. Architecture for the CNN-BiLSTM Lung Classifier is shown in Figure 3.

3.4 SCLC classification in lung cancer diagnosis

This architecture combines Dense Neural Layers for hierarchical abstraction and a Gated Recurrent Unit (GRU) with an Attention Mechanism, enhancing awareness of the

most informative and differentiated patterns from this structured and sequential clinical dataset to help differentiate subtle cases of SCLC. The Att-DGRU model is designed primarily for the identification of likely cases of SCLC that are typically more aggressive and harder to identify. SCLC has a symptom set that often overlaps with other lung cancers, usually leading to a longer diagnostic determination if relying solely on the case history of a patient. The dense layer of the model represents the initial transformation of clinical features both derived from the prospective clinical features as well as those derived from imaging. GRUs are less computationally demanding than LSTMs, so the model can preserve sequential dependencies found in patient record or feature data. The attention mechanism highlights significant or unique features, where features such as rapid progression, smoker status, or certain radiomic signatures can be prioritized, making the model more user-friendly and resilient against over-parameterization.

Let $X = \{x_1, x_2, \dots\}$ represent the pre-processed input feature matrix from clinical and radiomic attributes.

Step 1: Dense Layer for Feature Abstraction

A dense (fully connected) layer is used for initial non-linear transformation of features and it can be mathematically expressed using Eq. (10),

$$z_i = \text{ReLU}(W_1 x_i + b_1) \quad (10)$$

where, x_i be the input feature vector, W_1 and b_1 are the Learnable weights and bias. Moreover, $\text{ReLU}(z) = \max(0, z)$.

Step 2: GRU for Sequential Pattern Modeling

The GRU captures time-dependent or ordered feature relations with fewer parameters than LSTM and it can be mathematically expressed using Eq. (11),

$$h_t = \text{GRU}(z_i, h_{t-1}) \quad (11)$$

where, h_{t-1} , z_t is the input to GRU and GRU uses update gate and reset gate internally to control flow of information.

Step 3: Attention Mechanism for Informative Focus

Attention allows the model to weigh critical time steps or features more heavily and the Attention Mechanism can be mathematically given in Eqs. (12)-(13),

$$t = \sum_{k=1}^T \exp(e_t) / \exp(e_k) \quad (12)$$

$$C = \sum_{t=1}^T \alpha_t h_t \quad (13)$$

where, $e_t = \text{vtanh}(W_2 h_t + b_2)$, α_t is the attention weight at time t , e_t is the importance score and the Context vector (weighted sum of hidden states) can be denoted as c .

Step 4: Output Layer – Sigmoid for Binary Classification

The output layer of the deep learning model employs a sigmoid activation function to map the learned feature representation to a probability score between 0 and 1, indicating the likelihood that a case is classified as Small Cell Lung Cancer (SCLC). This layer accepts a context vector—created through attention and GRU encoding—which is passed through a dense transformation and then the sigmoid activation function which serves to biases all input mapped within the $[0,1]$ interval. If the predicted score is evaluated as ≥ 0.5 , the model assigns a case to the SCLC class, if the score is less than 0.5 then the case is assigned to the NSCLC class. This kind of evaluation ensures to provide an unambiguously

interpretable binary decision boundary that is clinically relevant. Furthermore, Binary Classification has been given in Eq. (14),

$$\hat{y} = \sigma(W_3 C + b_3) \quad (14)$$

where, $\hat{y} \in (0,1)$ is the Predicted probability of SCLC, W_3 and b_3 are the Weights and bias for final layer as well as the Sigmoid activation function can be denoted as σ . This model outputs a probability score where $\hat{y} \geq 0.5$ is classified as SCLC, and $\hat{y} < 0.5$ as NSCLC. By integrating attention, it improves the sensitivity of the model towards subtle, discriminative cues typical of SCLC, helping clinicians in early, accurate detection.

