IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

IEEE ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY SECTION

Received 9 August 2024, accepted 14 September 2024, date of publication 17 September 2024, date of current version 7 October 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3462629

APPLIED RESEARCH

Enhanced Lung Cancer Detection and TNM Staging Using YOLOv8 and TNMClassifier: An Integrated Deep Learning Approach for CT Imaging

ALAA WEHBE[1], SILVANA DELLEPIANE[1], AND IRENE MINETTI[2]
[1]Department of Naval, Electrical, Electronic and Telecommunications Engineering (DITEN), University of Genoa, 16126 Genoa, Italy
[2]EBIT—ESAOTE Group, 16153 Genoa, Italy

Corresponding author: Alaa Wehbe (alaa.wehbe@edu.unige.it)

The work supported in part by the PNRR Project "RAISE," through by the European Union - NextGeneration.

ABSTRACT This paper introduces an advanced method for lung cancer subtype classification and detection using the latest version of YOLO, tailored for the analysis of CT images. Given the increasing mortality rates associated with lung cancer, early and accurate diagnosis is crucial for effective treatment planning. The proposed method employs single-shot object detection to precisely identify and classify various types of lung cancer, including Squamous Cell Carcinoma (SCC), Adenocarcinoma (ADC), and Small Cell Carcinoma (SCLC). A publicly available dataset was utilized to evaluate the performance of YOLOv8. Experimental outcomes underscore the system's effectiveness, achieving an impressive mean Average Precision (mAP) of 97.1%. The system demonstrates the capability to accurately identify and categorize diverse lung cancer subtypes with a high degree of accuracy. For instance, the YOLOv8 Small model outperforms others with a precision of 96.1% and a detection speed of 0.22 seconds, surpassing other object detection models based on two-stage detection approaches. Building on these results, we further developed a comprehensive TNM classification system. Features extracted from the YOLO backbone were reduced using Principal Component Analysis (PCA) to enhance computational efficiency. These reduced features were then fed into a custom TNMClassifier, a neural network designed to classify the Tumor, Node, and Metastasis (TNM) stages. The TNMClassifier architecture comprises fully connected layers and dropout layers to prevent overfitting, achieving an accuracy of 98% in classifying the TNM stages. Additionally, we tested the YOLOv8 Small model on another dataset, the Lung3 dataset from the Cancer Imaging Archive (TCIA). This testing yielded a recall of 0.91, further validating the model's effectiveness in accurately identifying lung cancer cases. The integrated system of YOLO for subtype detection and the TNMClassifier for stage classification shows significant potential to assist healthcare professionals in expediting and refining diagnoses, thereby contributing to improved patient health outcomes.

INDEX TERMS Lung cancer classification, TNM staging, YOLOv8, CT images.

I. INTRODUCTION

According to the World Health Organization (WHO) [1], lung cancer ranks third in prevalence in Italy and has become one of the most common causes of cancer-related mortality worldwide. Lung cancer is divided into two main types: Small Cell Lung Cancer (SCLC) and Non-Small Cell Lung Cancer (NSCLC). NSCLC, which constitutes over 80 percent of cases, includes three subtypes: Adenocarcinoma (ADC), Squamous Cell Carcinoma (SCC), and Large Cell Carcinoma (LCC) [2]. ADC is the most prevalent subtype, accounting for 40 percent of NSCLC cases, and is located in mucus-secreting glands in the outer lung region. SCC is directly linked to smoking and is centered in the lungs, ranking third among

The associate editor coordinating the review of this manuscript and approving it for publication was Gina Tourassi.

NSCLC subtypes. Although observed in only 10 percent of cases, large-cell undifferentiated carcinoma exhibits rapid spread throughout the lungs [3].

Survival rates in lung cancer hinge on various factors, including **cancer stage**, age, general health, and treatment efficacy. Accurate classification and staging significantly affect the outcomes by shaping effective treatment strategies, reducing the time to diagnosis, and improving the quality of life for patients. Precision imaging techniques — Computed Tomography (CT) [4], Positron Emission Tomography combined with CT (PET/CT), and Magnetic Resonance Imaging (MRI) [5]— play a pivotal role in evaluating and distinguishing these subtypes.

In recent years, deep learning methods like convolutional neural networks (CNNs) have been particularly effective in analyzing CT images and identifying distinct features associated with different subtypes. For example, Pang et al. [16] proposed a densely connected network (DenseNet) to classify malignant tumors from CT images, followed by an adaptive boosting algorithm to improve classification accuracy. Similarly, Wang et al. [18] utilized deep residual neural networks with a transfer learning strategy to identify pathological types of lung cancer from CT images. However, these methods are based on classification without providing any visual information that helps doctors in understanding the output of those model. Therefore, object detection models based on deep learning algorithms have been employed to improve the accuracy of lung cancer subtype classification and localization. Object detection in deep learning encompasses two predominant paradigms: two-stage detection algorithms rooted in anchor boxes, and single-stage detection algorithms founded on anchor-free boxes. The former category includes well-known models such as R-CNN [20], Fast R-CNN [21], Faster R-CNN [22], R-FCN [23], and Mask R-CNN [24], Cascade R-CNN [34], Sparse-RCNN [35], and Improve Sparse RCNN [36] which typically achieve high accuracy but at the cost of increased detection time. In contrast, the latter category includes SSD [25] and YOLO algorithms [26], and TSD-YOLO [37], which employ a forward inference network to efficiently determine target locations and generate classification results.

In recent decades, computer-aided detection (CAD) systems have emerged as valuable tools in medical imaging, significantly aiding in the detection of lung nodules [6], [7], [8], [9]. Gunasekaran et al. [38] applied YOLOv5, optimizing hyperparameters and using augmentation techniques, achieving high accuracy and recall in lung cancer detection, outperforming previous methods, and demonstrating real-time clinical applicability. Liu [39] utilized the STBi-YOLO approach, derived from YOLO-v5 and incorporating stochastic-pooling-based spatial pyramid pooling, a bidirectional feature pyramid network, and an optimized EIoU loss function, achieving high accuracy and recall for lung nodule detection in CT images, outperforming YOLO-v3, YOLO-v4, YOLO-v5, Faster R-CNN, and SSD in experiments. Extensive research and development within the CAD domain have markedly improved the accuracy and efficiency of lung cancer diagnosis. For instance, the study by [10] innovatively classified lung cancer into squamous cell carcinoma (SCC), adenocarcinoma (ADC), and small cell lung cancer (SCLC) using a self-calibrated convolution module based on RCNN and a novel regression loss function. This method achieved an impressive mean average precision (mAP); however, it did not provide descriptions of tumor size or additional information regarding the stage of each cancer type. Similarly, Qi et al. [11] focused on the localization and classification of histological subtypes in lung cancer using a center-point-based approach, CenterNet, to train and localize CT images. While this method demonstrated efficiency in lung cancer detection by reducing false positives and negatives, it also lacked detailed information concerning the cancer stage. These studies underscore the potential of CAD systems in enhancing lung cancer diagnosis but highlight the need for comprehensive data on tumor size and staging to fully inform clinical decision-making.

