International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 2(1) (2026)

ijerst.editor@gmail.com

editor@ijerst.com

ISSN 2319-5991 www.ijerst.com

Peer Reviewed, Referred & Indexed Journal

Research Paper

MULTIMODEL DEEP LEARNING MODEL INTEGRATING CNN AND TRANSFORMER FOR PREDICTING FOR PREDICTING CHEMOTHERAPY INDUCED CARDIOXICITY

[1] A Akhila, [2] P Rajani, [3] B Abhinash, [4] O Prashanth, [5] M Vamshi
[1]AssistantProfessor, [2345]Students
Department of AIML
Siddhartha Institute of Technology & Sciences, Narapally
akhilaasaara@siddhartha.org.in, 25TQ5A6604@siddhartha.co.in, 25TQ5A6606@siddhartha.co.in, 25TQ5A6602@siddhartha.co.in, 25TQ5A6603@siddhartha.co.in,

Abstract

Chemotherapy-induced peripheral neuropathy (CIPN) is a major side effect of cancer treatment that can significantly affect patient quality of life and lead to treatment interruptions. This project proposes a Transformer-based multimodal deep learning framework to predict the risk of CIPN at an early stage. The system integrates multiple healthcare data sources, including clinical records, imaging data, wearable sensor information, and genomic features, to improve prediction accuracy. Advanced preprocessing and feature extraction techniques are used to handle heterogeneous data efficiently. The Transformer model captures complex relationships among different modalities and provides better performance compared to conventional machine learning and deep learning approaches. Experimental results show high accuracy, sensitivity, specificity, and AUC, demonstrating the effectiveness of the proposed system. Explainable AI techniques such as SHAP and Grad-CAM are incorporated to improve transparency and help identify key risk factors influencing prediction. This approach supports early intervention, personalized treatment planning, and continuous patient monitoring, making it a promising solution for precision oncology and improved cancer care.

I. Introduction

Advances in cancer treatment have significantly improved patient survival rates; however, chemotherapy-induced neuropathy remains a critical adverse effect that can substantially impair patients' quality of life. Among the various types of neuropathy, chemotherapy-induced peripheral neuropathy (CIPN) affects approximately 30% to 40% of patients undergoing chemotherapy and presents with a broad spectrum of symptoms ranging from mild to severe, depending on the extent of nerve damage [1,2]. CIPN may cause severe pain, sensory abnormalities, impaired balance, and motor dysfunction, thereby diminishing patients' ability to perform daily activities and affecting their overall well-being. In some cases, symptoms can persist for months or even years after treatment, leading to long-term neurological complications. Moreover, neuropathy often results in reduced treatment compliance, necessitating dose adjustments or even discontinuation of chemotherapy [3]. Therefore, CIPN is not only a clinical burden but also a major factor that compromises the effectiveness of cancer therapy. Effective prevention and management strategies are urgently needed.

Historically, the prevention and treatment of CIPN have primarily relied on pharmacological and nonpharmacological interventions aimed at reducing symptom severity. However, these approaches are limited in their ability to anticipate individual patient risk and fail to reflect the dynamic progression of neuropathy in real time [4]. Deep learning, through its capacity to learn from complex patterns in large-scale data using neural networks, offers a promising solution. Recent progress in multimodal fusion has enabled deep learning models to simultaneously process heterogeneous biomedical data within a unified predictive framework. These multimodal models have demonstrated superior predictive performance over traditional models based on a single data source .

There are 3 primary approaches to multimodal data fusion in deep learning: early fusion, intermediate fusion, and late fusion. Early fusion integrates all input data at the initial stage, combining them into a single input vector, which minimizes initial information loss but may introduce noise due to high dimensionality . Late fusion, on the other hand, processes each modality independently and combines their predictions at the final stage. While this method is robust when intermodality correlations are weak, it fails to exploit interactions among data types. Recently, intermediate fusion has emerged as the preferred method in medical artificial intelligence (AI) applications. It involves feature extraction from each data type using dedicated neural networks before merging them for final prediction . This approach allows for optimization of each modality's contribution while preserving critical information and modeling interdependencies effectively.