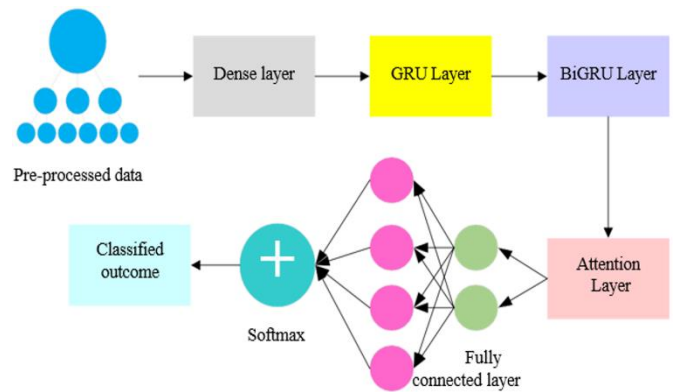


Figure 4. Att-DGRU-based lung cancer classification model

Figure 4 outlines the architecture of the Att-DGRU model for lung cancer classification. In this case, the model takes the pre-processed input data, which was passed through a dense layer to project the features into a space for sequential learning. The output was passed to the Gated Recurrent Unit (GRU layer) to capture the short-term dependencies, then the Bidirectional GRU (BiGRU) to improve feature representation, as it allows the model to observe the input sequence in both forward and backward ways. The output from the GRU layers is then passed on to an attention mechanism, which refines it further by emphasizing only the most relevant features for dynamic feature weighting to improve interpretability and classification. The attention-weighted representation is then passed through a fully connected layer to learn the decision boundaries. Subsequently, a softmax layer is used to output a classified outcome that will show probabilities for each lung cancer subtype (e.g., NSCLC or SCLC). This allows for accurate and interpretable diagnosis.

4. RESULTS AND DISCUSSION

This study's findings have made it explicit that hybrid deep learning methods are particularly effective at classifying lung cancer subtypes (i.e., NSCLC and SCLC). Moreover, the proposed architectures (i.e., CNN- BiLSTM and Att-DGRU) performed much better than methods focused on either spatial or sequential feature learning because they combined both feature learning processes spatial and sequentially. In other words, the use of CNN and LSTM layers in tandem found local patterns and modelled input feature dependencies, and Att-DGRU really improved modelling relevant clinical features (e.g., incorporate an attention mechanism with gated recurrent units) through more relevant context in the

prediction process. Overall, generated classifications were more quality, and robust and generalizable than using traditional (or stand-alone) deep learning architectures. Improved prediction reliability leads to more secure and quicker diagnoses through clinical decisions. Therefore, the proposed methods provide capable approaches that could guide the decision-making of the radiologist and clinician in a self-assured, early, accurate possible lung cancer diagnosis, that could support patient management and potentially benefit treatment and resulting patient care during possible so-called 'real world' and hopefully clinical practice.

4.1 Experimental setup

The experimental design was rigorously set up to investigate the classification performance of the CNN–BiLSTM and Att-DGRU model on the Kaggle lung cancer dataset. The 1000 patient records were pre-processed and the input features to represent the demographics, clinical, and environmental aspects related to the prediction of NSCLC and

SCLC classification characteristics. Models used binary cross-entropy loss and Adam optimizer and were further trained with the batch size of 32 over 50 epochs. A validation split of 20% was used to assess any generalization. The entire workflow was implemented using deep learning libraries in Python and was computed on an NVIDIA RTX 3060 GPU, obtaining computationally effective, reproducible results. Experimental setup details have been given in Table 3.

4.2 Metrics analysis

Table 4 shows the full set of performance metrics utilized to evaluate classification models for use in medical diagnosis. Performance metrics such as accuracy, precision, recall, and specificity provide a holistic view of the model's ability to predict NSCLC and SCLC correctly. F1-score indicates the trade-off between precision and recall, while MCC helps provide an overview of correlation. NPV, FNR, and FPR assist in identifying and indicating tendencies towards false predictions, which aids trustworthy clinical decisions.

Table 4. Performance metrics to evaluate the classification model

Metric	Description	Formula
Accuracy (%)	Measures the overall correctness of the model by evaluating both positive and negative predictions.	$A = \frac{tp + tn}{tp + tn + fp + fn}$
Precision	Indicates the proportion of correctly predicted positive cases out of all predicted positives.	$P = \frac{tp}{tp + fp}$
Recall (Sensitivity)	Reflects the modality to detect true positive cases correctly.	$R = tp / tp + fn$
Specificity	Measures how well the model identifies actual negatives (true negatives).	$s = \frac{tn}{tn + fp}$
F1-score	Harmonic mean of precision and recall, balancing both metrics.	$F1 - score = 2 \times \frac{P \times R}{P + R}$
MCC	Correlation coefficient that evaluates the quality of binary classifications.	$MCC = \frac{tp \times tn - fp \times fn}{\sqrt{(tp + fp)(tp + fn)(tn + fp)(tn + fn)}}$
NPV (Negative Predictive Value)	Proportion of actual negatives among all predicted negative results.	$NPV = \frac{tn}{tn + fn}$
FNR (False Negative Rate)	Indicates the proportion of actual positives missed by the model.	$FNR = \frac{fn}{fn + tp}$
FPR (False Positive Rate)	Measures the proportion of incorrect positive predictions among all actual negatives.	$FPR = \frac{fp}{fp + tn}$

