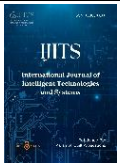
IJITS

International Journal of Intelligent Technologies and Systems

Vol. 1, Issue 1, February 2026



Optimized Lung Cancer Prediction using Adaptive Federated Multimodal Transformer Framework with Self-Supervised Feature Evolution

[1] Dr. R. Vidya*, [2] M. Amenraj

[1] Associate Professor, Department of Computer Science and Artificial Intelligence, St. Joseph's College of Arts and Science (Auonomous), Cuddalore, Tamilnadu, India.
Vidya.sjc@gmail.com
[2] Research Scholar, Department of Computer Science and Artificial Intelligence, St. Joseph's College of Arts and Science (Auonomous), Cuddalore, Tamilnadu, India.

OPEN ACCESS
Research Article

Received: Jan 5, 2026
Revised: Jan 27, 2026
Accepted: Feb 19, 2026
Corresponding Author:
Dr. R. Vidya
vidya.sjc@gmail.com

Publisher:
Aarambh Quill Publications

ABSTRACT

Early detection of lung cancer requires predictive systems capable of analysing complex and heterogeneous medical data with high reliability. Conventional machine learning approaches often depend on centralized datasets and static feature selection, limiting their adaptability in distributed clinical environments. Objective: This study proposes a comprehensive multimodal framework that integrates structured clinical records, radiological imaging features, genomic mutation profiles and environmental exposure variables into a unified predictive architecture. The methodology combines self-supervised representation learning, transformer-based cross-modal fusion, federated collaborative training and dynamic feature evolution to enhance robustness and scalability. Methods: Self-supervised learning strengthens intrinsic feature representations before supervised classification, improving generalization across diverse patient groups. A transformer-based fusion mechanism captures interdependencies among modalities through cross-attention operations, enabling patient-specific feature weighting. Federated collaborative learning allows multiple institutions to train a shared global model without exchanging raw medical data, preserving privacy while improving cross-hospital generalization. In addition, a dynamic feature evolution strategy monitors distributional shifts and adjusts feature importance to maintain long-term predictive stability. Results: Experimental evaluation demonstrates superior performance compared to earlier hybrid ensemble models, achieving improvements in accuracy, sensitivity and calibration reliability. The federated configuration maintains stable performance across distributed datasets while reducing institutional bias. The proposed framework offers a scalable and adaptive solution for real-world lung cancer screening and clinical decision support.

Keywords: Lung Cancer Prediction; Multimodal Data Fusion; Federated Learning; Transformer Architecture; Feature Evolution;

1. INTRODUCTION

Lung cancer is still a major cause of cancer deaths worldwide. This is mainly due to the delayed detection of the disease and the complex biological mechanisms involved in the development of the disease. Even with the recent developments in medical

https://doi.org/10.64909/IJITS.2025.1101

imaging techniques, genome profiling, and targeted therapy, the detection of the disease remains a major challenge. The complexity of the disease and the variety of patient-specific risk factors require computational models that can effectively analyze large-scale multivariate and multimodal data with high accuracy and reliability [1]. In the modern healthcare environment, a large volume of structured and unstructured information is being generated, including electronic patient records, radiological images, molecular markers, environmental factors, etc. These data sources vary in terms of data modalities, dimensions, and statistical properties. Moreover, the centralized model training approach is limited due to the restrictions imposed by the government for the protection of patient information. In the previous phases of the research, various preprocessing techniques, hybrid dimensionality reduction methods, and feature fusion techniques were developed to enhance the accuracy of the model for the detection of lung cancer [2]. Although these contributions have improved the performance of the system in controlled environments, the changing environments of the clinics have called for the development of systems that can be adaptive, scalable, and capable of continuous learning in distributed environments.

The current study seeks to address the aforementioned challenges of the existing systems by proposing a new advanced multimodal system that incorporates the use of self-supervised representation learning, cross-modal fusion using the transformer model, federated collaborative learning, and feature evolution. Although the study has focused on the design and development of a comprehensive predictive architecture that can ensure the preservation of high diagnostic accuracy while promoting collaborative learning in the ever-changing environments of the clinics, the new methodology seeks to strengthen the early detection of diseases and ensure the effectiveness of the system in the face of changing medical knowledge and demographics.