In addition to distinguishing between lung cancer subtypes, another significant challenge is the accurate staging of lung cancer, which includes classification based on TNM (Tumor, Node, and Metastasis) stages. The TNM staging system classifies cancer based on three key components: Tumor (T), Node (N), and Metastasis (M). The first component, Tumor (T), describes the size and extent of the primary tumor, and accurate measurement of tumor size is critical as it directly impacts the T-stage classification. The Node (N) component indicates whether cancer has spread to nearby lymph nodes and the extent of such spread. Lymph node involvement is crucial in determining disease progression and influences treatment decisions. Finally, the Metastasis (M) component describes whether cancer has spread to distant parts of the body, with the presence of metastasis signifying advanced disease and a poorer prognosis. Accurate TNM staging is essential for determining the appropriate treatment strategy and predicting patient outcomes. To the best of our knowledge, no clear study has targeted both the major challenges—distinguishing lung cancer subtypes and accurate TNM staging—within a single framework. Various approaches have been implemented for the TNM problem. Kirienko et al. [30] developed a CNN-based algorithm to classify lung cancer lesions as T1-T2 or T3-T4 on FDG-PET/CT images, achieving an accuracy of 82.6%. This work only considers T-staging without addressing N or M staging. Moitra et al. [31] developed a 1D CNN model for automated staging and grading of NSCLC using the NSCLC Radiogenomics Collection from TCIA, achieving satisfactory performance with accuracy and ROC-AUC scores higher than other leading machine learning methods. This study focuses on the TNM staging of NSCLC without focusing on SCLC. A notable study by Tyagi et al. [29] introduced LCSCNet (Lung Cancer Stage Classification Network), a multi-level 3D deep convolutional

neural network designed to classify T, N, and M labels. This network achieved average accuracies of 96.23% for T-Stage, 97.63% for N-Stage, and 96.92% for M-Stage classification, along with an overall stage classification accuracy of 97%.

Our research aims to develop a reliable method for staging lung cancer based on TNM criteria using advanced image analysis techniques. By incorporating the TNM staging system into our detection framework, we aim to provide a comprehensive tool for lung cancer diagnosis and staging, thereby improving clinical decision-making and patient management.

A. CONTRIBUTIONS

The main contributions of this paper are as follows:

- Developed a comprehensive method for lung cancer subtype classification and staging using a 3D volumetric CT imaging approach.
- Utilized the latest YOLO architecture for lung cancer subtype classification (SCLC, ADC, and SCC), achieving high precision and speed. Extensive experiments were conducted to validate the performance of the proposed method against baseline models such as YOLOv6, YOLOv7, and Faster R-CNN, as well as other models specific to lung cancer.
- Designed a neural network-based classifier for TNM (Tumor, Node, Metastasis) staging, leveraging features extracted from the YOLO backbone, and evaluation of the models compared to state-of-art.

II. METHODOLOGY

The methodology of our research encompasses several critical steps to develop a reliable and efficient system for lung cancer subtype classification and TNM staging using 3D volumetric CT imaging. The overall process is illustrated in Figure 1, starting from a DICOM study the volumetric CT where extracted and preprocessed, the 2D slices images were fed to object detection, and the tracking algorithms, once we have a detection of the abnormality (Lung Cancer Subtypes) an id is given to this object to start tracking the whole study, finally we end with **i** number of slices with the predicted bounding box, the highest area of the bounding box where chosen and the feature extraction and reduction for this slices where made then fed to another classifier to make the prediction of the of TNM stages. In the rest of this paper, first, we will discuss every part of the systems briefly, after that in the Evaluation section we will test the effectiveness of the system proposed by evaluating each models; object detection, and the classifier using different metrics.

A. DATASETS PREPARATION

1) LUNG PET-CT-DX

The Lung PET-CT-DX [12] is a publicly available database provided by the Cancer Imaging Archive (TCIA) [3]. This dataset comprises DICOM images, including CT and PET-CT scans of individuals with lung cancer, accompanied by XML annotation files indicating the location of tumors through boundary boxes around regions of interest. The data was retrospectively acquired from 350 patients clinically suspected of lung cancer, who subsequently underwent standard-of-care lung biopsy and PET/CT scans. Each study within this dataset includes one CT volume, one PET volume, and fused PET and CT images. Expert radiologists, with varying levels of experience, annotated the location of each tumor, ensuring a comprehensive dataset. The project focuses on CT volumes, which are categorized as follows: Adenocarcinoma denoted by 'A,' consisting of 264 studies; Squamous Cell Carcinoma denoted by 'G,' comprising 62 studies; and Small Cell Carcinoma denoted by 'B,' 44 studies. Additionally, an Excel file contains detailed descriptions of each patient, including TNM staging as represented in Figure 5. Each study yields multiple slices, each annotated with a boundary box around the region of interest and properly classified into the designated group as illustrated in Figures 2,3,4.

For the purposes of this research, all DICOM studies were converted to JPEG images with sizes $512 \times 512 \times 3$. From the dataset, 170 studies from Group A, 30 studies from Group B, and 50 studies from Group G were selected to form the training set. The remaining images were allocated to the testing set.

TABLE 1. Distribution of training and testing sets on the lung PET-CT-DX dataset.

Group	Total Studies	Training Studies	Testing Studies
A	264	170	94
G	62	50	12
B	44	30	14

2) NSCLC DATASET (LUNG3)

The Lung3 [40] dataset, available through the Cancer Imaging Archive (TCIA), is a comprehensive public resource. It consists of DICOM images, including CT scans of patients with non-small cell lung cancer (NSCLC), and is supplemented with clinical data detailing each patient's TNM stage, tumor location, and group affiliation. The dataset was retrospectively gathered from 89 patients who had undergone surgical treatment for NSCLC, with each study containing one CT volume. For this research, the dataset was used to validate the proposed methodology. After excluding irrelevant classes, we focused on 54 studies that are pertinent to our investigation. Similarly to the first dataset, all DICOM studies were converted to JPEG images with sizes $512 \times 512 \times 3$.

