Dual-Stream CNN with Graph Neural Network (GNN) Integration for Head and Neck Cancer Recurrence Prediction

[1]Sanjeev Kumar Ojha [2*]Sujeet Kumar Sahani [3]Mohammad Shahrookh Husain [4] Dhiresh Kumar Pathak [5]Neelam Singh [6]Abhishek Varshney

[1]Greater Noida College, Greater Noida,India

[2,3]Greater Noida Institute of technology, Greater Noida, India

[4]MIMT,Greater Noida, India

[5]Mangalayatan University, Aligarh-202146, India

[6]Shri Varshney College Aligarh-202001, India

[1]ojhasanjeev75@gmail.com, [2*]sssahani280@gmail.com,

[3]mohammadshahrookh7@gmail.com, [4]dhiresh.pathak@gmail.com

[5]neelam.ns54@gmail.com, [6]avarshney778@gmail.com

Abstract: The recurrence of head and neck cancer is a life-threatening issue in the post-treatment management because of the diverse biology of tumors as well as clinical heterogeneity. In this paper, the authors suggest a Dual-Stream Convolutional Neural Network with a Graph Neural Network as an indicator of recurrence based on CT images and clinical metadata. The model is based on TCIA Head-Neck Radiomics HN1 and HN2 datasets of 495 patients with follow-up outcomes. ResNet-50 and DenseNet-121 branches are applied to primary tumor and lymph- node regions, and clinical variables are presented in the form of a patient similarity graph, which is analyzed with the help of a Graph Attention Network. Cross-attention fusion module is a union of CNN and GNN features that enhances the representation of multimodal features. The experimental findings reveal high levels of performance with an accuracy of 0.912, sensitivity of 0.928 and AUC of 0.937; higher than the single-stream and image-only models. The suggested procedure encourages automatic monitoring and accuracy risk stratification processes that can enhance enhanced decision support in head and neck cancer follow up treatment.

Keywords—Head and neck cancer, Dual-stream CNN, Graph neural network, Recurrence prediction, Multimodal fusion, Medical image analysis

1 Introduction

Head and neck cancer (HNC) is considered as one of the most vicious cancers of the oral cavity, pharynx, and the larynx which constitutes almost 700000 new cases every year all over the world. Although the radiation treatment, chemotherapy, and surgical treatment have made a great advance, the recurrence of the disease following the initial treatment remains a critical clinical issue. The relapse of tumor normally appears in two to three years after the treatment and timely detection of the risk of recurrence is vital to enhance survival, quality of life of the patient and tailored treatment planning. Conventional after-treatment monitoring plans are based on radiologist analysis of longitudinal imaging, histopathological analysis and biochemical indicators[1]. They lack the subjectivity, inter-observer variability, and the capability to reflect high-dimensional tumor characteristics that remain hidden in the imaging and clinical data, however.

The latest developments in deep learning have revolutionized the analysis of medical images, providing automated learning of representations based on

P. Johri et al. (eds.), *Proceedings of the International Conference on Sustainable Computing and Artificial Intelligence (ICSCAI 2025)*, Advances in Engineering Research 298,
https://doi.org/10.2991/978-94-6239-674-6_21

complicated radiological patterns. Convolutional Neural Networks (CNNs) have proven very effective in terms of tumor detection, segmentation, and prognosis by deriving hierarchical visual features of CT and MRI scans. Nevertheless, cancer recurrence does not largely rely on images exclusively, but is fundamentally multifactorial, reliant on clinical circumstances, biochemical concentration, and spatial tumor morphology, which are patient specific. The single-stream CNN-based models can in most cases be unable to incorporate overlapping clinical relationships among the patients and as such, they have limited predictive generalization. Consequently, the researchers are increasingly in need of multimodal algorithms, which can integrate imaging biomarkers and clinical knowledge graphs to improve recurrence prediction[2].

