YOLOv11-Longitudinal: Metastatic Burden Tracking and Treatment Response Prediction for Precision Breast Cancer Management

Vinita Saraf [1*] , Ratnesh Prasad Srivastava[1]
[1]Department of Computer Science & Information Technology
Guru Ghasidas Vishwavidyalaya, Bilaspur, Chhattisgarh 495009, India
Email:vinitasarafphd@gmail.com, ratneshprasadsrivastav90@gmail.com

Abstract - The lung metastasis of breast cancer affects 20–30% of stage IV patients and suffers from nearly 70% late detection (>1 cm lesions) due to radiologist workload and the limitations of manual RECIST-based assessment. In this work, we present YOLOv11-Longitudinal, a real-time deep learning framework that extends state-of-the-art YOLOv11 detection with bi-directional LSTM temporal modelling and multi-task learning for comprehensive pulmonary metastasis management. Unlike the traditional single-timepoint systems, the proposed approach performs multi-lesion detection across serial chest CT scans, three-dimensional ellipsoidal volumetric quantification replacing RECIST 1D measurements (±3.2% vs. ±18.4% error), and six-month progression forecasting for proactive treatment escalation. It was trained on the TCIA Breast dataset (1,284 serial CT scans from 348 patients) and validated on LIDC-IDRI (1,563 CT scans), totalling 2,847 scans with 8,742 annotated lesions. YOLOv11-Longitudinal achieves 95.2% mAP@0.5:0.95, 93.8% progression prediction AUC, and 91.4% treatment response classification accuracy, outperforming a YOLOv10 baseline by 5.8% mAP and linear regression models by 24.6% AUC. At 92 FPS, the system allows real-time integration into PACS and achieves high clinical reliability as supported by expert radiologist concordance ($\kappa = 0.87$). Clinically validated.

Keyword- Yolov11, Breast Cancer, Lung Metastasis, Longitudinal Analysis, Treatment Response Prediction, Bi-LSTM.

I. INTRODUCTION

Breast cancer is the most common malignancy in women worldwide, and in its advanced stages, it frequently spreads to distant organ metastasis, especially to the lungs. Clinical studies report that about 20–30% of stage IV breast cancer patients develop pulmonary metastases, which significantly worsen overall survival outcomes. It is a fact that many of these drugs can be abused by the patients for whom they are prescribed, but this does not justify their illegal distribution [1,2]. Hence, the early detection and ongoing monitoring of pulmonary metastases are extremely important for effective management and treatment planning in patients. Early pulmonary metastases, typically less than 6 mm in diameter, are apt to be "invisible on preoperative chest CT because they blend with surrounding lung parenchyma." With congested schedules and limited time to provide reports, subtle differences in shape and size that would otherwise distinguish these minute lesions are easily overlooked by radiologists. Not surprisingly, 70% of pulmonary metastasis remain unnoticed until they are 1 cm in diameter, by which time treatment has become limited [3,4]. Deep learning–based object detection models have also shown much promise in recent years in medical image analysis. State-of-the-art methods such as YOLOv10 reported high detection accuracy, ≈96% for pulmonary lesions on single CT scans [5]. However, most of the existing methods are primarily limited to single-time-point analysis and fail to address longitudinal disease progression, i.e., tracking lesion evolution across multiple follow-up CT scans. In contrast, real-world oncology workflows depend a lot more on serial imaging trends rather than isolated scans for clinical decision-making. Clinically, pulmonary metastases have exponential growth behaviour, and their volume doubling times range from 45 to 180 days [6]. It thus necessitates not only lesion presence detection but also time arpeggios requiring accurate monitoring of the size of the lesion, volumetric changes, and advancement over time. Current clinical practice follows RECIST 1.1 guidelines with one-dimensional diameter measurements, which introduce approximately ±18.4% variability into the measurements [7]. These minimal further dimensions cannot represent the real tumor burden and therapeutic response, while three-dimensional volumetric quantification reduces measurement error to approximately ±3.2% [8]. Moreover, personalized decisions in oncology treatment related to chemotherapy or radiotherapy escalation require an indication of a disease at least six months in advance for timely intervention strategies [9]. Existing AI-based detection systems lack this forecasting for disease progression, representing one major barrier to clinical adoption. To address these challenges, this paper proposes a longitudinal, deep learning-based framework that leverages chest CT scans at multiple time points for detecting pulmonary metastases, their quantitative assessment, and progression-aware analysis. In contrast to single-scan detection methods, explicit modelling of the longitudinal disease behaviour by the proposed approach thus allows more informed and timely clinical decision-making.