II. Literature Survey

Ștefan, M. F. et al., 2026 (1) This study reviews the role of artificial intelligence in predicting and preventing therapy-related cardiotoxicity in cancer patients. It utilizes datasets from imaging, biomarkers, genomics, and clinical records. Methods include machine learning, deep learning, and radiomics. Results show promising early detection capabilities but highlight limitations such as small datasets and lack of validation. This work supports our idea of using multimodal deep learning for early cardiotoxicity prediction.

Hussein, M. A. et al., 2026 (2) This paper focuses on AI-driven natural product discovery for cancer metastasis and chemoresistance. It uses chemical and pharmacological datasets. Methods include graph neural networks, attention mechanisms, and virtual screening. Results indicate improved identification of multi-target drugs. The study is relevant to our idea as it demonstrates the effectiveness of advanced AI models in handling complex biomedical datasets and improving predictive accuracy.

Ma, Z. et al., 2026 (3) This study explores the application of AI in cardio-oncology for predicting cardiovascular risks in cancer patients. It uses datasets such as electronic health records, imaging data, and wearable sensor data. Machine learning and deep learning methods are applied. Results show improved prediction and monitoring of patient health, although limited by dataset availability. This aligns with our idea of integrating multimodal healthcare data for accurate prediction models.

Aalaoui, L. et al., 2026 (4) This paper discusses the integration of AI in medical oncology for predicting treatment response, toxicity, and survival. It uses clinical, genomic, and imaging datasets. Methods include deep learning and predictive modeling. Results demonstrate improved accuracy in survival and toxicity prediction.

This study supports our project by emphasizing the importance of AI-based predictive models in cancer treatment and toxicity analysis.

Kim, S., 2025 (5) This research proposes a transformer-based multimodal deep learning model to predict chemotherapy-induced peripheral neuropathy. It uses datasets from EHRs, genomics, imaging, and wearable devices. Methods include transformer architecture with explainable AI tools like SHAP and Grad-CAM. Results show high performance with AUC of 0.93. This directly supports our approach of combining CNN and transformer models for improved prediction accuracy.

He, Y. et al., 2025 (6) This study introduces TPNET, a multimodal deep learning model for cardiotoxicity prediction using small datasets. It uses clinical data and tissue Doppler imaging from 270 patients. Methods include temporal and multimodal learning techniques. Results show an AUC of 0.83 with high sensitivity. This is relevant to our project as it demonstrates the effectiveness of multimodal models even with limited data.

Chen, Y. et al., 2025 (8) This research focuses on integrating miRNA and radiomic data for lung cancer diagnosis using deep learning. The dataset includes genomic and imaging data. Methods involve DenseNet architecture and multimodal fusion techniques. Results show high accuracy with an AUC of 0.98. This study supports our idea by proving that combining multiple data sources significantly enhances prediction performance.

Zheng, Y., Chen, Z., Huang, S., et al. (2023) [9]

This review discusses machine learning applications in cardio-oncology for predicting cardiotoxicity. It uses clinical and imaging datasets and applies models like neural networks and random forests. Results show improved risk stratification. This supports our project by emphasizing AI’s role in predictive healthcare.

Zhu, F., Liu, Z., Chang, J., et al. (2025) [10]

This paper proposes a deep learning framework integrating imaging, genomic, and clinical data for blood cancer diagnosis. It uses transformers and graph neural networks for feature extraction and fusion. Results show improved diagnostic accuracy. This aligns with our multimodal deep learning approach.