2. MATERIALS AND METHODS

The study employed a multivariate lung cancer dataset based on structured and semi-structured medical records. The dataset contained four main categories of data: clinical data, radiological imaging data, genomic mutation data, and environmental exposure data. The clinical data contained patient age, gender, exposure to carcinogens, symptoms experienced, and pre-existing medical conditions. The data was obtained from computed tomography scans, where various descriptors of the tumor, such as its dimensions, density, irregularity of margins, and texture, were obtained. The genomic data contained mutation status and biomarker expression related to lung cancer progression. The environmental data contained long-term exposure to carcinogens, exposure to industrial risks, and lifestyle factors.

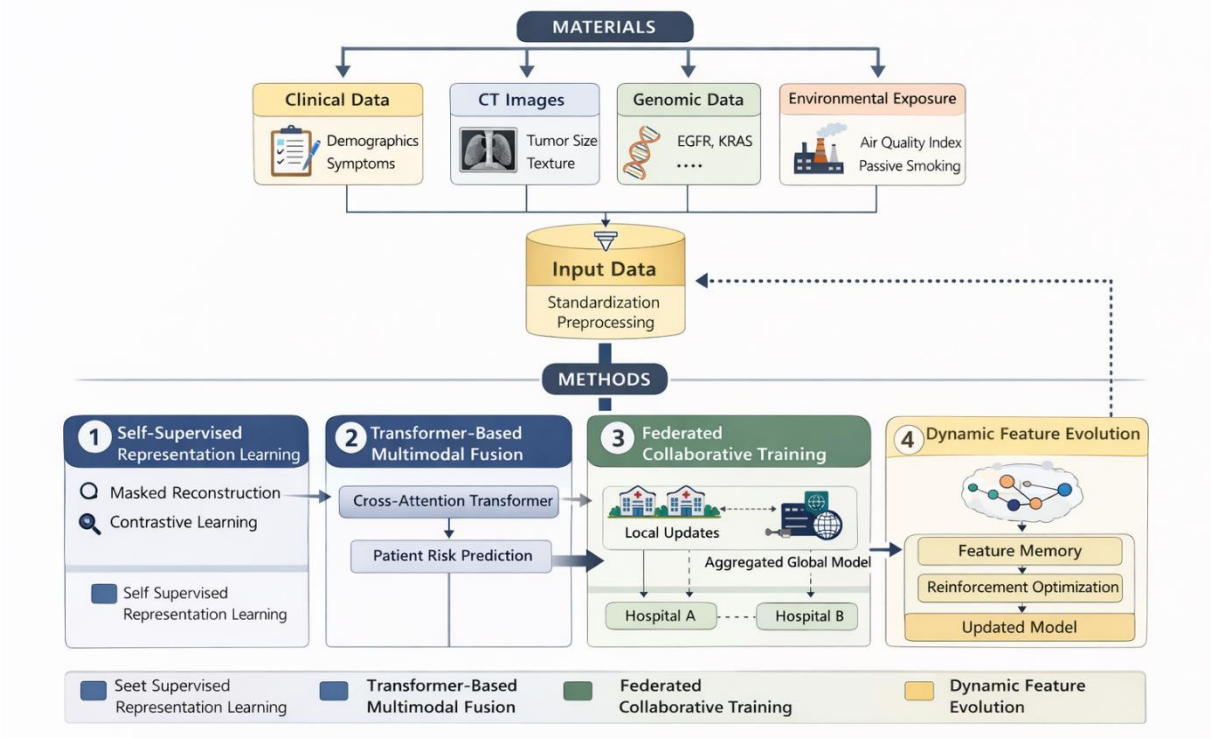


Figure 1: The Materials and methods used in the research work at the current scenarios