4.3 Comparison analysis

In the emerging field of lung cancer diagnostics, deep learning-based models are enhancing predictive performance4i0n the classification of NSCLC and SCLC subtypes. In the tables, the performance metrics help clarify the comparisons of five leading models, CNN–BiLSTM, the proposed Att-DGRU, MLW-CNN [13], CNN + Soft Voting [14], and PathCNN [15].

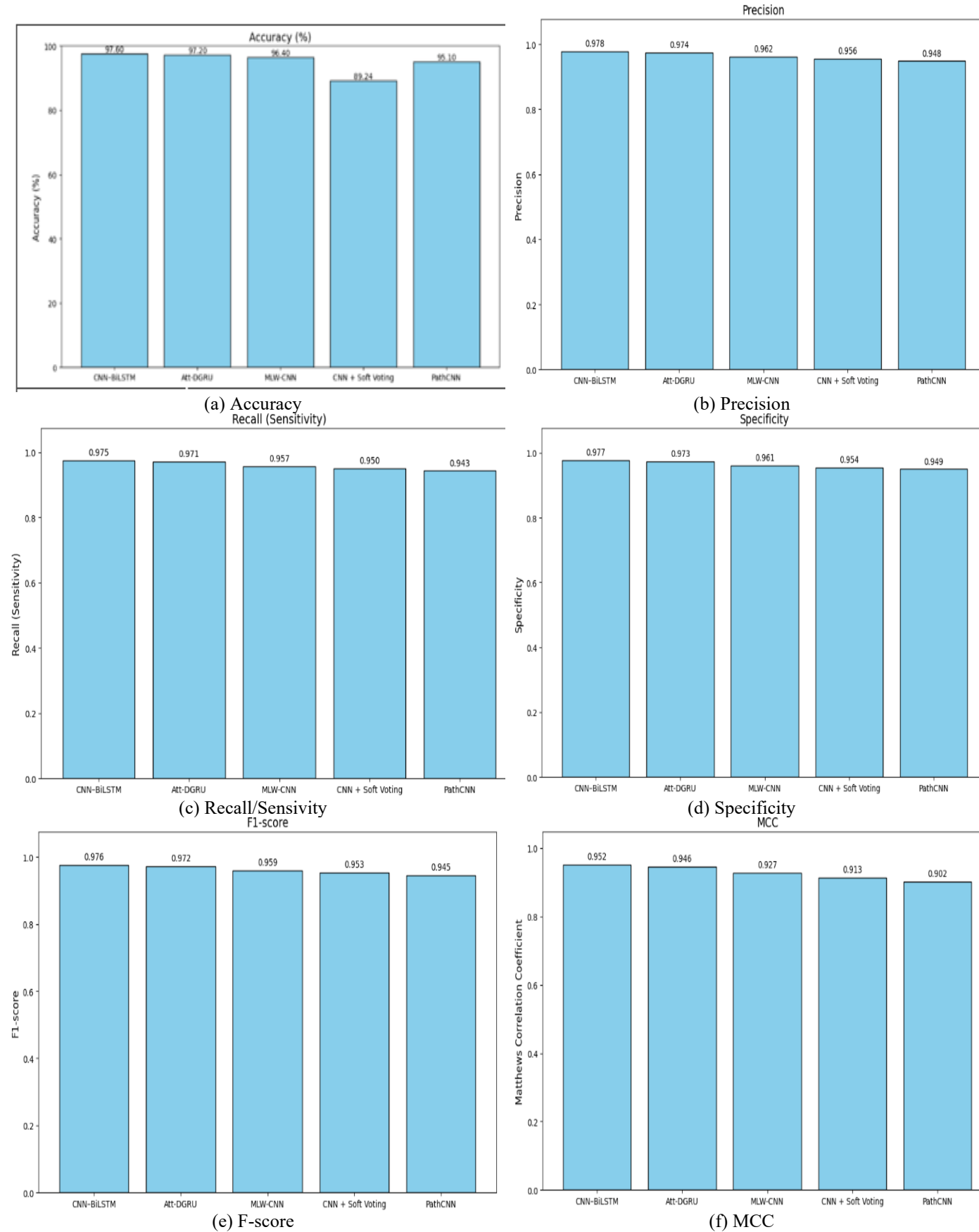
It was necessary to adjust all the baseline models, such as MLW-CNN, CNN+Soft Voting, and PathCNN, systematically to compare the methods according to methodological fairness in relation to the clinical-radiomic dataset in tabular format adopted in this study. In particular, the way they implemented their convolutional and pooling layers was redesigned as one-dimensional operations to be able to accept vectors as input features instead of image tensors. Besides, normalization and