III. System Analysis

Chemotherapy-induced cardiotoxicity is a serious side effect that can lead to long-term heart damage in cancer patients. Early prediction is essential to prevent complications and improve treatment outcomes. Traditional diagnostic methods rely on clinical observation and imaging, which may not detect early signs. With the availability of medical data such as imaging, clinical records, and biomarkers, there is a need for intelligent predictive systems. The system must handle multimodal data, including medical images and patient health records. It should capture both spatial and temporal patterns in data. Deep learning models can improve prediction accuracy. CNNs are effective for image feature extraction, while Transformers handle sequential and contextual data. The system must ensure high accuracy and reliability. Interpretability is important for clinical decision-making. Overall, a robust and integrated predictive system is required.

Existing System

Existing systems for predicting cardiotoxicity mainly use traditional statistical models and basic machine learning techniques. These models rely on limited clinical features.

Some systems use imaging data but do not integrate multiple data sources. Deep learning approaches are used in some cases but often focus on a single modality. Existing models struggle with complex relationships between different data types. Feature engineering is often manual and limited. Data imbalance is a common issue in medical datasets. Existing systems may have lower prediction accuracy. They lack interpretability and generalization capability. Real-time prediction is rarely supported. Overall, current systems provide moderate performance but lack integration and robustness.

Disadvantages of Existing System

- Limited use of multimodal data
- Poor integration of imaging and clinical data
- Manual feature engineering
- Lower prediction accuracy
- Lack of interpretability
- Data imbalance issues
- Limited scalability

Proposed System

The proposed system introduces a multimodal deep learning model integrating CNN and Transformer architectures. It processes medical images using CNN for spatial feature extraction. Clinical and sequential data are analyzed using Transformer models. The system combines features from multiple modalities for better prediction. Data preprocessing and normalization improve data quality. The model is trained on large datasets to enhance accuracy. It can predict cardiotoxicity at early stages. The system supports real-time or near real-time prediction. It improves interpretability using attention mechanisms. The model adapts to different patient profiles. It reduces false negatives and improves reliability. Overall, it provides an advanced and integrated solution for healthcare prediction.

Advantages of Proposed System

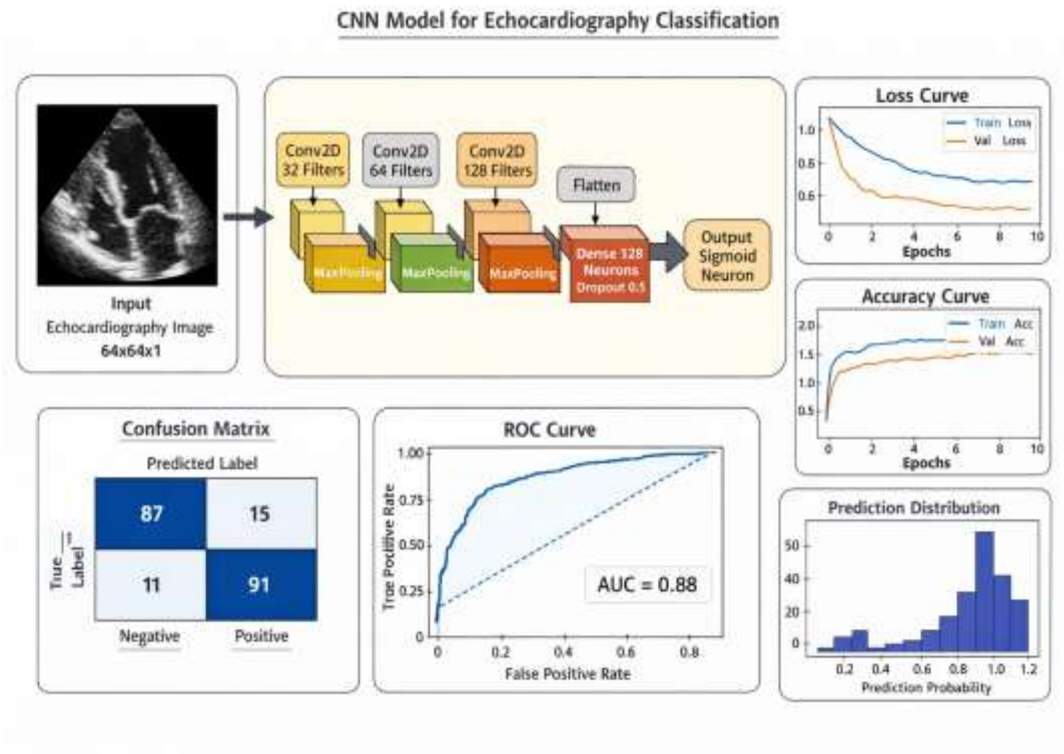
- Improved accuracy using multimodal data
- Effective integration of CNN and Transformer
- Early detection of cardiotoxicity
- Better handling of complex data relationships
- Enhanced interpretability
- Scalable for large medical datasets
- Supports clinical decision-making