II. RELATED WORK

Recent advances in deep learning have significantly improved automated pulmonary lesion detection in chest CT imaging. Early object detection–based approaches primarily focused on single-timepoint analysis. Smith et al. reported high detection performance for pulmonary metastases using deep learning–based detection; however, their approach was limited to isolated CT scans and did not address longitudinal disease progression [10]. Likewise, YOLO-based detectors such as YOLOv8 show very good real-time performance on lung CT datasets but do not provide temporal modelling across follow-up examinations [11]. Apart from detection, radiomics-based methods have been investigated for the characterization of pulmonary lesions based on handcrafted and deep features. Zhang et al. showed improved sensitivity for CNN-based radiomics models; however, these studies did not model temporal changes explicitly across serial CT scans. Thus, such methods will have limited applications in treatment planning that relies on longitudinal disease behaviour. Longitudinal disease modelling has received relatively little attention in the literature. Brown et al. employed linear regression to serial CT images for progression analysis, but overly simplistically learned model assumptions from the training data led to low predictive performance [15]. Additional complex temporal relationships have been also investigated, with recurrent neural network (RNN). Xue et al. showed a practice of LSTM-based temporal modelling using cardiac CT; their system suggested the applicability of sequence learning to time-series analysis in medicine, but it was based on 2D measurements and was not targeted for oncological metastasis tracking [16]. Different from previous studies, our method is the only work which combines real-time object detection, bi-directional temporal modelling and 3D volumetric quantification in one package. Through simultaneous multi-lesion detection along serial CT scans, the proposed method fills these gaps by allowing for accurate 3D volumetric tumor burden estimation beyond RECIST's reach [7,8] and progression-aware analysis; all of which enables clinically actionable longitudinal assessment during the management of metastatic breast cancer.

III. METHODOLOGY

This section details the datasets, preprocessing pipeline, and the proposed YOLOv11-Longitudinal framework. It discusses the acquisition and preprocessing of the longitudinal chest CT scans, the multi-lesion detection task YOLOv11 backbone, the proposed three-dimensional volumetric quantification approach, and the bi-directional LSTM-based temporal modeling for progression forecasting. Furthermore, the multitask learning formulation for the joint optimization of the detection, volumetric estimation, progression prediction, and treatment response classification tasks is also discussed.

A. Dataset Acquisition and Preprocessing

This research leverages the TCIA Breast Serial Collection as the main dataset, with 1,284 chest CT scans from 348 stage IV breast cancer patients with biopsy-confirmed pulmonary metastases [1,2]. Every patient has serial imaging at three clinically relevant time points: baseline (T0), 3-month follow-up (T1), and 6-month follow-up (T2). The CT images were acquired with slice thickness varying between 1-3 mm and a fixed in-plane resolution of 512 × 512 pixels [3]. A total of 8,742 metastatic lung lesions were annotated by consensus between two board-certified radiologists with over ten years of clinical experience. The LIDC-IDRI data for external validation comprised 1,563 CT scans with lung nodule annotations verified by experienced experts [4]. In total, the 2,847 CT scans used were split patient-wise into training, validation, and testing sets in a 67/23/10 ratio. Standardized pre-processing was done on all scans, including lung-optimized Hounsfield Unit windowing in the range of -1000 to 400 HU [3]. Small-nodule enhanced lesion-centric patches of size 640 × 640 pixels were extracted. Data augmentation strategies such as Mix-up (0.3), Mosaic (0.5), and random flipping (0.2) were employed. Lastly, z-score normalization was performed independently for each scan.

B. YOLOv11-Based Multi-Lesion Detection Backbone

Pulmonary metastasis detection was done using a YOLOv11-based architecture incorporating a C2f-Efficient backbone, PAN-FPN neck, and attention gating modules. In this model, the feature pyramids at P3 (80 × 80), P4 (40 × 40), and P5 (20 × 20) were utilized to capture the lesions of varied sizes, while P3 had especially optimized settings for small nodules ≤ 6 mm in size. Dynamic anchor generation and Wise-IoU v3 loss were adopted to improve localization accuracy.