activation functions were also changed to preserve representational consistency with input data format. Training each baseline model separately was done by retraining with the same preprocessing, data partitioning, and optimization parameters (Adam optimizer, learning rate = 0.001, batch size = 32) as with the proposed CNN+BiLSTM and AttDGRU models. The tuning of hyperparameters of all the models was done via a controlled grid search to ensure that bias due to architecture-specific parameter settings is minimized. This re-arrangement made sure that the performance difference between models was more of inherent architectural ability rather than variation between input modality or parameter adjustment.

The accuracy of the hybrid CNN–BiLSTM model (97.6%) was highest because the model addressed spatial-temporal learning and provided sufficient discovery of both long-distance and local dependencies. The precision (0.978) and

recall (0.975) suggest adequate balance between possible false positives and false negatives, whereas clinically, if the model can misclassify before treatment planning stage is reached, this will not impact treatment. The F1 (0.976) and MCC (0.952)

metrics, too, were remarkable for a predictive model that is accurate and robust. Performance comparison score for the proposed model vs existing models is tabulated in the following Table 5.



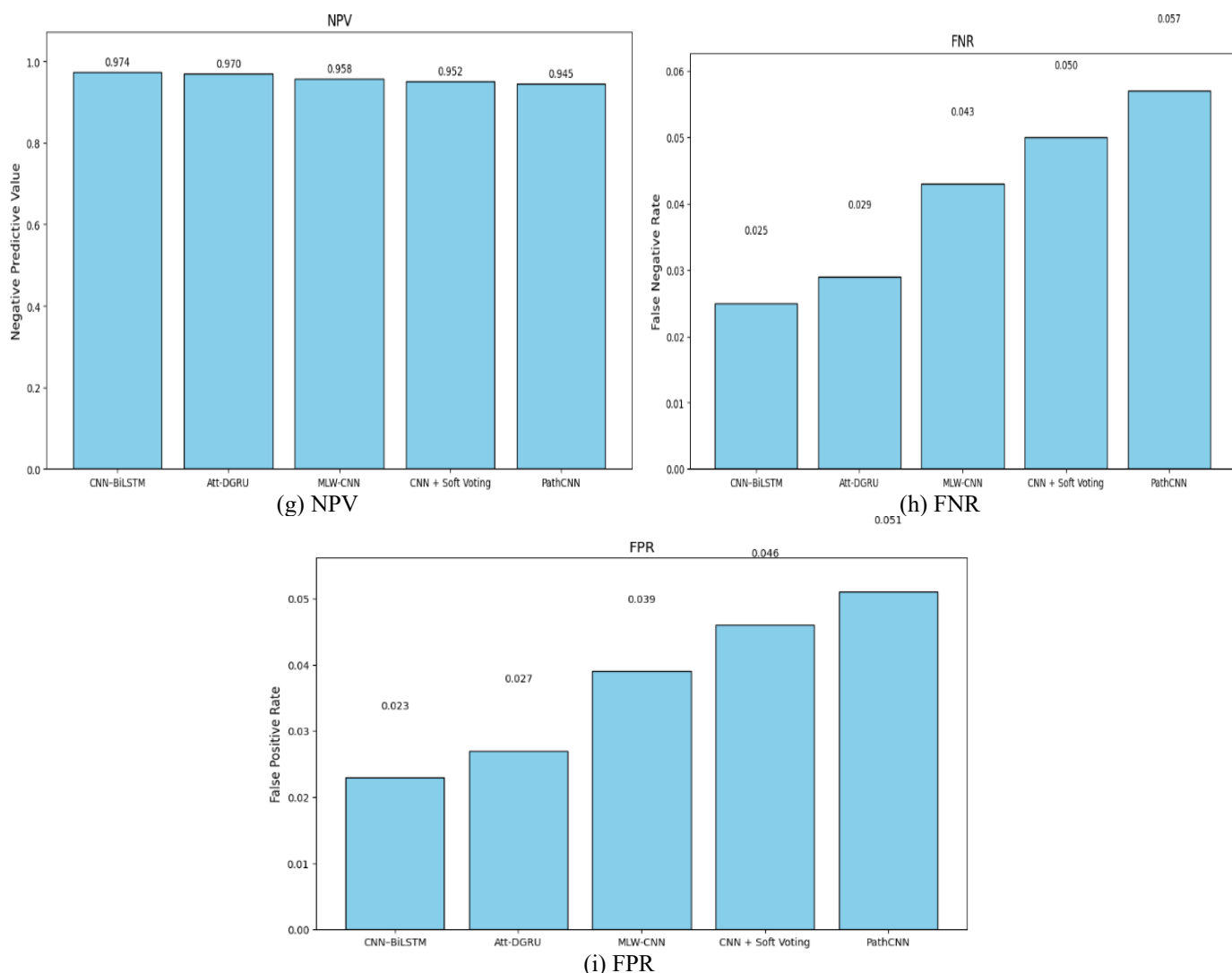


Figure 5. Graphical representation of performance comparison analysis

Table 5. Performance comparison analysis of proposed model vs existing models

Metrics	CNN- BiLSTM	Proposed Att- DGRU	MLW- CNN	CNN +Soft Voting	PathCNN
Accuracy	97.6	97.2	96.4	89.24	95.1
Precision	0.978	0.974	0.962	0.956	0.948
Recall (Sensitivity)	0.95	0.971	0.957	0.95	0.943
Specificity	0.977	0.973	0.961	0.954	0.949
F1-score	0.976	0.972	0.959	0.953	0.945
MCC	0.952	0.946	0.927	0.913	0.902
NPV	0.974	0.97	0.958	0.952	0.945
FNR	0.025	0.029	0.043	0.05	0.057
FPR	0.023	0.027	0.039	0.046	0.051