3) DATA AUGMENTATION

The Lung PET-CT-DX dataset, as mentioned in [12], had an imbalanced distribution among patient groups: 70% belong to group A, 18% to group B, and 12% to group G. After data extraction from the training set, this results in 3,784 instances for group A, 472 slices for group B, and 754 slices for group G. To address this imbalance, data augmentation techniques were introduced to increase the

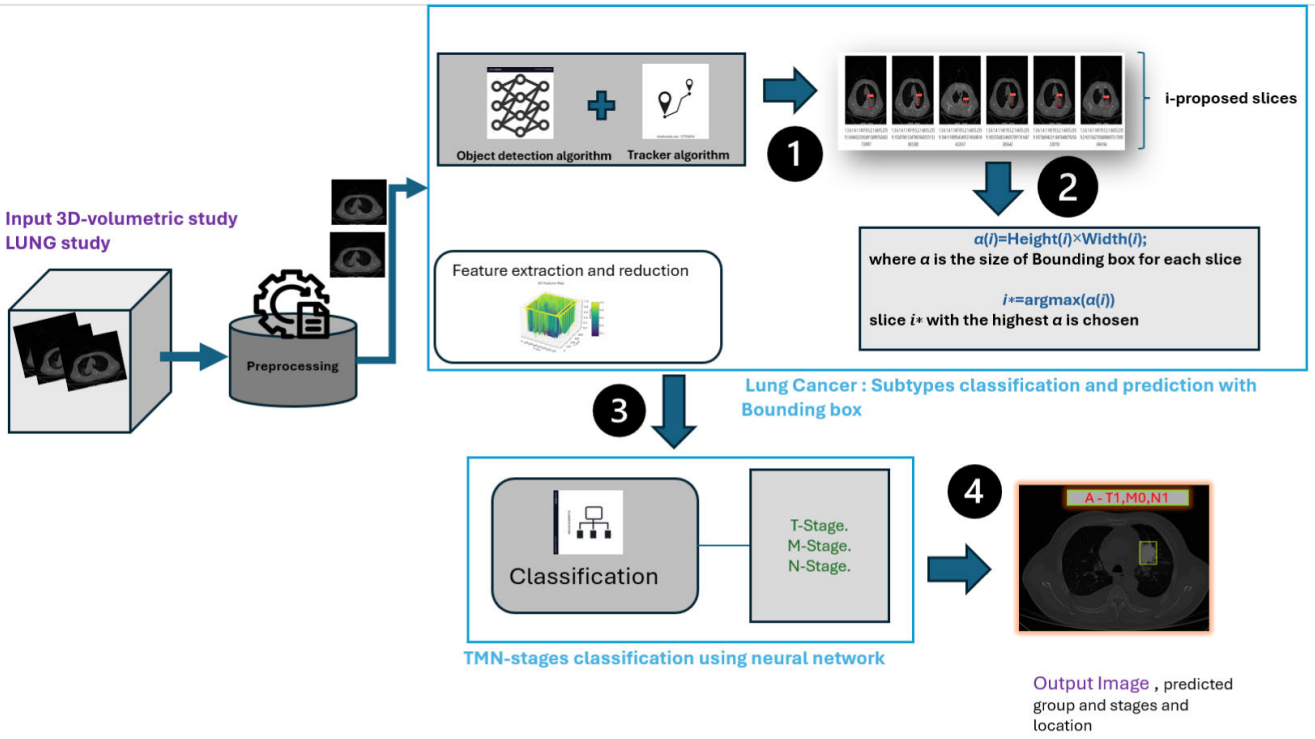

FIGURE 1. Proposed frame work.

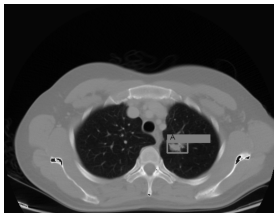

FIGURE 2. Group A: CT lung cancer adenocarcinoma with bounding box.

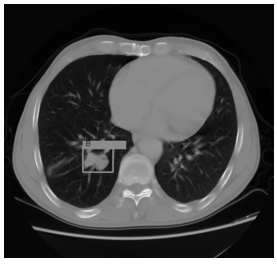

FIGURE 3. Group B: CT lung cancer small cell with bounding box.

FIGURE 4. Groupe G:CT lung squamous cancer cell with bounding box.

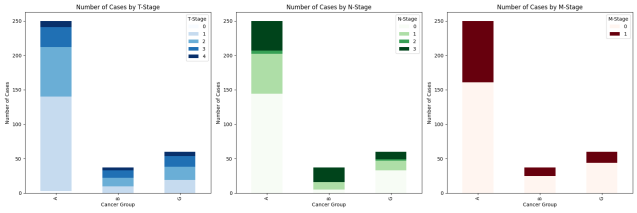

FIGURE 5. Distribution of Cases by Cancer Group and Stages (T, N, M). The subplots represent the number of cases for each T, N, and M stage, respectively, across different cancer groups.

number of samples in the minority classes and improve model generalization.

The data augmentation techniques employed two image processing methods using Roboflow [15], a framework for creating computer vision models:

1) **Gaussian Noise Addition:**

$$I_{\text{aug}}(m, n) = I(m, n) + N(\mu = 0, \sigma = 0.05) \quad (1)$$

where $N(\mu, \sigma)$ denotes Gaussian noise with mean μ and standard deviation σ. Adding Gaussian noise helps to simulate real-world variations and imperfections, enhancing the model's robustness to noise and small

variations in image quality. This technique encourages the model to learn more generalized features rather than memorizing the exact training data.

2) **Rotation:**

$$I_{\text{aug}}(m, n) = \text{Rotate}(I(m, n), \text{angle} = 5^\circ) \quad (2)$$

Rotation is performed with an angle of 5°. This method introduces slight variations in the orientation of the images, which helps the model become invariant to small changes in orientation. A rotation angle of 5 degrees is chosen to ensure that the images remain similar to the originals, avoiding excessive distortions while still providing useful variations for training.

Group B, for instance, is augmented 8 times using these techniques, while group G undergoes augmentation 5 times. This strategy helps to balance the number of instances among the groups and mitigates the impact of class imbalance. Post-augmentation, the training set contains:

A (ADC): 3,784 instances
B (SCLC): 3,770 instances
G (SCC): 3,776 instances

B. OBJECT DETECTION AND TRACKING

1) YOLOv8

It was in 2015 that Redmon et al. [15] created the object detection algorithm You Only Look Once. An input image is divided into SxS grid cells by the algorithm which then forecasts B bounding boxes and the class probabilities that correspond to each cell. By converting the object detection models problem into a regression problem Yolo can detect objects quickly and accurately [26].

YOLO through the last year has gone through continuous improvements and optimization, until we have the YOLO version eight that was released in 2023 by Ultralytics [19]. Even though no clear paper describes the YOLOv8 architecture, the network structure has already been partially explained. The Yolov8 is formed from three essential parts; the backbone, the Neck, and the head. The backbone is a deep-learning architecture that acts as a feature extractor. It is formed from a normal convolution layer followed by SiLu activation function and C2F blocks which contain bottleneck blocks with two types; one Similar to the ResNet [28] uses skip connection and the other without using skip connections. However the neck is not explicitly mentioned in the YOLOv8 documentation, but if we break down the algorithm we see that the model used Spatial Pyramid Pooling Fast (SPPF) which is a modification of Spatial Pyramid Polling (SPP) with a higher speed, C2F blocks and concat blocks. Finally, the neck combines the features acquired from the various layers of the backbone model. The head predicts the classes and the bounding box regions which is the final output produced by the object detection model. We have the detection block when the detection happens, different from the previous version, Yolov8 is an anchor-free model the prediction happens in the grid cell.