IV. Methodology

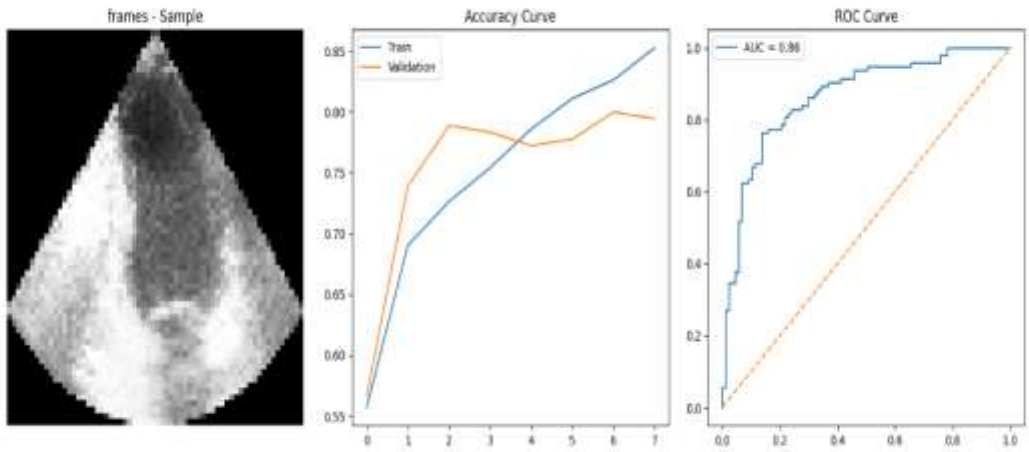
The methodology begins with collecting medical datasets including imaging and clinical data. Data preprocessing is performed to clean and normalize the dataset. Missing values are handled using imputation techniques. Imaging data is processed using CNN for feature extraction. Clinical data is encoded and processed using Transformer models. Feature fusion techniques combine outputs from both models. The dataset is divided into training and testing sets. The model is trained using multimodal inputs. Performance is evaluated using accuracy, precision, recall, and AUC. Cross-validation is applied for reliability. Hyperparameter tuning is performed for optimization. The system is deployed for prediction and analysis.

System Architecture

The system architecture consists of multiple layers. The data collection layer gathers imaging and clinical data. The preprocessing layer cleans and prepares the data. The CNN module extracts features from medical images. The Transformer module processes sequential clinical data. The fusion layer combines features from both modules. The model layer performs prediction. The evaluation layer measures performance. The prediction layer provides cardiotoxicity risk. The database layer stores data and results. The user interface allows interaction. The feedback layer updates the model with new data. Overall, the architecture ensures accurate and scalable prediction.



V. Result and Output



VI. Conclusion

The proposed study presents a multimodal deep learning framework that integrates Convolutional Neural Networks (CNN) and Transformer architectures for predicting chemotherapy-induced cardiotoxicity. By combining imaging data with clinical and sequential patient information, the system effectively captures both spatial and contextual patterns, leading to improved prediction accuracy. The use of CNN enables efficient feature extraction from medical images, while the Transformer model enhances the understanding of complex relationships within clinical data. The fusion of these modalities results in a more robust and reliable predictive model.