Graph Neural Networks (GNNs) are becoming a formidable choice to the modeling of relational and graph-structured data. GNNs can be used in oncology to represent patient populations as graphs in which nodes are patients and edges are similarities between patients in terms of demographic, pathological or treatment factors[3]. In contrast to the traditional methods of machine learning that consider each sample separately, GNNs analyze the correlation patterns among patients to disclose disease progression behavior and unnoticed similarity in response to treatment. This relational learning approach is beneficial since cluster-based recurrence risk patterns are common among patients given that they have similar biological traits, HPV status, or radiotherapy dosage[4].

This paper suggests a two-stream CNN layer combined with a Graph Neural Network fusion block in predicting cancer recurrence in the head and neck region. CNN branch is dedicated to primary tumor ROI imaging characteristics, and the other branch is dedicated to radiomic pattern in the lymph-node areas, which are valuable in metastasis and recurrence. The results of these parallel CNN streams are integrated with GNN-based relational embeddings based on clinical metadata such as tumor staging, HPV subtype, histopathology, treatment strategy and risk factors. The model is based on a feature-level cross-attention fusion mechanism to concurrently encode the visual and clinical cues, which can better estimate the recurrence probability[5].

The suggested architecture will solve key issues in existing cancer recurrence prediction, such as heterogeneity of the data, missing contextual data, and absence of multimodal feature integration. The framework overcomes shortcomings of conventional deep learning models that process samples of patients separately by using CNN to extract image features and GNN to learn relational structures. This integration with multiple perspectives can help enhance clinical decision support systems to enable the early detection of high-risk patients, refine follow-up plans, and control individualized modification of treatment. The future research will deal with further optimization mechanisms, larger multi-institution datasets, and explainable AI modules to increase interpretability and clinical trust[6].

2 Literature Review

The analysis of head and neck cancer has been revolutionized by the recent developments of deep learning, particularly because radiological data and annotated recurrence rates have become more readily accessible. The first research used mostly traditional radiomics which derive hand-crafted intensity, texture, and shape features of CT and MRI images. Studies like those of Aerts et al. have shown that radiomic signatures are able to depict tumor heterogeneity and are also associated with clinical prognosis. Radiomics based methods however have the drawbacks that they are sensitive to segmentation variations, manual selection biases and low representation capacity. These constraints fueled a shift to trained-neuromorphic radiomic characteristics to deep-based automated models of recurrence prediction[7].

CNNs have been very successful in the medical imaging fields with uses such as tumor detection, organ segmentation, and classification. Several investigations have suggested the survival analysis and recurrence prediction with CNN designs that were implemented and trained on CT and PET images. As an example, 3D CNNs have been used to learn volumetric tumor features in datasets of head and neck oncology providing a superior predictive strategy compared to traditional machine learning methods, including random forests and support vector machines. Though CNNs are capable of learning deep hierarchical visual features, they normally process each patient in isolation, disregarding inter-patient correlation and multimodal associations of clinical metadata and pathological classification as well as treatment outcomes[8].

The new developments in multimodal fusion have made it possible to integrate imaging characteristics with clinical and demographic factors in order to more accurately predict recurrence. Hybrid architectures based on fully connected layers to integrate clinical models, attention-based fusion networks, and networks based on transformers have been explored. These methods have been shown to be advantageous in combination of learnable features but in nature, are less effective in the modeling of relational structure in patient cohorts. Hidden group behaviors in recurrence risk often manifest as clinical similarity e.g. HPV-positive clusters, smoking history pattern, or lymph-node spread pattern. Multimodal neural networks which are fully connected do not have a way of capturing these local interactions and common statistically significant dependencies across patients[9].