One notable advantage of YOLOv8 over its predecessors is its prediction on three levels—high resolution for smaller objects, medium resolution for mid-sized objects, and lower resolution for larger objects. Therefore, with the notable improvement in both speed, by introducing SPPF, and accuracy, owing to the adaptation of the anchor-free model, Yolov8 was chosen along with other object detection models.

2) YOLO TRACKING AND TRAINING

ByteTrack [33] is a robust multi-object tracking (MOT) algorithm designed to address the challenges of occlusion and motion blur in video sequences. Traditional MOT methods often discard low-confidence detection boxes, leading to missed detections and fragmented trajectories. ByteTrack, however, introduces a novel approach by associating nearly all detection boxes, including those with low confidence scores, to enhance tracking performance. The key innovation of ByteTrack lies in its two-stage data association process. Initially, high-confidence detection boxes are matched with tracklets based on motion or appearance similarity. In the second stage, unmatched low-confidence detection boxes are matched to the remaining tracklets. This method ensures that true objects, even those with low detection scores due to occlusion or motion blur, are correctly associated and tracked, while background detections are filtered out. When combined with YOLO, ByteTrack leverages YOLO's high detection accuracy and enhances tracking performance by effectively managing detection boxes with varying confidence levels. For this project, we utilized the latest version of YOLO—version 8 developed by Ultralytics [19]. YOLOv8 offers a suite of models tailored to diverse requirements, including Nano, Small, Medium, Large, and Extra Large options. These models provide varying trade-offs between computational efficiency and accuracy, allowing us to select the most suitable model for our specific needs. The collected dataset was partitioned into training and testing sets, as shown in Table 1. For each study, the entire study was included in the training phase, with both annotated and non-annotated images used as background. The training set was then divided into 80% training and 20% validation.

C. BOUNDING BOX SELECTION

For each proposed slice, a bounding box is determined. The area of the bounding box is calculated as follows:

$$\alpha(i) = \text{Height}(i) \times \text{Width}(i) \quad (3)$$

The slice with the highest bounding box area is determined by:

$$i^* = \arg\max(\alpha(i)) \quad (4)$$

is selected for further processing. This ensures that the most significant nodule is considered since distinguishing the TNM stages requires a more detailed study.

D. FEATURE EXTRACTION AND REDUCTION

After selecting the most significant slices from each CT scan, features were extracted using the backbone of a pre-trained YOLO model. This process captures high-dimensional data representing the characteristics of the CT images, as presented in the feature extraction code [32]. To reduce computational complexity and retain essential features, Principal Component Analysis (PCA) was applied. PCA transforms the high-dimensional data into a lower-dimensional space by projecting it onto principal components, which preserve the most significant variance. Let X be the high-dimensional data matrix, then PCA projects X onto a lower-dimensional space:

$$X_{\text{reduced}} = XW \tag{5}$$

where W is the matrix of principal components. This reduced feature set serves as the input for subsequent classifiers, ensuring efficient and effective TNM stage classification.

E. TNM STAGING CLASSIFICATION

The TNMClassifier model is designed to classify lung cancer stages based on Tumor (T), Node (N), and Metastasis (M) criteria. The model uses a series of fully connected layers to process the PCA-reduced features extracted from CT images.

The architecture presented in Figure 6 is composed of three fully connected layers with 512, 256, and 128 neurons, respectively. Each layer uses the ReLU activation function to introduce non-linearity. A dropout layer with a dropout rate of 0.5 is included after the third fully connected layer to prevent overfitting. The final classification is achieved through three separate output layers, each corresponding to one of the TNM stages. These output layers use the softmax activation function to produce probability distributions over the respective classes.

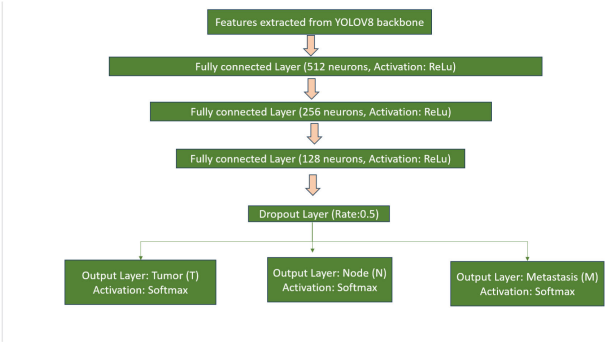

FIGURE 6. TNMClassifier neural network architecture.

III. EXPERIMENTAL EVALUATION

In this section, we first introduce the experimental settings of the YOLO, then conduct a set of experiments to compare our model with other leading object detection models and methods used in the literature in terms of Recall, Precision, F1 score, mAP, and speed of detection. The evaluation is divided into two parts: the first part is dedicated to the evaluation of the model used for the classification and localization of the Lung Cancer Subtype, and the second part refers to the evaluation of the TNM stage classification.

Note: When we show the results of our models it's all dedicated and made on the testing set.

A. EXPERIMENTAL SETTINGS

The YOLOv8 is based on Pytorch v2.1.1 deep learning network and trained on NVIDIA RTX 4050. The programming language of this model is Python; GPU is accelerated using CUDAv12.1 and CuDNNv8.9.6. The model parameter settings are shown in Table 2 below.

TABLE 2. Model parameters settings.

Parameters	Values
Learning Rate	0.01
Epoch	200
Batch Size	16

B. LUNG CANCER SUBTYPE CLASSIFICATION AND LOCALIZATION EVALUATION

In this section, we present the training results of the YOLO v8 model used for the detection and classification of lung cancer subtypes. The model's performance was evaluated based on various metrics including box loss, classification loss, precision, recall, and mean Average Precision (mAP) across different epochs.

The results indicate that the model's performance improved significantly over the training period, achieving high precision and recall values, and steadily increasing mAP metrics. The loss values for both box and classification tasks decreased consistently, indicating effective learning and optimization of the model. The performance of YOLOv8 was evaluated using different models architectures for instance in Figure 9; Nano, Small, Medium, and Large. Different metrics were used including precision, Recall, mAP, F1 score, and speed of detection to evaluate those models. The experimental results in Table 3 show that all the models achieved a high performance and there is no big difference between them. We mark that the YOLOv8-Small achieved the highest mAP with 0.971, and YOLOv8-Nano had the lowest time of detection with 0.119s per image. But when we compare those two models we highlight that YOLO-small gets higher precision and recall. Therefore we have conducted another deep study on the YOLO-small and we calculated the confusion matrix to see how the model perform on each group.