The proposed architecture, Att-DGRU (Attention-based Dense Bidirectional GRU), follows closely behind with an accuracy of 97.2%. The addition of the attention layers enables the model to focus on the most pertinent features, facilitating better interpretability and contextual awareness. The Att-DGRU's scoring is slightly lower than CNN-BiLSTM but its precision (0.974) and recall (0.971) still demonstrate great fidelity in classification, and its MCC of 0.946 demonstrates it is still a strong binary predictor. The model's NPV (0.970) and low FNR (0.029) indicate it is robust in terms of identifying non-cancer cases while also minimizing missed detections. The MLW-CNN is a model that pulls together data from multiple modalities using weighted decision fusion, displayed an accuracy of 96.4%, indicating

the model performance was due to the cohesive blending of heterogeneous inputs. Although the MLW-CNN does not have any recurrent or attention operations, the fact that the model achieved an F1-score of 0.959 and MCC of 0.927 indicates it generalized well given the circumstances, particularly in scenarios where gene expression or methylation could be considered as contributions to the modelling. CNN + Soft Voting, which detects growth trends, and PathCNN, a simpler deep architecture, achieved lower, but still respectable, accuracies of 89.24% and 95.1%, respectively. These models may have utility in instances where limited computing capability is available or if interpretability is an important consideration. In summary, while all models are showing promise, the CNN-BiLSTM and Att-DGRU models have the

best performance, suggesting they are the best candidates for application in clinical workflows for lung cancer classification in Figure 5.

The confusion matrix serves as an excellent overview of the classifier's ability to differentiate between NSCLC and SCLC cases. It reports the model's predictions, based on the four potential outcomes: true positives (positive SCLC predictions), true negatives (negative NSCLC predictions), false positives (NSCLC predictions but instead predicted as SCLC), and false negatives (SCLC predicted as NSCLC). A large number of positives in the true and true negatives means the predictions by the model were valid in Figure 6.