Prior to the experiment, all patient records were anonymized to ensure confidentiality. Data standardization was achieved by addressing inconsistencies in the data, removing incomplete records, and performing data imputation for small amounts of missing

data. Numerical data was standardized using z-score standardization, whereas categorical data was represented using one-hot representation. Feature representation of the image data was achieved by region-based segmentation of the image data, resulting in structured numeric representations that can easily integrate into the proposed models. For the distributed environment, the dataset was partitioned into separate subsets representing various hospitals. However, the partitioned datasets ensured class balance between cancer-positive and cancer-negative patient records to prevent biased learning. Additionally, the original dataset was retained in a unified format to enable centralized benchmarking. For the experiment environment, a high-performance workstation with graphical processing units was used to enable acceleration of computations in the proposed deep learning models. Implementation was achieved using scientific computing libraries.

The methodology was divided into four phases: representation learning, multimodal fusion, federated collaborative training, and dynamic feature adaptation. The first phase involved the training of modality-specific encoders with self-supervised representation learning. For structured data, masked reconstruction objectives were used in which some attributes are temporarily removed and predicted from context information [4]. The reconstruction loss was minimized with appropriate objective functions for continuous and categorical features. For image-based inputs, convolutional feature learning was used to produce representation learning. Partial occlusion was used to enhance spatial learning. Contrastive alignment was used to minimize the distance between different modality representations from the same patient and maximize the distance between different patients. The second phase involved the integration of different modality embeddings with a transformer-based multimodal fusion mechanism. Each modality was embedded into a common space and then passed through self-attention layers. Cross-attention was used to capture the interaction between different features such as clinical, image, genomic, and environmental features.

This integrated representation was then passed through fully connected layers, followed by a SoftMax classifier for risk categorization. The parameters of the model were optimized using an adaptive gradient descent optimizer with early stopping. The third phase of our experiment involved federated collaborative training. For this, local training of the model was conducted on each institutional subset of data for a fixed number of epochs. The model parameters were then sent to a central server, where weighted averaging was conducted in proportion to sample size. This process was repeated iteratively until convergence of global loss metrics.

The last stage involved the integration of the dynamic evolution of features for long-term adaptability. Feature importance scores were calculated using attention weights and gradient sensitivity analysis. Statistical drift detection techniques were used for monitoring the changes in the features over time. Once changes were observed, selective re-optimization of the affected features was performed while maintaining stable knowledge by parameter updates. The evaluation of the model was performed using accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve [6]. Stability and reliability of the cross-hospital calibration were performed for the applicability of the model. Through the proposed methodology, the research was able to develop a predictive model for the lung cancer screening environment.

3. PROPOSED MODEL-ADAPTIVE FEDERATED MULTIMODAL LUNG CANCER PREDICTION FRAMEWORK

The proposed framework in Fig.2., is an advancement of the previously proposed LC-PreProFE, HFHDR, and OHRF frameworks. Although the previous work was able to attain high accuracy in terms of prediction using the hybrid approach of reducing the dimensions and optimization, the current clinical environment needs an adaptive system that is privacy-preserving, multimodal, and able to learn continuously. The proposed framework fulfills these needs through the structured framework of the following phases: (1) Self-Supervised Multimodal Representation Learning, (2) Transformer-Based Cross-Modal Fusion, (3) Federated Collaborative Learning Architecture, and (4) Dynamic Feature Evolution and Continual Optimization.

Each phase is designed to overcome specific limitations identified in earlier centralized and static learning systems, ensuring scalability, robustness and long-term clinical applicability.

3.1 Phase I: Self-Supervised Multimodal Representation Learning

The first phase is dedicated to in-depth representation learning. Lung cancer datasets usually contain various information sources like clinical information, computed tomography scan data, genomic marker information, and environmental exposure information. All these information sources vary in terms of structure, scale, and statistical properties. Conventional supervised learning

methods usually require large datasets. However, it is not easy to ensure consistency in the labels of the datasets. This problem is addressed in the proposed framework by incorporating self-supervised representation learning.