When we looked at the confusion matrix on the testing set for Yolov8s (Figure 8) to understand the behavior of the model better. The matrix demonstrated how well the model identified cases of squamous cell carcinoma (G) and adenocarcinoma (A), with high True Positives (549 and 580, respectively). With 435 True Positives, the model also showed proficiency in detecting Small Cell Carcinoma (B). False Positives in the True Background row were the result

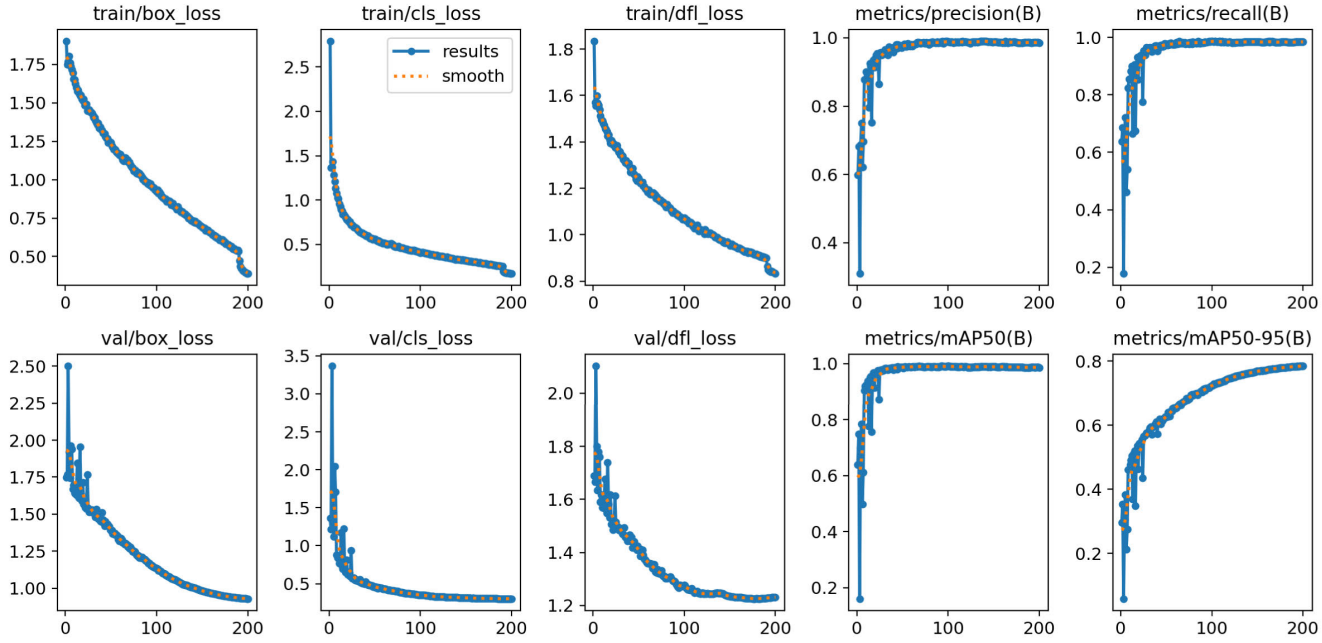

FIGURE 7. Training results of the YOLO v8 model showing the box loss, classification loss, precision, recall, and mAP metrics over 200 epochs.

of several misclassifications, especially when predicting Background cases. These results highlight the necessity of ongoing improvement and optimization and the importance of correctly identifying between non-cancerous areas and different forms of lung cancer.

TABLE 3. Performance metrics and speed of detection for different models.

Model	Precision	Recall	mAP at 0.5	F1 Score	Speed (s)
YOLO nano	0.954	0.94	0.97	0.946	**0.119**
YOLO small	**0.961**	**0.945**	**0.971**	**0.952**	0.22
YOLO medium	**0.961**	**0.945**	0.968	**0.952**	0.53
YOLO large	0.957	0.942	0.966	0.949	0.62

1) COMPARISON WITH OTHER OBJECT DETECTION MODELS

After conducting a comparative analysis of various object detection models on the same dataset with identical experimental settings, including baseline models such as Faster R-CNN, YOLOv6, and YOLOv7, and models specifically used in the domain of lung cancer detection like YOLOv5 Small, and STBi-YOLO we found that the YOLOv8 Small model outperformed all other models in terms of precision,AP at 0.5, and F1 score,except for STBi in terms of recall.The Table 4 summarized the results. When comparing our approach to other state-of-the-art methods, we found that few studies focus on localizing tumors, especially those dedicated to the three targeted groups. To provide a relevant comparison, we selected two studies that used similar CT imaging modalities and chose to retrain those models (SC-Dynamic R-CNN and Center Key Point) on

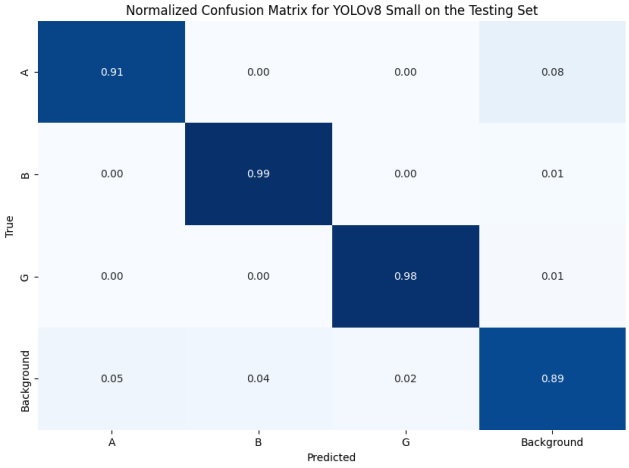

FIGURE 8. Normalized confusion matrix for YOLOv8 Small on the testing set.

the Lung PET-CT-DX dataset. The YOLOv8 model showed superior performance, further validating the effectiveness of our approach. The results are illustrated in Table 4.

C. EVALUATION OF YOLOv8 SMALL MODEL ON THE LUNG3 DATASET

To further validate our approach, we assessed the YOLOv8 small model on the Lung3 dataset, which includes DICOM images and clinical data from 89 patients who underwent surgical treatment for non-small cell lung cancer (NSCLC). Out of these 89 patients, 54 belong to NSCLC subtypes A and G. This evaluation aimed to determine the model's

TABLE 4. Performance metrics for different models on lung PET-CT-DX dataset.