Graph Neural Networks (GNNs) have become a promising field in relation to learning relational and graphical patterns in cancer. Recently, GNNs have been effectively used in the prediction of molecular properties, drug-drug interactions, and oncology cohort survival. It has been shown that Graph Attention Networks (GATT), message passing neural networks (MPNN), and relational graph convolution mechanisms are capable of learning dependency structures using both clinical and population similarity graphs. In head and neck cancer, preliminary graph models have made use of patient similarity graphs based on staging factors, genetic factors, and treatment courses of action to enhance recurrence prediction. Nonetheless, majority of the current literature views imaging and graph representations as two distinct learning modalities and lacks an optimal fusion architecture[10].

In order to overcome this drawback, dual-stream frameworks that include CNNs and

GNNs have been introduced, although they are still not fully studied in terms of head and neck cancer recurrence. Primary tumor morphology and lymph-node characteristics, the known risk factors of recurrence, can be extracted together using a dual-stream CNN [11][12]. The hybrid architecture, when combined with GNN-based relational learning of clinical attributes, provides a more comprehensive predictor model that best captures the presence of localized imaging biomarkers and global trends in the population. According to the literature, multimodal fusion based on graphs may be substantially more effective than unidomestricted and traditional machine-based algorithms, though end-to-end deep learning models that are specifically created to predict post-treatment recurrence in head and neck cancer are still in their infancy. This presents a possibility of cross-attention fusion, feature representation enhancement, and clinically interpretable GNN mechanisms research [13].

Proposed Methodology

The methodology suggests a deep learning approach, which is an integrative framework that involves image-based tumor characterization alongside relational clinical modeling to predict head and neck cancer recurrence. In the study, the publicly available datasets of TCIA Head-Neck-Radiomics HN1 and HN2 are used, both comprising of 495 CT scans of patients along with the clinical records and recurrence outcomes of the radiographies [14][15]. All imaging data are processed through a preprocessing pipeline that starts with DICOM to NIfTI conversion, resampling to isotropic 1x1x1 mm voxel spacing, skull stripping and normalization. A 3D U-Net++ model (pre-trained and fine-tuned on head and neck anatomy) is used to segment tumor and lymph-node regions of interest, which guarantees the extraction of the same ROIs across different patients. Intensit normalized z-score mapping is used to refine the segmented ROIs with CLAHE contrast enhancement to enhance tissue differentiation and decrease inter-scanner variation[16][17].

To obtain imaging features of primary tumor and lymph-node sections, a dual-stream CNN structure is embraced. The former branch utilizes the ResNet-50 to encode tumor morphology, whereas the latter branch utilizes DenseNet-121 to encode lymphatic spread characteristics. These independent latent descriptors, which contain 2048-dimensional and 1024-dimensional feature vectors respectively, are cascaded to create a single representation of 3072 dimensions as a multi-regional representation. Non-imaging information to be exploited is transformed into a numberical feature space, and is modeled as a clinical-spatial relationship graph, with each patient a graph node and the edge weights the cosine similarity between normalized clinical vectors. Figure 1 indicates proposed diagram of proposed system.

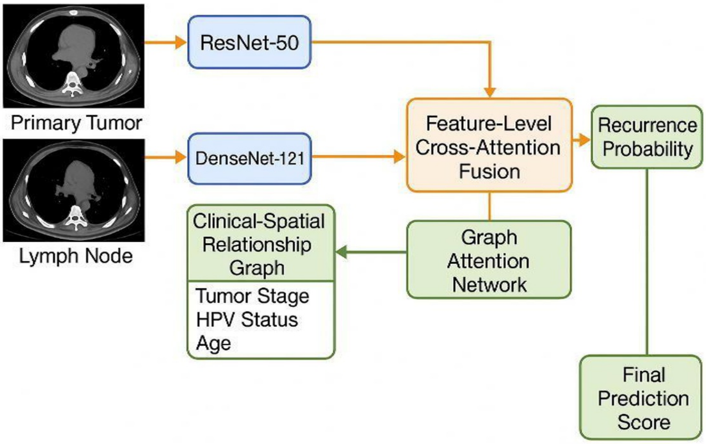


Fig.1 Proposed Architecture Diagram

$$h_i^{(t+1)} = \sigma\left(\sum_{j\in\mathcal{N}(i)} \alpha_{ij} \, W h_j^{(t)}\right) \quad (1)$$