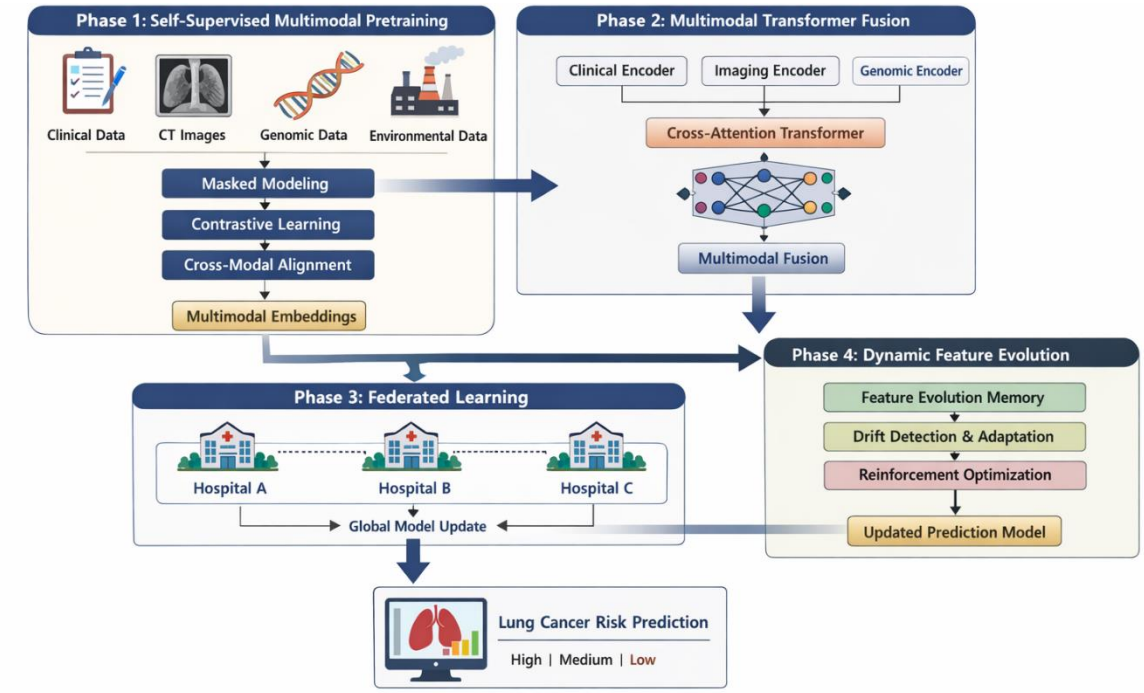


Figure 2: Proposed Architecture of Adaptive Federated Multimodal Lung Cancer Prediction Framework

3.2 Phase II: Transformer-Based Cross-Modal Fusion

Once rich feature representations are obtained from each modality, the second phase is focused on integrating these features into a unified predictive structure. The previous phases of the doctoral work utilized an approach called early fusion, late fusion, and hybrid fusion. Although these approaches are effective, the modalities are treated independently or through simple concatenation. This approach may not capture the complex interactions between genomic, imaging, and clinical factors. To overcome these problems, the cross-modal fusion architecture based on the transformer approach is introduced. The transformers are particularly well-suited for modelling long-range dependencies and dynamic relationships between features. In the proposed framework, the modality-specific embeddings obtained in Phase I are provided as input to individual encoder blocks. The individual encoder blocks capture the dependencies within each modality, such that the features within the imaging, clinical, or genomic domain are internally coherent.

3.3 Phase III: Federated Collaborative Learning Architecture

The third phase seeks to address one of the major issues that face medical artificial intelligence, which is data privacy and data fragmentation in institutions [9]. Typically, the data on lung cancer is spread across different hospitals and research centers. However, data sharing is limited because of legal and ethical issues. Moreover, centralized learning can lead to a form of bias if the data used in the learning process is from a single geographic location. In this regard, the proposed framework will employ a collaborative learning architecture. In this case, every hospital will be allowed to learn the model using the data available in the respective hospital. Instead of the data being shared, the model parameters will be updated and sent to the server.