Model	Precision	Recall	mAP at 0.5	F1 Score
Faster-RCNN	0.914	0.891	-	0.9
YOLOv6	0.933	0.932	0.957	0.93
YOLOv7	0.94	0.92	0.945	0.92
YOLOv5 small [38]	0.922	0.894	0.936	0.907
SC-Dynamic RCNN [10]	0.75	0.78	0,85	0,76
STBi-YOLO [39]	0.931	**0.951**	-	0.94
Center key point [11]	0.71	0.821	0.851	0.761
YOLOv8 small	**0.961**	0.945	**0.971**	**0.952**

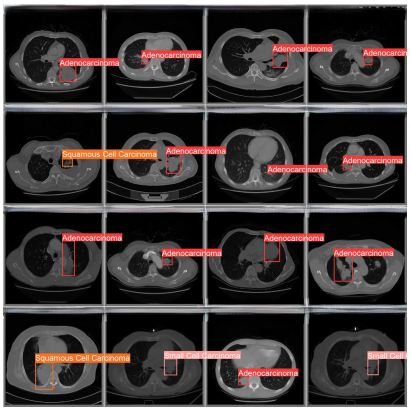

FIGURE 9. Detection and Classification of Lung Cancer Subtypes using YOLOv8. The figure illustrates the bounding boxes and labels for different lung cancer subtypes including Adenocarcinoma, Squamous Cell Carcinoma, and Small Cell Carcinoma across various CT image slices.

effectiveness in detecting and classifying NSCLC in CT scans beyond the initial training set.The YOLOv8 small model, originally trained on the Lung PET-CT-DX dataset, was tested on the Lung3 dataset. The results of this evaluation showed that the model achieved a precision of 0.89 and a recall of 0.91. These metrics demonstrate the model's high effectiveness in identifying NSCLC cases within the Lung3 dataset, indicating its generalizability and robustness in different clinical settings.

D. TNM STAGE CLASSIFICATION EVALUATION

In this section, we present the training and validation loss and accuracy for the TNM classification model. The model was trained over 30 epochs, and the resulting performance metrics are shown in Figure 10. Similarly to the training of the YOLO model, the data were divided into three-part training, validation, and testing while taking in consideration that in the testing set, we don't have slices that belong to the same patient in training and testing. During this process the features where extracted from the pre-trained YOLO and applied the feature reduction using PCA and fed to the TNMclassifier.

1) CLASSIFICATION RESULTS

After training the model, we evaluated it using precision, recall, and F1-score for each TNM classification stage on the Lung PET-CT-DX dataset. The results are presented in Tables 5, 7, and 13 in the testing set. For T-stage classification,

TABLE 5. T classification report.

T-Stage	Precision	Recall	F1-score	Slices
1	0.98	0.99	0.99	964
2	0.98	0.97	0.97	686
3	0.98	0.98	0.98	459
4	0.99	0.96	0.98	218
Accuracy	0.98			
Macro avg	0.98	0.98	0.98	2327
Weighted avg	0.98	0.98	0.98	2327

TABLE 6. N classification report.

N-Stage	Precision	Recall	F1-score	Slices
0	0.97	0.99	0.98	984
1	0.99	0.97	0.98	681
2	0.99	0.97	0.98	80
3	0.97	0.97	0.97	582
Accuracy	0.98			
Macro avg	0.98	0.98	0.98	2327
Weighted avg	0.98	0.98	0.98	2327

TABLE 7. M classification report.

M-Stage	Precision	Recall	F1-score	Slices
0	0.99	0.99	0.99	1474
1	0.97	0.98	0.98	853
Accuracy	0.98			
Macro avg	0.99	0.84	0.88	2327
Weighted avg	0.98	0.98	0.98	2327

TABLE 8. Average results using different supervised classifiers.

Model	Precision	F1-score	Accuracy
SVM	0.86	0.91	0.93
XGBM	0.9	0.93	0.952
KNN	0.861	0.768	0.77
Our method	**0.98**	**0.98**	**0.98**

the model achieved an overall accuracy of 98%. The T1 class showed a precision, recall, and F1-score of 99%. T2 had a precision of 98%, recall of 97%, and F1-score of 97%. T3 achieved precision at 98%, recall of 98%, and F1-score of 98%. T4 had precision of 99%, recall of 96%, and F1-score of 98%.In N-stage classification, the overall accuracy was 98%. N0 had a precision of 97%, recall of 99%, and F1-score of 98%. N1 showed high metrics with precision, recall, and F1-score in average of 98%. N2 achieved perfect precision and recall at 99%, resulting in an F1-score of 98%. N3 had a precision of 97%, recall of 97%, and F1-score of 97%, indicating some class confusion.For M-stage classification, the model also achieved an overall accuracy of 98%. M0 had a precision of 99%, recall of 99%, and F1-score of 99%. M1 showed a precision of 97%, recall of 98%, and F1-score of 98%. We also computed confusion matrices presented in the following(Figures 11,13,12).

2) COMPARISON WITH OTHER AND EXISTING METHODS

Table 8 summarizes the average results (average P, F1-score, accuracy over all the stage and group) of different supervised

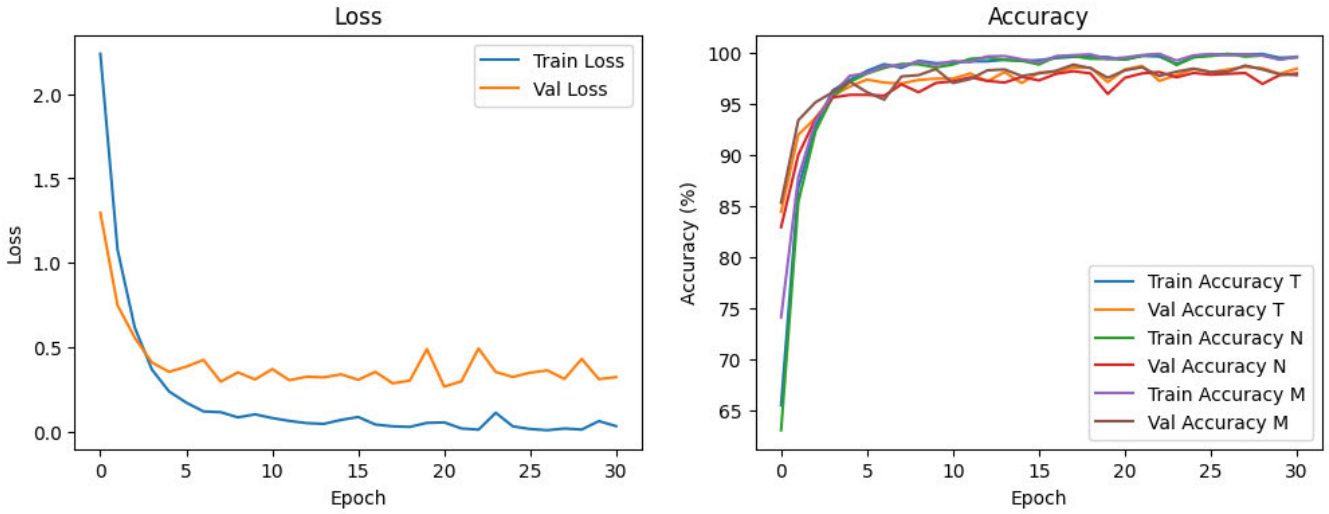

FIGURE 10. Training and validation loss and accuracy for the TNM classifier.

TABLE 9. Comparaision with classification approach.