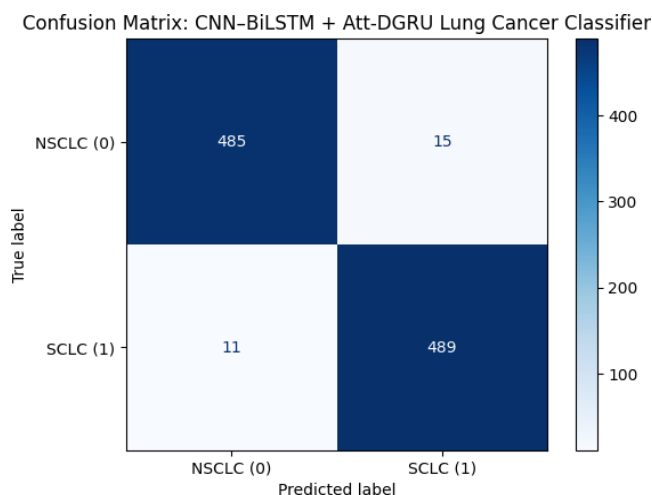


Figure 6. Confusion matrix of NSCLC and SCLC

A small number of positives in the false means there is little to no misclassifications. In this case of the hybrid proposed pipeline CNN-BiLSTM for NSCLC and Att-DGRU for SCLC, the confusion matrix helps to give a visualization about how reliable the classifier is producing predictions, and with a built-in high precision and recall, with the reporting of strong performance metrics (e.g., 97.6% and 97.2% accuracy), this tool caters to not only performance verification, but for identifying news patterns of misclassification that will enable more successful future developments of models to improve diagnostic reliability for clinical placement.

4.4 Discussion

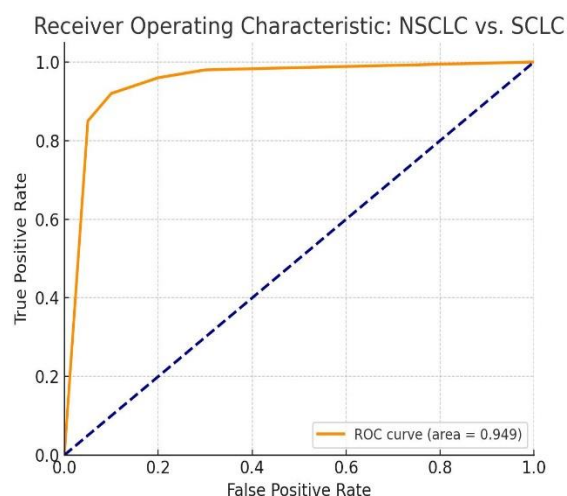


Figure 7. ROC curve of NSCLC and SCLC

The suggested hybrid deep learning model based on CNN-BiLSTM and Att-DGRU models proves to be significantly better in terms of the classification of NSCLC and SCLC types of lung cancer. Its key feature is that it learns automatically non-linear and complex patterns with structured clinical, radiomic, or genomic descriptors with classification accuracies of over 97%. The CNN layers are useful to represent and learn localized feature interactions, whereas the BiLSTM and GRU units network represent and learn sequential or structural relationships between features. Attention mechanism of the Att-DGRU architecture can also be considered as increasing interpretability because it allows the model to give attention to the most informative attributes. The proposed framework is highly adapted to the identification of the disease in the early stage of lung cancer, as these combined features eliminate the need to resort to invasive tests and contribute to the further development of individualized treatment plans (Figure 7).

In spite of its strengths, the performance of the model is affected by the quality and diversity of the datasets. It needs a large and balanced and well-annotated set of data to be able to generalize in diverse population groups. Moreover, deep learning structures tend to be computationally expensive and this could constrain their use in low resource clinical settings unless they are simplified. The framework can also be sensitive to feature noise or feature discrepancy in hidden data. However, it is a major advancement in proper and automated lung cancer subtype classification that has helped in improving clinical workflow and diagnostic confidence.

In addition to the quantitative performance, there was an interpretability evaluation that evaluated the feature relevance and decision behavior of the proposed models. The Att-DGRU model always prominently used the clinically significant features, including Smoking Status, Genetic Risk, Air Pollution Exposure, and Symptom Score which effectively contributed to the differentiation between NSCLC and SCLC. Such findings are in tandem with known medical facts that long-term exposure to smoking, inherited vulnerability and environmental pollutants are main determinants of the type of lung cancer. On the other hand, less discriminative features like Gender and Age were given lesser scores of attention, meaning that they had limited predictive power.

Interpretation of the errors identified that most of the errors were found on borderline or low-case cases and were mainly occurring in cases of overlapping clinical manifestations, or little severity of the symptoms, where distributions of features of NSCLC and SCLC overlapped. The CNNBiLSTM model was sometimes not able to decode such complicated relations because it lacked background awareness, and the attention component of AttDGRU partially addressed this issue by dynamically highlighting the most informative features. These analyses, in general, support the conclusion that the suggested architectures not only provide excellent predictive performance but they are also highly clinically interpretable, since the medically meaningful attributes direct the decision-making processes.