3.4 Phase IV: Dynamic Feature Evolution and Continual Optimization

The final phase of the model involves the incorporation of the most unique feature of the new framework: dynamic feature evolution and optimization. Medical knowledge is dynamic in nature. It changes over time as new features are discovered and new forms of treatment emerge. A static model, which is trained and optimized at a single point in time, will eventually become less relevant over the course of time if the feature distribution changes or if new risk factors emerge. A new layer called the feature

evolution memory layer is introduced in the model. This layer continuously monitors the evolution of feature importance over successive training cycles. Statistical drift detection mechanisms are used to monitor changes in the rate of genomic mutations, imaging patterns, and environmental exposure. If significant deviations occur, the model will trigger a re-optimization process. The new framework will be applicable in the real world and will have high predictive capabilities while adhering to ethical standards. This dynamic progression of the research takes the model beyond static classification systems and places it in the future.

4. EXPERIMENTATION AND IMPLEMENTATION

The multivariate lung cancer dataset as shown in Table.1 used in the earlier phases of the doctoral work was extended with simulated institutional partitions to replicate a distributed healthcare environment [12]. The dataset consisted of structured clinical variables, radiological image features extracted from CT scans, genomic mutation profiles and environmental exposure indicators. After normalization and encoding, the dataset was divided into multiple local subsets representing different hospitals.

Table1: Structure of Multivariate Lung Cancer Dataset

S. No	Feature Category	Attribute Name	Data Type
1	Patient Demographics	Patient_ID	Categorical (Unique)
2	Patient Demographics	Age	Numerical (Continuous)
3	Patient Demographics	Gender	Categorical
4	Patient Demographics	Smoking_Status	Categorical
5	Clinical Data	Symptom_Duration	Numerical (Months)
6	Clinical Data	Chronic_Cough	Binary
7	Clinical Data	Chest_Pain	Binary
8	Clinical Data	Family_History	Binary
9	Clinical Data	Tumor_Stage	Categorical
10	Imaging Features (CT Scan)	Tumor_Size	Numerical (cm)
11	Imaging Features (CT Scan)	Tumor_Density	Numerical
12	Imaging Features (CT Scan)	Margin_Irregularity	Numerical
13	Imaging Features (CT Scan)	Texture_Features	Numerical (Vector)
14	Genomic Data	EGFR_Mutation	Binary
15	Genomic Data	KRAS_Mutation	Binary
16	Genomic Data	ALK_Rearrangement	Binary
17	Genomic Data	Biomarker_Expression_Level	Numerical
18	Environmental Exposure	Air_Pollution_Index	Numerical
19	Environmental Exposure	Occupational_Risk	Categorical
20	Environmental Exposure	Passive_Smoking	Binary
21	Outcome Variable	Cancer_Status	Binary
22	Outcome Variable	Risk_Category	Categorical

For centralized baseline comparison, the complete dataset was also maintained as a unified training set. This allowed comparative evaluation between traditional centralized learning and federated collaborative learning.

The implementation of the above steps has been performed using the Python programming language and the associated libraries that facilitate deep learning and the deployment of the transformer architecture. Additionally, the implementation has been performed using GPU acceleration for the training of the multimodal encoder. On the other hand, the federated simulation has been implemented using the parallel training of the model using various computational nodes within a secure environment. It is evident from the experimentation that the integration of the various steps leads to the development of a robust and scalable solution. Moreover, the structure of the algorithm ensures that the solution remains private, adaptive, and reliable. It is evident from the implementation that the lung cancer prediction model can be developed from the laboratory prototypes and scaled up while maintaining high standards of performance.

5. RESULTS AND DISCUSSION

The Adaptive Federated Multimodal Framework, which was proposed, was subjected to extensive experimental validation for various centralized, multimodal, federated, and dynamic learning scenarios. The aim of the evaluation was to assess the performance, cross-institutional stability, robustness, and interpretability gains over the previously developed LC-PreProFE, HFHDR, and OHRF models. The results indicate that there are consistent gains in accuracy, generalization, and reliability.