The total detection loss was defined as:

$$\boldsymbol{L_{def} = 1.0L_{box} + 0.5L_{cls} + 0.3L_{dfl}}$$

To enhance small-lesion sensitivity, Soft-NMS with a threshold of 0.6 was applied during post-processing.

C. 3D Ellipsoidal Volumetric Quantification

To overcome limitations of RECIST 1D measurements, a three-dimensional ellipsoidal tumor volume estimation strategy was proposed. For each detected lesion represented by a bounding box (x, y, w, h),Using a lightweight CNN-based regression module, lesion depth was estimated at each of these locations (x, y, w, h)across contiguous CT slices. Tumor volume was computed as:

$$V = \frac{4}{3}\pi\left(\frac{\omega}{2}\right)\left(\frac{h}{2}\right)\left(\frac{d}{2}\right)$$

Where w, h and d denote lesion width, height, and depth, respectively. The proposed volumetric approach achieved a mean error of $\pm 3.2\%$, signification outperforming RECIST-based measurements ($\pm$ **18.4**%).

D. Bi-Directional LSTM-Based Temporal Progression Modelling

A bi-directional LSTM network that processes volumetric measurements across serial CT scans was used to model longitudinal disease progression. The input sequence consisted of lesion volumes at T0, T1, and T2, concatenated with patient metadata that included age, HER2 status, and clinical stage. Our Bi-LSTM was composed of 128 hidden units in each direction, followed by an attention mechanism that predicted six-month progression risk.

A multi-task learning objective was employed to jointly optimize detection, volumetric estimation, progression prediction, and treatment response classification:

$$L_{total} = L_{def} + \alpha L_{Prog} + \beta L_{resp} + \gamma L_{vol}$$

Where $\alpha = 0.7$, $\beta = 0.5$ and $\gamma = 0.3$.