Method	Dataset	Classification Task	Accuracy
Kirienko et al. [30]	Private data	T-stage as T1/T2 and T3/T4	82.6%
Moitra et al. [31]	NSCLC-Radiogenomics	TNM stage	96%
Tyagi et al. [29]	Lung PET-CT Dx	TNM stage	97%
Our method	Lung PET-CT Dx/Lung3	TNM stage	**98%/91%**

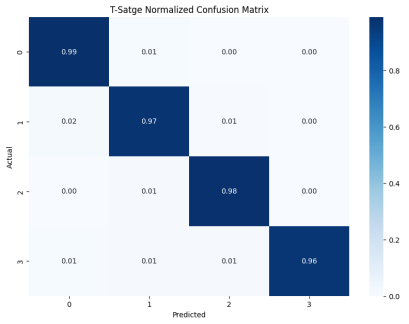

FIGURE 11. T Staging.

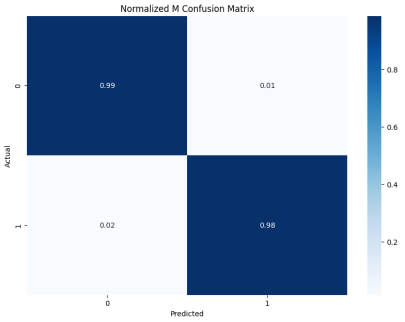

FIGURE 12. M Staging.

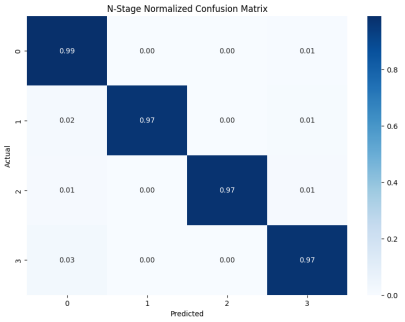

FIGURE 13. N Staging.

classifiers used for TNM stage classification, evaluating precision, F1-score, and accuracy. The Support Vector Machine (SVM) model achieved the lowest performance with a precision of 0.86, an F1-score of 0.91, and an accuracy of 0.93. The Gradient Boost Machine Learning (XGBM) model showed an improvement, with a precision of 0.90, an F1-score of 0.93, and an accuracy of 0.768. K-nearest neighborhoods show decreasing in performance, attaining a precision of 0.861, an F1-score of 0.768, and an accuracy of 0.77. The **proposed architecture** achieved the highest in terms of all metrics and an impressive accuracy with 0.98.

The table 9 shows that different methods and datasets were used across studies. Notably, for the TNM stage classification using the Lung PET-CT Dx dataset and Lung3 datasets, our method achieves the highest accuracy at 98% on the Lung PET-CT-DX dataset, and 91% on the Lung3 dataset. This demonstrates the effectiveness of our approach compared to Tyagi et al. [29], which reported 97% accuracy using the Lung PET-CT Dx dataset. This improvement highlights

the robustness of our feature extraction and classification techniques.

IV. CONCLUSION

In this paper, we introduced an automated classification and detection model for lung cancer using YOLOv8 and the TNMClassifier. We addressed the challenge of imbalanced data by applying data augmentation techniques, which improved model generalization. YOLOv8, particularly the small model variant, demonstrated superior performance with a high mAP of 97.1%, outperforming other object detection models like Faster R-CNN and earlier YOLO versions, as well as other models used in the literature. Additionally, our TNMClassifier effectively classified the tumor, node, and metastasis stages with high precision and recall, achieving an overall accuracy of 98%. These results underscore the efficacy of our integrated approach in enhancing the accuracy and efficiency of lung cancer detection and staging. Future work will focus on the feature extracted from Yolo and design a Content-based Image Retrieval (CBIR) that will justify the classification process by giving additional information from previous cases.

REFERENCES

[1] F. Passiglia, M. Calandri, and F. Guerrera, "Lung cancer in Italy," *J. Thoracic Oncol.*, vol. 14, no. 12, pp. 2046–2052, 2019.

[2] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics 2017," *Cancer J. Clin.*, vol. 67, pp. 7–30, Jan. 2017.

[3] (2023). *Non–Small Cell Lung Cancer*. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK562307

[4] I. Sluimer, A. Schilham, M. Prokop, and B. van Ginneken, "Computer analysis of computed tomography scans of the lung: A survey," *IEEE Trans. Med. Imag.*, vol. 25, no. 4, pp. 385–405, Apr. 2006, doi: 10.1109/TMI.2005.862753.

[5] B. Hochhegger, E. Marchiori, O. Sedlaczek, K. Irion, C. P. Heussel, S. Ley, J. Ley-Zaporozhan, A. S. Souza, and H.-U. Kauczor, "MRI in lung cancer: A pictorial essay," *Brit. J. Radiol.*, vol. 84, no. 1003, pp. 661–668, Jul. 2011.

[6] A. El-Baz, G. Gimel'farb, R. Falk, and M. Abo El-Ghar, "A new CAD system for early diagnosis of detected lung nodules," in *Proc. IEEE Int. Conf. Image Process.*, vol. 2, Oct. 2007, pp. 461–464, doi: 10.1109/ICIP.2007.4379192.

[7] K. Manikandan, "Blob based segmentation for lung CT image to improving CAD performance," in *Proc. Int. Conf. Recent Trends Inf. Technol.*, Apr. 2014, pp. 1–6, doi: 10.1109/ICRTIT.2014.6996157.

[8] N. A. Pande and D. Bhoyar, "A comprehensive review of lung nodule identification using an effective computer-aided diagnosis (CAD) system," in *Proc. 4th Int. Conf. Smart Syst. Inventive Technol. (ICSSIT)*, Tirunelveli, India, Jan. 2022, pp. 1254–1257, doi: 10.1109/ICSSIT53264.2022.9716327.

[9] R. Indumathi and R. Vasuki, "Lung cancer detection using CAD system," in *Proc. 3rd Int. Conf. Advance Comput. Innov. Technol. Eng. (ICACITE)*, May 2023, pp. 206–210, doi: 10.1109/ICACITE57410.2023.10182740.

[10] X. Wang, L. Wang, and P. Zheng, "SC-dynamic R-CNN: A self-calibrated dynamic R-CNN model for lung cancer lesion detection," *Comput. Math. Methods Med.*, vol. 2022, pp. 1–9, Mar. 2022, doi: 10.1155/2022/9452157.

[11] J. Qi, Z. Deng, G. Sun, S. Qian, L. Liu, and B. Xu, "One-step algorithm for fast-track localization and multi-category classification of histological subtypes in lung cancer," *Eur. J. Radiol.*, vol. 154, Sep. 2022, Art. no. 110443.

[12] P. Li, S. Wang, T. Li, J. Lu, F. Huang, and D. Wang, "A large-scale CT and PET/CT dataset for lung cancer diagnosis (lung-PET-CT-Dx) [Data set]," *Cancer Imaging Archive*, 2020.