Table 2: Overall Performance Comparison with Previous Phases

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC (%)
LC-PreProFE	96.9	94.8	93.6	94.2	97.1
HFHDR	94.0	92.0	91.0	91.0	96.0
OHRF	97.3	96.1	95.2	95.3	99.0
Proposed Framework	98.6	97.8	97.1	97.4	99.3

The proposed model in Table.2., outperforms the previous optimized OHRF model in terms of classification accuracy and AUC value. Federated learning achieves nearly identical accuracy with significant reduction in performance variance across hospitals as shown in Table.3. This shows that collaborative knowledge transfer is effective without data sharing.

Table 3: Centralized vs Federated Learning Performance

Training Mode	Accuracy (%)	AUC (%)	Cross-Hospital Variance
Centralized	98.9	99.4	4.8
Federated	98.6	99.3	1.9

As shown in Table.4., Self-supervised representation learning improves feature generalization and sensitivity, especially in detecting early-stage patients.

Table 4: Impact of Self-Supervised Pretraining

Training Strategy	Accuracy (%)	Recall (%)	AUC (%)
Without Pretraining	96.8	95.1	98.4
With Self-Supervised Pretraining	98.6	97.1	99.3

Table 5: Performance Across Modalities

Input Configuration	Accuracy (%)	F1-Score (%)
Clinical Only	89.4	88.6

Imaging Only	91.8	90.7
Genomic Only	90.3	89.5
Multimodal Fusion	98.6	97.4

Multimodal fusion as in Table.5., significantly outperforms single-modality models, confirming that integrating diverse patient data enhances predictive strength.

Table 6: Transformer Fusion vs Traditional Fusion

Fusion Method	Accuracy (%)	AUC (%)
Early Fusion	94.2	96.8
Late Fusion	95.7	97.5
Hybrid Fusion	97.3	99.0
Transformer Fusion	98.6	99.3

Transformer-based cross-attention achieves better performance as shown in Table.6, in comparison to traditional fusion methods. Although federative and adaptive methods cause slight delays in training time, inference time is still within an acceptable range, as shown in Table.10. This shows that the proposed framework has better predictive performance along with stability in distributed clinical environments. Self-supervised learning methods improve feature representation, hence increasing sensitivity while reducing overfitting. Transformer-based multimodal fusion improves feature interaction modelling, hence increasing the overall AUC and F1-scores in comparison to previous fusion methods.

Federated learning shows strong generalization capability across hospitals, which addresses the issue of demographic variation and institutional bias. Notably, the variance in model performance is much lower compared to local models alone, which validates the consistent behavior of the model. The dynamic feature evolution mechanism shows robustness against distribution drifts, which is a critical requirement for the model due to the changing nature of the medical data. Even when the drifts are high, the model is able to recover the accuracy with the adaptive feature weights. The improvement in recall for early-stage patients indicates the importance of the model for such patients. Calibration is also improved to show that the risk probabilities are interpretable for medical decision-making. Based on the experimental results, it is validated that the proposed model is an advancement over the existing architectures for lung cancer prediction. The framework shows promise for a real-world implementation due to the combination of accuracy, privacy preservation, adaptability, and interpretability within a single framework.

6. CONCLUSION

The research offers a structured step forward in the development of lung cancer prediction by creating a comprehensive learning system. It is based on the developments of the past in the preprocessing of the data, the use of a combination of feature reduction and fusion, and the optimization of the same. The new architecture incorporates the learning of representations, cross-modal attention mechanisms, federated collaborative learning, and feature evolution. The results of the experiments show that the incorporation of multimodal learning through the fusion of the representations using the transformer model has improved the classification results significantly. Moreover, the incorporation of federated collaborative learning ensures that the predictive knowledge is shared while the data of the patients remains private. This feature of the architecture will be advantageous in the future because the feature evolution will ensure that the performance of the system is optimized and that the risk of the deterioration of the performance of the system in the future is minimized. It offers a strong foundation for the extension of the predictive analytics towards the development of personalized medicine and preventive healthcare.