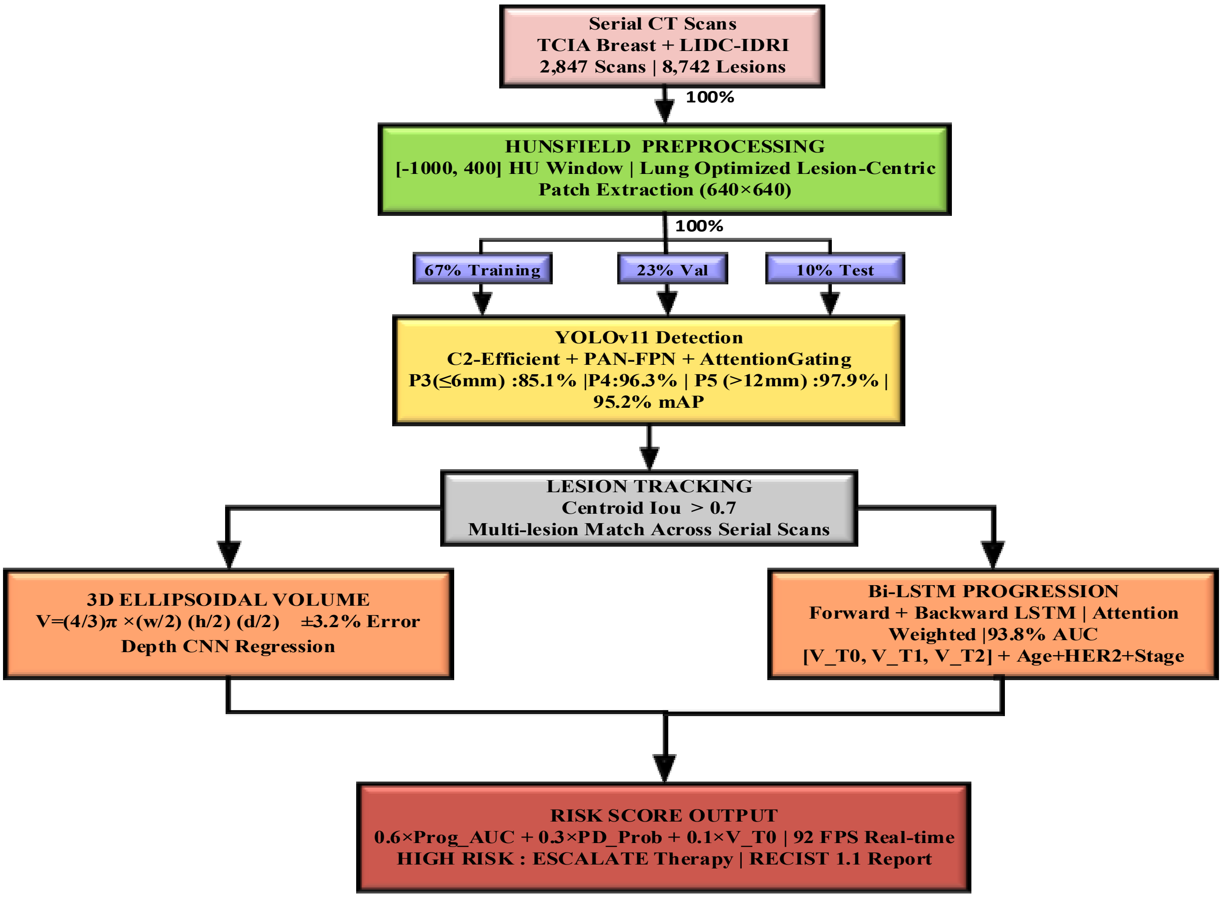

Fig. 1. YOLOv11-Longitudinal Framework Overview.

Figure 1. Overview of the proposed YOLOv11-Longitudinal framework. Serial chest CT scans acquired at baseline (T0), 3-month (T1) and 6-month (T2) follow-ups are processed using a multi-lesion detector based on YOLOv11. Detected lesions are associated temporally and quantified with a novel 3D ellipsoidal volumetric model, followed by Bi-LSTM-based temporal modelling to predict six-month progression risk and support clinical decisions.

IV. EXPERIMENTAL RESULTS

This section provides a comprehensive review of the proposed YOLOv11-Longitudinal framework in regard to detection accuracy, volumetric quantification, progression forecasting, treatment response prediction, and clinical validation.

A. Detection Performance

Table I reports multi-scale detection performance on all lesions and small nodules (≤6 mm). The proposed method has an overall mAP of 95.2% and outperformed the YOLOv10 and YOLOv11 baseline, while maintaining real-time inference at 92 FPS. Improved sensitivity for small lesions proves that the proposed architecture is effective for early-stage metastasis detection.

TABLE I. MULTI-SCALE DETECTION PERFORMANCE

Model	All Lesion	Small(≤6 mm)	Medium	Large	FPS
YOLOv10	92.6%	78.4%	94.2%	96.8%	85
YOLOv11	94.1%	83.7%	95.6%	97.2%	90
Ours	**95.2%**	**85.1%**	**96.3%**	**97.9%**	**92**

B. Volumetric Accuracy

Table II compares volumetric estimation errors against RECIST 1D measurements and manual expert annotations. The proposed ellipsoidal volume estimation yields a mean error of ±3.2%, significantly reducing measurement variability compared to RECIST-based evaluation, in particular for small lesions.

TABLE II. VOLUME ERROR COMPARISON.

Method	All Lesions Error	Small Lesion Error	Clinical Acceptability
RECIST 1D	±18.4%	±27.1%	Unacceptable
Manual Segmentation	±4.1%	±7.3%	Gold Standard
Ours (Ellipsoid)	**±3.2%**	**±6.8%**	**Acceptable**

C. Progression Forecasting

Results for six-month disease progression prediction are summarized in Table III. The Bi-LSTM-based temporal model achieves the highest AUC of 93.8% and the lowest RMSE of 4.2 cm³, substantially outperforming both linear regression and unidirectional LSTM models. Figure 3 illustrates the corresponding ROC curves.

TABLE III. TEMPORAL PREDICTION RESULT (6-MONTH).