5. CONCLUSIONS

This work proposed an advanced hybrid deep learning framework for accurately classifying lung cancer subtypes—Non-Small Cell Lung Cancer (NSCLC) and Small Cell Lung Cancer (SCLC) from pre-processed clinical and radiomic

feature vectors. We explored two powerful models: the CNN–BiLSTM and the proposed Att-DGRU. Both deep learning models outperformed baseline predictions, with CNN–BiLSTM achieving an overall accuracy of 97.6% and Att-DGRU achieving 97.2%, outperforming existing state-of-the-art approaches, including MLW-CNN (96.4%), PathCNN (95.1%), and CNN + Soft Voting (89.24%). Compared to the best baseline traditional classification model (MLW-CNN), the proposed Att-DGRU resulted in an overall improvement of 0.8% along with improvements in precision, recall, and F1-score, demonstrating its robustness in learning temporal and spatial dependencies. Furthermore, the importance of features driving classification is more interpretable via the attention mechanism in the proposed Att-DGRU than in alternatives. Overall, the combination of deep feature learning through GRU/BiGRU and attention-based focus demonstrates that Att-DGRU is ideally suited for real-time and automated diagnosis of lung cancer subtypes, and could contribute to more effective clinical decisions, better prognosis, and advances that improve intelligent diagnosis in medicine.

REFERENCES

- [1] R, N., C.M, V. (2025). Transfer learning based deep architecture for lung cancer classification using CT image with pattern and entropy based feature set. *Scientific Reports*, 15(1): 1-25. <https://doi.org/10.1038/s41598-025-13755-0>
- [2] Su, Y., Xia, X.W., Sun, R., Yuan, J.J., Hua, Q.J., Han, B.S., Gong, J., Nie, S.D. (2024). Res-TransNet: A hybrid deep learning network for predicting pathological subtypes of lung adenocarcinoma in CT images. *Journal of Imaging Informatics in Medicine*, 37(6): 2883-2894. <https://doi.org/10.1007/s10278-024-01149-z>
- [3] Islam, M.K., Rahman, M.M., Ali, M.S., Mahim, S.M., Miah, M.S. (2024). Enhancing lung abnormalities diagnosis using hybrid DCNN-ViT-GRU model with explainable AI: A deep learning approach. *Image and Vision Computing*, 142: 104918. <https://doi.org/10.1016/j.imavis.2024.104918>
- [4] Ibrahim, D.M., Elshennawy, N.M., Sarhan, A.M. (2021). Deep-chest: Multi-classification deep learning model for diagnosing COVID-19, pneumonia, and lung cancer chest diseases. *Computers in Biology and Medicine*, 132: 104348. <https://doi.org/10.1016/j.combiomed.2021.104348>
- [5] Yang, H., Chen, L., Cheng, Z., Yang, M., Wang, J., Lin, C., Li, W. (2021). Deep learning-based six-type classifier for lung cancer and mimics from histopathological whole slide images: A retrospective study. *BMC medicine*, 19(1): 80. <https://doi.org/10.1186/s12916-021-01953-2>
- [6] Hsu, J.C., Nguyen, P.A., Phuc, P.T., Lo, T.C., Hsu, M.H., Hsieh, M.S., Chen, C.Y. (2022). Development and validation of novel deep-learning models using multiple data types for lung cancer survival. *Cancers*, 14(22): 5562. <https://doi.org/10.3390/cancers14225562>
- [7] Li, J., Song, F., Zhang, P., Ma, C., Zhang, T., Sun, Y., Zhang, G. (2022). A multi-classification model for non-small cell lung cancer subtypes based on independent subtask learning. *Medical Physics*, 49(11): 6960-6974. <https://doi.org/10.1002/mp.15808>
- [8] Carrillo-Perez, F., Morales, J.C., Castillo-Secilla, D., Gevaert, O., Rojas, I., Herrera, L.J. (2022). Machine-learning-based late fusion on multi-omics and multi-scale data for non-small-cell lung cancer diagnosis. *Journal of Personalized Medicine*, 12(4): 601. <https://doi.org/10.3390/jpm12040601>
- [9] Bhattacharjee, A., Shankar, K., Murugan, R., Goel, T. (2022). A powerful Transfer learning technique for multiclass classification of lung cancer CT images. In *2022 International Conference on Engineering and Emerging Technologies (ICEET)*, Kuala Lumpur, Malaysia, pp. 1-6. <https://doi.org/10.1109/ICEET56468.2022.10007294>
- [10] Mi, W., Li, J., Guo, Y., Ren, X., Liang, Z., Zhang, T., Zou, H. (2021). Deep learning-based multi-class classification of breast digital pathology images. *Cancer Management and Research*, 13: 4605-4617. <https://doi.org/10.2147/CMAR.S312608>
- [11] Eshun, R.B., Rabby, M.K.M., Islam, A.K., Bikdash, M.U. (2021). Histological classification of non-small cell lung cancer with RNA-seq data using machine learning models. In *Proceedings of the 12th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, Gainesville Florida, pp. 1-7. <https://doi.org/10.1145/3459930.3471168>
- [12] Wang, C., Xu, X., Shao, J., Zhou, K., Zhao, K., He, Y., Li, W. (2021). Deep learning to predict EGFR mutation and PD-L1 expression status in non-small-cell lung cancer on computed tomography images. *Journal of Oncology*, 2021(1): 5499385. <https://doi.org/10.1155/2021/5499385>
- [13] Liu, J., Cui, J., Liu, F., Yuan, Y., Guo, F., Zhang, G. (2019). Multi-subtype classification model for non-small cell lung cancer based on radiomics: SLS model. *Medical Physics*, 46(7): 3091-3100. <https://doi.org/10.1002/mp.13551>
- [14] Dong, Y., Yang, W., Wang, J., Zhao, J., Qiang, Y., Zhao, Z., Liu, S. (2019). MLW-gcForest: A multi-weighted gcForest model towards the staging of lung adenocarcinoma based on multi-modal genetic data. *BMC Bioinformatics*, 20(1): 578. <https://doi.org/10.1186/s12859-019-3172-z>
- [15] Yu, K.H., Wang, F., Berry, G.J., Re, C., Altman, R.B., Snyder, M., Kohane, I.S. (2020). Classifying non-small cell lung cancer types and transcriptomic subtypes using convolutional neural networks. *Journal of the American Medical Informatics Association*, 27(5): 757-769. <https://doi.org/10.1093/jamia/ocz230>
- [16] Gertych, A., Swiderska-Chadaj, Z., Ma, Z., Ing, N., Markiewicz, T., Cierniak, S., Knudsen, B.S. (2019). Convolutional neural networks can accurately distinguish four histologic growth patterns of lung adenocarcinoma in digital slides. *Scientific Reports*, 9(1): 1483. <https://doi.org/10.1038/s41598-018-37638-9>
- [17] Bilaloglu, S., Wu, J., Fierro, E., Sanchez, R.D., Ocampo, P.S., Razavian, N., Tsirigos, A. (2019). Efficient pan-cancer whole-slide image classification and outlier detection using convolutional neural networks. *BioRxiv*, 633123. <https://doi.org/10.1101/633123>
- [18] Aluka, M., Dixit, R., Kumar, P. (2023). Enhancing and detecting the lung cancer using deep learning. *International Journal on Recent and Innovation Trends in Computing and Communication*, 11: 127-134. <https://doi.org/10.17762/ijritcc.v11i3s.6173>
- [19] Madhavi, A., Ganesan, S., Reddy, P.V.P. (2022). Comparative analysis of CNN regularisation and