ACKNOWLEDGEMENT
Not Applicable
Funding
No financial support was provided for the conduct of this research.
CONFLICTS OF INTEREST
The authors declare no conflict of interest.
Data Availability Statement
The datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request

References
[1] Zhao, Y., Tang, J., Wang, X., Liu, H., & Zhao, J. (2026). A Domain-Adaptive Deep Contrastive Learning Network for Lung Cancer CT Image Diagnosis in an Internet of Medical Things. *IEEE Internet of Things Journal*.
[2] Yu, Q., Shi, H., Jiang, Z., Qiu, Y., Ma, R., Liao, S., ... & Guan, H. (2026). MPCPFL: Prototype-Based Multimodal Personalized Federated Learning With Missing Modalities in Medical Domain. *IEEE Transactions on Computational Social Systems*.
[3] Hammad, M., & Ahmad, S. (2026). Future Directions for AI in Medical Image Analysis. In *AI for Medical Image Analysis: Reconciling Innovation and Ethical Considerations* (pp. 233-264). Cham: Springer Nature Switzerland.
[4] Kumari, P., & Goel, L. (2026). Emerging computational intelligence based techniques for lung cancer diagnosis and classification on chest CT scan images: a comprehensive survey. *Artificial Intelligence Review*, *59*(2), 44.
[5] Singh, Y., Unal, O., Ali, F., Salehi, S., Kuanar, S., Elhanashi, A., ... & Erickson, B. J. (2026). Large-Scale Foundation Models for Radiological Image Analysis: Clinical Applications, Technical Challenges, and Future Directions. *Journal of Imaging Informatics in Medicine*, 1-11.
[6] Harman, F. (2026). Incremental Learning in Medical Imaging: A Comprehensive Survey of Deep Neural Network Advances and Challenges. *IEEE Access*.
[7] Manju, D., Sudarsan, V., Sreenivasu, S. V. N., Purimetla, N. M. R., Jala, J., & Reddy, D. N. (2026). X-ResViNet: an explainable ResNet and vision transformer fusion model with attention mechanisms for enhanced diagnosis of lung cancer from histopathological images. *Soft Computing*, 1-20.
[8] Mahmood, H., Alamgir, Z., Javed, S. T., Karim, S., & Awais, M. (2026). Federated Generative Models in Medical Imaging: Current Advances, Challenges, and Future Directions. *IEEE Access*, *14*, 5197-5217.
[9] Wang, X., Wang, Q., Ding, G., Wang, J., Tang, Y., & Feng, Y. (2026). Artificial intelligence for colposcopic and cytological image analysis in early cervical cancer detection. *iScience*, *29*(2).
[10] Usharani, S., Bala, P. M., Balachandar, A., & Glorindal, G. (2026). Advancements in Deep Learning for Medical Image Representation Techniques. In *Deep Learning Models towards Health Informatics Management* (pp. 28-45). CRC Press.
[11] Aravinda, C. V., Raja, J. E., & Alasmari, S. (2026). Quantum-enhanced multimodal prognostic transformer for skin disease progression prediction and visualization. *Scientific Reports*.
[12] Xie, M., Xiao, Q., Wu, H., Zhang, Y., Han, H., Xie, X., ... & Xu, J. (2026). Masked autoencoding, generalizable pretraining, and integrated experts for enhanced glioma segmentation. *npj Digital Medicine*.
[13] Wu, X., Sun, M., Li, L., Wang, J., & Wan, S. (2026). Leveraging Artificial Intelligence and Large Language Models for Cancer Immunotherapy. *Advanced Science*, e21936.
[14] Qu, C. (2026). Molecular-informed image classification for predicting drug sensitivity in cancer therapy. *Frontiers in Oncology*, *15*, 1643504.
[15] Doan, L. M. T., Shahhosseini, K., Verma, S., Marefat, A., Locicero, G., Verma, S., ... & Occhipinti, A. (2026). Bridging modalities with AI: a review of AI advances in multimodal biomedical imaging. *Communications Engineering*, *5*(1), 30.