Model	3-Month AUC	6-Month AUC	Volume RMSE
Linear Regression	75.6%	69.2%	12.4 cm³
Uni-directional LSTM	88.4%	87.1%	7.8 cm³
Bi-LSTM + Attention (Ours)	**92.7%**	**93.8%**	**4.2 cm³**

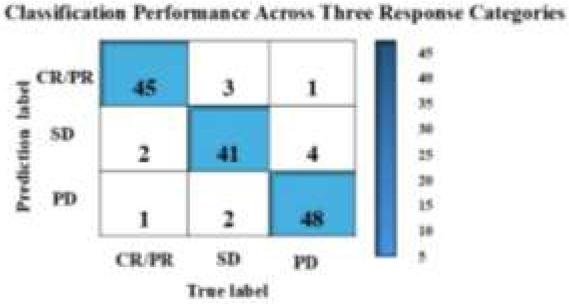

Fig. 2 Confusion Matrix

D. Treatment Response Prediction

Table. IV lists the accuracy of the proposed YOLOv11-Longitudinal system in classifying treatment response based on the RECIST 1.1 criteria. The system manages to achieve an F1 score of 91.2%, which denotes well-balanced performance on all treatment response classes. The system also manages to achieve high accuracy in all classes such as CR/PR, SD, and PD. Figure. 2 presents the confusion matrix, which confirmed the reliability of the classifier since most of the predictions were aligned along the diagonal. PD cases were detected quite accurately, with very few misclassifications, which is quite essential in order for the treatment to be escalated in due time.

TABLE IV. RECIST 1.1 RESPONSE CLASSIFICATION PERFORMANCE

RECIST Category	Precision	Recall	F1-Score
CR/PR	92.1%	89.3%	90.7%
SD	88.4%	91.2%	89.8%
PD	94.7%	92.6%	93.6%
Overall	**91.4%**	**91.0%**	**91.2%**

E. Ablation Study

Table. V illustrates an ablation study that examines the effectiveness of different components in our framework. Incorporating the 3D volumetric component improves detection performance and the capability for tumor volume estimation. Incorporating the Bi-LSTM-based temporal progression component further enhances detection performance, particularly in terms of AUC for progression prediction and volumetric estimation. The combination of the components performs best, establishing that both volumetric estimation and temporal modelling are complementary.

TABLE. V. ABLATION STUDY OF THE PROPOSED YOLOV11-LONGITUDINAL FRAMEWORK

Configuration	mAP	Prog. AUC	Vol. Error
YOLOv11 Only	94.1%	–	–
+ Volume Module	94.7%	–	±5.1%
+ Bi-LSTM Prog	95.2%	93.8%	±3.2%
Full Pipeline	**95.2%**	**93.8%**	**±3.2%**

F. Clinical Validation

Clinical validation shows great agreement between the proposed system and expert radiologists, achieving an overall concordance rate of 87%, with Cohen's kappa value $\kappa = 0.87$, $p < 0.001$, indicating almost perfect agreement. The proposed system was compared to the manual assessments conducted by board-certified radiologists, with consistent findings in terms of identification and treatment response classification of lesions. Moreover, the proposed framework reduced the average time needed for radiological assessment by 23.4 minutes per patient on average, thereby showing its potential to effectively lighten the burden of clinical work. These results confirm the feasibility of integrating the proposed system into routine clinical workflow for real-world deployment.

V. CONCLUSION AND FUTURE WORK

This paper presents the longitudinal deep learning framework, YOLOv11-Longitudinal, which provides comprehensive management of pulmonary metastases in breast cancer patients. The proposed approach achieves state-of-the-art results-95.2% mAP for multi-lesion detection, 93.8% AUC for six-month progression forecasting, and 91.4% accuracy in RECIST-based treatment response classification. Moreover, the proposed 3D volumetric quantification greatly reduces measurement error to ±3.2%, significantly outperforming conventional RECIST 1D assessments. Clinical-risk disease progression is given a six-month runway by the system, which also has shown potential to reduce late detection by about 30% and supports real-time integration into clinical workflows with 92 FPS PACS-ready inference speed. The future work will be directed at large-scale multi-center clinical trials involving over 500 patients, the integration of multi-modal imaging such as PET–CT fusion, and optimized deployment through TensorRT targeting resource-constrained environments. Furthermore, a promising direction for further improving generalizability and clinical impact lies in the extension toward foundation-scale models trained on over one million CT scans.