[13] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, L. Tarbox, and F. Prior, "The cancer imaging archive (TCIA): Maintaining and operating a public information repository," *J. Digit. Imag.*, vol. 26, no. 6, pp. 1045–1057, Dec. 2013.

[14] D. Mason, "pydicom/pydicom: Pydicom v2.4.0," Zenodo, Jun. 2023, doi: 10.5281/zenodo.8034250.

[15] B. Dwyer and J. Nelson. (2022). *Roboflow (Version 1.0)*. [Online]. Available: https://roboflow.com

[16] S. Pang, Y. Zhang, M. Ding, X. Wang, and X. Xie, "A deep model for lung cancer type identification by densely connected convolutional networks and adaptive boosting," *IEEE Access*, vol. 8, pp. 4799–4805, 2020, doi: 10.1109/ACCESS.2019.2962862.

[17] C. Li, L. Li, H. Jiang, K. Weng, Y. Geng, L. Li, Z. Ke, Q. Li, M. Cheng, W. Nie, Y. Li, B. Zhang, Y. Liang, L. Zhou, X. Xu, X. Chu, X. Wei, and X. Wei, "YOLOv6: A single-stage object detection framework for industrial applications," 2022, *arXiv:2209.02976*.

[18] S. Wang, L. Dong, X. Wang, and X. Wang, "Classification of pathological types of lung cancer from CT images by deep residual neural networks with transfer learning strategy," *Open Med.*, vol. 15, no. 1, pp. 190–197, Mar. 2020.

[19] G. Jocher, A. Chaurasia, and J. Qiu. (2023). *Ultralytics*. [Online]. Available: https://github.com/ultralytics/ultralytics

[20] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.

[21] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, Chile, Dec. 2015, pp. 1440–1448, doi: 10.1109/ICCV.2015.169.

[22] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017, doi: 10.1109/TPAMI.2016.2577031.

[23] J. Dai, "R-FCN: Object detection via region-based fully convolutional networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 1–29.

[24] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988, doi: 10.1109/ICCV.2017.322.

[25] Q. Shuai and X. Wu, "Object detection system based on SSD algorithm," in *Proc. Int. Conf. Culture-Oriented Sci. Technol. (ICCST)*, Oct. 2020, pp. 141–144, doi: 10.1109/ICCST50977.2020.00033.

[26] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788, doi: 10.1109/CVPR.2016.91.

[27] C.-Y. Wang, A. Bochkovskiy, and H.-Y. Mark Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 7464–7475.

[28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[29] S. Tyagi and S. N. Talbar, "LCSCNet: A multi-level approach for lung cancer stage classification using 3D dense convolutional neural networks with concurrent squeeze-and-excitation module," *Biomed. Signal Process. Control*, vol. 80, Feb. 2023, Art. no. 104391.

[30] M. Kirienko, M. Sollini, G. Silvestri, S. Mognetti, E. Voulaz, L. Antunovic, A. Rossi, L. Antiga, and A. Chiti, "Convolutional neural networks promising in lung cancer T-parameter assessment on baseline FDG-PET/CT," *Contrast Media Mol. Imag.*, vol. 1, pp. 1–6, Oct. 2018.

[31] D. Moitra and R. Kr. Mandal, "Classification of non-small cell lung cancer using one-dimensional convolutional neural network," *Expert Syst. Appl.*, vol. 159, Nov. 2020, Art. no. 113564.

[32] *YOLOv8 Feature Extraction*. Accessed: Jan. 2024. [Online]. Available: https://github.com/Alaawehbe12/Feature-extraction-YOLOv8

[33] Y. Zhang, "Bytetrack: Multi-object tracking by associating every detection box," in *Proc. Eur. Conf. Comput. Vis.*, 2022, pp. 1–26.

[34] K. Xie, S. Ge, Q. Ye, and Z. Luo, "Traffic sign recognition based on attribute-refinement cascaded convolutional neural networks," in *Advances in Multimedia Information Processing-PCM 2016*. Cham, Switzerland: Springer, 2016, pp. 201–210.

[35] P. Sun, R. Zhang, Y. Jiang, T. Kong, C. Xu, W. Zhan, M. Tomizuka, L. Li, Z. Yuan, C. Wang, and P. Luo, "Sparse R-CNN: End-to-end object detection with learnable proposals," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 1–20.

[36] T. Liang, H. Bao, W. Pan, and F. Pan, "Traffic sign detection via improved sparse R-CNN for autonomous vehicles," *J. Adv. Transp.*, vol. 2022, pp. 1–16, Mar. 2022.
[37] S. Du, W. Pan, N. Li, S. Dai, B. Xu, H. Liu, C. Xu, and X. Li, "TSD-YOLO: Small traffic sign detection based on improved YOLO v8," *IET Image Process.*, vol. 18, no. 11, pp. 2884–2898, Sep. 2024.
[38] K. P. Gunasekaran, "Leveraging object detection for the identification of lung cancer," 2023, *arXiv:2305.15813*.
[39] K. Liu, "STBi-YOLO: A real-time object detection method for lung nodule recognition," *IEEE Access*, vol. 10, pp. 75385–75394, 2022.
[40] H. J. W. L. Aerts et al., "Data from NSCLC-radiomics," Cancer Imag. Arch., 2019, doi: 10.7937/K9/TCIA.2015.PF0M9REI.

ALAA WEHBE was born in Lebanon. He received the bachelor's degree in electronics engineering and the master's degree in signal, telecom, and image processing (STIP) from Lebanese University, in 2021 and 2022, respectively. He is currently pursuing the Ph.D. degree with the University of Genoa, Italy.

SILVANA DELLEPIANE was born in Greenwich Village. She received the Laurea (M.Sc.) degree (cum laude) in electronic engineering and the Ph.D. degree in electronic and computer engineering, in 1986 and 1990, respectively. She became a University Researcher with the Universita degli Studi di Genova, in 1992, where she is currently an Associate Professor. She has been teaching signal theory, recognition theories and techniques, and statistical methods. She is currently teaching electrical communications and digital image processing in the courses of telecommunications engineering, biomedical engineering, and history of art and preservation of cultural heritage. She has been responsible for signal analysis during rehabilitation in the Rehab@home EU Project. She is a member of the IEEE BISP Technical Committee and an Honorary Member of SISS—Societa Italiana per lo Studio dello Stroke.

IRENE MINETTI received the Ph.D. degree in electronic, computer science, robotics, and telecommunications engineering from the University of Genoa in 2011, master's degree in telecommunications engineering (summa cum laude) in 2006. She has over a decade of experience at EBIT, a company within the ESAOTE group, where she initially worked on designing turnkey Healthcare IT solutions. Since 2014, she has been the Radiology Product Manager, focusing on strategic marketing, product development, and supporting regulatory and software quality functions. She is actively involved in scientific research, dissemination, and engagement with Key Opinion Leaders. She is also a co-author of multiple publications and patents.

Open Access funding provided by 'Università degli Studi di Genova' within the CRUI CARE Agreement