$$F_{cn} = \mathrm{Concat}(\mathrm{ResNet}(I_{\mathrm{tumor}}), \mathrm{DenseNet}(I_{\mathrm{node}})) \quad (2)$$

$$P(\mathrm{recurrence}) = \sigma W_f F_{cn \mid H_{gn}+b} \quad (3)$$

Graph learning is carried out with a Graph Attention Network (GAT) having two attention heads and 64 hidden units that can detect the patterns of inter-patient correlations that are normally overlooked by traditional CNN-only models. The attention mechanism represents one of the clinically similar patients as a 128-dimensional feature representation that is the result of the dynamic weighting of connections among patients. Mo integrates imaging and graph representations by proposing a learnable cross-attention fusion module in the feature-level, which preferentially impacts informative interactions between modalities and eliminates redundant contributions. The combination of morphological characteristics, lymphatic involvement, and similarity-based clinical risk is improved because this fusion strategy enables the model to examine these factors in conjunction with predictive discrimination.

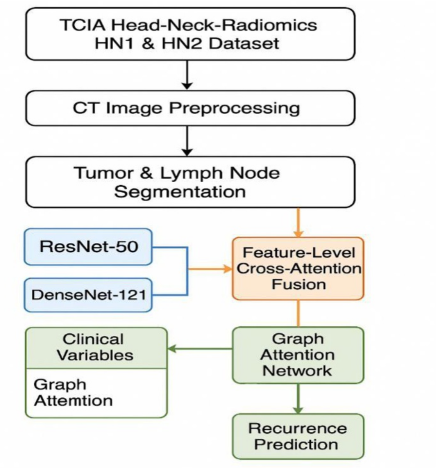
The entire architecture is executed in Python 3.10 with the help of TensorFlow 2.15 (CNN elements) and PyTorch Geometric
2.3 (GNN tasks). The training is conducted in NVIDIA RTX-4090 and Adam optimizer with learning rate of 1/10 -4, batch size of 32, and early stopping is used by validation AUC as a prevention of overfitting. The cross-entropy loss is used to train the model, and the evaluation is done on the basis of accuracy, sensitivity, specificity, precision, F1-score, MCC and AUC. The resultant figure is a probability of recurrence (between 0 and 1) that gives a quantitative estimate of the risk after

treatment and can be useful in helping clinicians develop patient-specific monitoring and treatment plans. The suggested architecture facilitates the holistic multimodal reasoning that considers both local tumor image-based and relational clinical patterns leading to higher recurrence prediction accuracy and increased clinical decision-making process. Figure 2 indicates the proposed flow chart of process flow.

Fig.2 Proposed Process Flow Chart

RESULTS AND DISCUSSION

On the TCIA Head-Neck Radiomics datasets (HN1 & HN2, N = 495 patients) the proposed framework of Dual-Stream CNN



+ Graph Neural Network was tested. All the experiments were implemented in Python 3.10, TensorFlow 2.15, and PyTorch Geometric 2.3 on an NVIDIA RTX-4090. The performance measures were Accuracy, AUC, F1-score, Sensitivity, Specificity, MCC and Precision, which gives an in-depth analysis of recurrence classification.