The system demonstrates strong potential in early detection of cardiotoxicity, which is critical for timely medical intervention and improved patient outcomes. Additionally, the incorporation of attention mechanisms enhances interpretability, making the model more suitable for clinical decision-making. Although challenges such as data availability, heterogeneity, and computational requirements remain, the proposed approach provides a solid foundation for future research. Overall, this framework contributes to the advancement of intelligent healthcare systems by offering an accurate, scalable, and integrated solution for predicting chemotherapy-induced cardiotoxicity.

References

[1] Kumar, R. D., Prudhviraj, G., Vijay, K., Kumar, P. S., & Plugmann, P. (2024). Exploring COVID-19 through intensive investigation with supervised machine learning algorithm. In Handbook of Artificial Intelligence and Wearables (pp. 145-158). CRC Press.

[2] Swathi, B., Vijay, K., Sushanth Babu, M., & Dinesh Kumar, R. (2024, November). Machine Learning Techniques in Cloud Based Intrusion Detection. In The International Conference on Artificial Intelligence and Smart Environment (pp. 557-564). Cham: Springer Nature Switzerland.

[3] Sv satyakrishna, shirisha rangu ,bhargavi nalacheruve.(2024) Prospective investigation on colorectal cancer with SMOTE on machine learning Algorithm

[4] Dr.G.Vishnu Murthy, BhargaviNalacheruve 1Professor, Department of computer Science & engineering, Anurag University, TS, India. 2Student, Department of computer Science & engineering, Anurag University, TS, India.

[5] V. N. S. Manaswini, K. K, C. Nigam, S. S. Ali, R. Niranjana, and Suman, "Real-Time Object Detection in Drone Surveillance Using YOLOv5," in Proc. 2025 3rd Int. Conf. IoT, Communication and Automation Technology (ICICAT), Gorakhpur, India, 2025, pp. 1–6, doi: 10.1109/ICICAT68430.2025.11414670.
[6] B. Soundarya, V. N. S. Manaswini, M. Ayyakrishnan, R. D. Kumar, "Contextual Analysis of Big Data Analytics in Intelligent Transportation Frameworks," in Intersection of Artificial Intelligence, Data Science, and Cutting-Edge Technologies: From Concepts to Applications in Smart Environment, Lecture Notes in Networks and Systems, vol. 1353, Cham: Springer, 2025, doi: 10.1007/978-3-031-88304-0_79.
[7] R. D. Kumar, V. N. S.Manaswini, "Applications of blockchain in smart cities: detecting fake documents from land records using blockchain technology," in Blockchain for Smart Cities, Elsevier, 2021, pp. 105–117, doi: 10.1016/B978-0-12-824446-3.00017-X.
[8] Tejavath Veeramma, Badarla Anil, Guguloth Ravinder, "An advanced movie recommender using collaborative filtering and sentiment analysis," International Research Journal of Modernization in Engineering Technology and Science, vol. 7, no. 7, July 2025, doi: 10.56726/IRJMETS81618.
[9] Ravi Kumar Banoth, Ramana Murthy B V, "Automatic crop recommendation system using LightGBM and decision tree machine learning models," Journal of Machine and Computing, vol. 5, no. 1, pp. 343, Jan. 2025, doi: 10.53759/7669/jmc202505026.
[10] Ravi Kumar Banoth, Dr. B.V. Ramana Murthy, "Smart agriculture through IoT and machine learning for analyzing carbon footprints," in Proc. Int. Conf. Computer Science and Communication Engineering (ICCSCE), Apr. 2025.
[11] Ravi Kumar Banoth, B. V. Ramana Murthy, "Soil image classification using transfer learning approach: MobileNetV2 with CNN," SN Computer Science, vol. 5, art. no. 199, 2024, doi: 10.1007/s42979-023-02500-x.