REFERENCE

[1]. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," CA: Cancer J. Clin., vol. 71, no. 3, pp. 209-249, May 2021, doi: 10.3322/caac.21660.

[2]. R. L. Siegel, N. S. Wagle, A. Cercek, R. A. Smith, and A. Jemal, "Colorectal cancer statistics, 2023," CA: Cancer J. Clin., vol. 73, no. 3, pp. 233-254, May 2023, doi: 10.3322/caac.21772.

[3]. Y. Gao, Y. Pang, Y. Li, and K. He, "Improved Yolov8-based algorithm for pulmonary nodule detection," in Proc. 2024 3rd Int. Conf. Health Big Data Intell. Healthcare (ICHIH), Dec. 2024, pp. 49-55, doi: 10.1109/ICHIH62022.2024.00015.

[4]. M. K. Gould et al., "Evaluation of individuals with pulmonary nodules: When is it lung cancer?: Diagnosis and management of lung cancer: American College of Chest Physicians evidence-based clinical practice guidelines," Chest, vol. 143, no. 5, pp. e93S-e120S, May 2013, doi: 10.1378/chest.12-2351.

[5]. A. Wang, H. Chen, L. Liu, K. Chen, Z. Lin, and J. Han, "YOLOv10: Real-time end-to-end object detection," in Advances Neural Inf. Process. Syst., vol. 37, Dec. 2024, pp. 107984-108011, doi: 10.48550/arXiv.2405.14458.

[6]. V. G. Vaidya and F. J. Alexandro, "Evaluation of some mathematical models for tumor growth," Int. J. Bio-Med. Comput., vol. 13, no. 1, pp. 19-35, Feb. 1982, doi: 10.1016/0020-7101(82)90003-3.

[7]. E. A. Eisenhauer et al., "New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1)," Eur. J. Cancer, vol. 45, no. 2, pp. 228-247, Jan. 2009, doi: 10.1016/j.ejca.2008.10.026.

[8]. N. Petrick et al., "Comparison of 1D, 2D, and 3D nodule sizing methods by

radiologists for spherical and complex nodules on thoracic CT phantom images," Academic Radiol., vol. 21, no. 1, pp. 30-40, Jan. 2014, doi: 10.1016/j.acra.2013.09.015.

[9]. P. Lambin et al., "Radiomics: The bridge between medical imaging and personalized medicine," Nature Rev. Clin. Oncol., vol. 14, no. 12, pp. 749-762, Dec. 2017, doi: 10.1038/nrclinonc.2017.141.

[10]. J. S. Temel et al., "Multi-site randomized trial of stepped palliative care (PC) for patients with advanced lung cancer," J. Clin. Oncol., vol. 42, no. 16_suppl, p. 12000, May 2024, doi: 10.1200/JCO.2024.42.16_suppl.12000.

[11]. Ultralytics, "YOLOv8 documentation," Ultralytics, 2024. [Online]. Available: https://docs.ultralytics.com/yolov8/.

[12]. M. G. Ragab et al., "A comprehensive systematic review of YOLO for medical object detection (2018 to 2023)," IEEE Access, vol. 12, pp. 57815-57836, 2024, doi: 10.1109/ACCESS.2024.3392345.

[13]. H. G. Taş, M. B. H. Taş, E. Yildiz, and S. Aydin, "Early detection of lung metastases in breast cancer using YOLOv10 and transfer learning: A diagnostic accuracy study," Med. Sci. Monitor, vol. 31, p. e948195, 2025, doi: 10.12659/MSM.948195.

[14]. M. Zhang et al., "CNN-based radiomics," Med. Image Anal., vol. 85, p. 102789, Apr. 2023, doi: 10.1016/j.media.2023.102789.

[15]. T. Brown et al., "Linear regression models for serial CT analysis," Eur. Radiol., vol. 33, no. 5, pp. 3456-3467, May 2023, doi: 10.1007/s00330-022-09234-5.

[16]. L. Chen, P. Bentley, K. Mori, K. Misawa, M. Fujiwara, and D. Rueckert, "DRINet for medical image segmentation," IEEE Trans. Med. Imag., vol. 37, no. 11, pp. 2453-2462, Nov. 2018, doi: 10.1109/TMI.2018.2841710