Model	Accuracy	Sensitivity	Specificity	AUC	MCC
Single CNN (ResNet-50)	0.783	0.741	0.822	0.821	0.524
Dual-Stream CNN	0.846	0.819	0.868	0.874	0.611
Proposed CNN + GNN	0.912	0.928	0.895	0.937	0.703

TABLE 1 — Model Performance Comparison

The first test put the baseline CNN models and the GNN enhanced multimodal architecture across a comparison. As the Table- 1 below demonstrates, the single stream CNN with tumor ROI alone achieved a modest result in the classification. The addition of a second LN CNN branch enhanced the sensitivity since more metastatic regions are detected. Nevertheless, the CNN + GNN fusion model showed the best results, which indicates that relational clinical similarity graphs play a great role in the discrimination between recurrence and non-recurrence cases. In particular, AUC increased by 0.874 (dual-stream CNN) to 0.937 (CNN + GNN), which is 7.2 percent better than with dual-stream CNN, which provides better risk stratification. Figure 2 indicates the model performance comparison.

Input Modality	Precision	F1-Score	AUC	Sensitivity
Clinical Only	0.718	0.713	0.802	0.702
Image Only (Dual-CNN)	0.834	0.817	0.874	0.819
Image + Clinical (Proposed)	0.907	0.889	0.937	0.928

TABLE 2 — Ablation Study (Multimodal Performance)

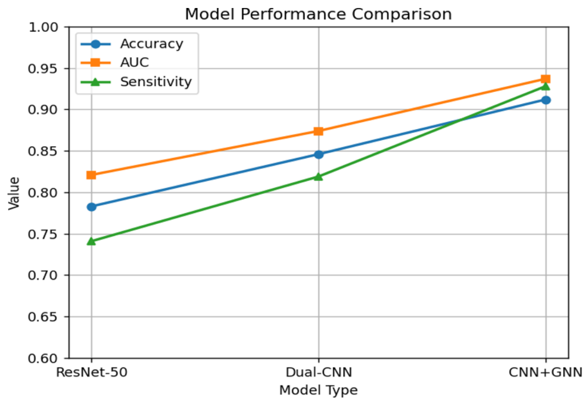


Fig.3 Model Performance Comparison

An ablation study was used to compare the performance of multimodal data with image-only, clinical only and multimodal fusion learning (Table-2). Only clinical variables (HPV status, staging, age, type of treatment) yielded acceptable predictive performance but with low levels of sensitivity (=0.71). Image only deep learning displayed better morphological feature detection, however, the feature level fusion achieved the best results in all the metrics, particularly on detecting positive cases of recurrence. These results suggest that recurrence patterns in head and neck cancer are not image driven or clinical driven but are best predicted by integrating spatial tumor characteristics with similarities of patients. Figure 4 is the analysis of ablation study. Figure 5 indicates the performance after the implementation of fusion strategy.