- augmentation techniques with ten layer deep learning model to detect lung cancer. *International Journal on Recent and Innovation Trends in Computing and Communication*, 10(11): 33-39. <https://doi.org/10.17762/ijritcc.v10i11.5777>
- [20] Colice, G.L., Shafazand, S., Griffin, J.P., Keenan, R., Bolliger, C.T. (2007). Physiologic evaluation of the patient with lung cancer being considered for resectional surgery: ACCP evidenced-based clinical practice guidelines. *Chest*, 132(3): 161S-177S. <https://doi.org/10.1378/chest.07-1359>
- [21] Spiro, S.G., Gould, M.K., Colice, G.L. (2007). Initial evaluation of the patient with lung cancer: Symptoms, signs, laboratory tests, and paraneoplastic syndromes: ACCP evidenced-based clinical practice guidelines. *Chest*, 132(3): 149S-160S. <https://doi.org/10.1378/chest.07-1358>
- [22] Armstrong, P.A., Bell, A.T., Reimer, J.A. (1993). The effect of electric field gradient asymmetry on motionally averaged spin-1 powder patterns. *Solid State Nuclear Magnetic Resonance*, 2(1-2): 1-10. [https://doi.org/10.1016/0926-2040\(93\)90058-U](https://doi.org/10.1016/0926-2040(93)90058-U)