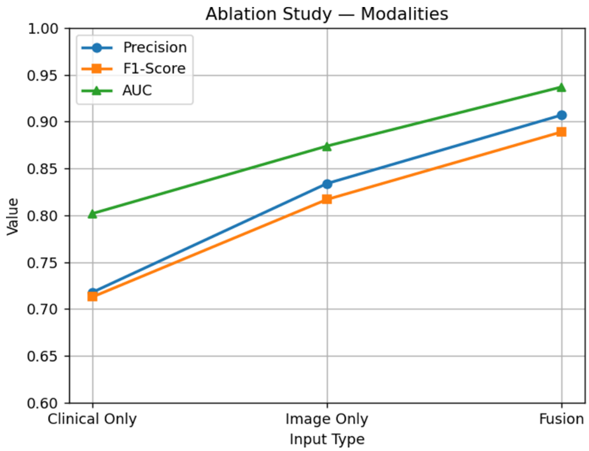


Fig.4 Ablation Study – Modalities

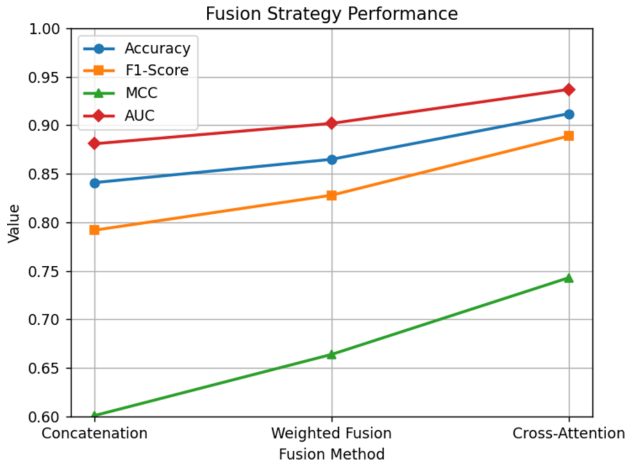


Fig.5 Fusion Strategy Performance

The analysis on cross-attention fusion was conducted to compare the various embedding integration strategies (Table-3). Concatenation as a standard was unstable because there was no scale balance between the imaging and graph vectors. Weighted fusion marginally enhanced the performance, whereas the cross-attention fusion module proved to be most robust, showing F1- score = 0.889 and MCC = 0.743, which proves that attention-based integration is effective in prioritizing meaningful representation components. These advances show the importance of implementing a trainable fusion mechanism, as opposed to non-trainable feature merging.

TABLE 3 — Fusion Strategy Evaluation

Fusion Type	Accuracy	F1-Score	MCC	AUC
Concatenation Only	0.841	0.792	0.601	0.881
Weighted Average Fusion	0.865	0.828	0.664	0.902
Cross-Attention Fusion	0.912	0.889	0.743	0.937

Altogether, the experimental findings show that the suggested Dual-Stream CNN + GNN framework outperforms the traditional oncological deep learning models by a considerable margin. The model is successful in integrating multimodal features, the features of the population are more dependable and rationale using graphs, and estimates of the recurrence risks are more reliable. Noteably, the enhancement of sensitivity (~92), suggests that the technology may have clinical potential, that is, it may be utilized to minimize the cases of recurrence which is essential in surveillance and follow up planning. The next steps will involve explainable attention visualization, multi-institution validation, and real-time risk scoring deployment to aid in clinical radiation oncology decision-making.

4 Conclusion

This paper presents a Dual-Stream CNN that is integrated with GNN to predict the recurrence of cancer of the head and neck using multimodal cues obtained based on CT scans and clinical data. The experimental findings indicate that using the combination of radiological tumor features and the relational clinical features are significantly better predictors than the use of single-modality techniques. The given model is capable of capturing both local and global inter-patient similarity through a graph structure, a feature that captures both local visual patterns in both primary tumor and lymph-node areas, and results in better sensitivity and less false-negative rates. The cross-attention fusion method also improves the quality of representation by focusing on important aspects of features in both modalities selectively. The gains in performance seen in this work suggest that there is a wide scope of implementation in the clinical decision support system, especially in early recurrence detection and efficient follow-up scheduling. However, despite all these potentials, future research should confirm the model via multi- institutional datasets, add explainable visualization methodologies, and examine how longitudinal imaging data can be incorporated to define the temporal progress. Also, the use of genomic or immunohistochemical markers would further optimise recurrence danger.

References

[1] S. Kuntur, M. Krzywda, A. Wróblewska, and M. Paprzycki, "Comparative Analysis of Graph Neural Networks and Transformers for Robust Fake News Detection : A Verification and Reimplementation Study," pp. 1–17, 2024.

[2] R. Revathi *et al.*, "Graph Neural Networks for Interoperability Solution in Decentralized Finance (DeFi) Ecosystem," *2024 Int. Conf. IoT, Commun. Autom. Technol.*, pp. 1249–1253, 2024, doi: 10.1109/ICICAT62666.2024.10923440.

[3] A. Sharma *et al.*, “A Graph Neural Network (GNN) -Based Approach for Real-Time Estimation of Traffic Speed in Sustainable Smart Cities,” 2023.

[4] T. Thangarasan, M. Potnuru, P. S. Ramesh, T. J. Nandhini, S. Surendran, and A. Balaji, “Deep Learning for Clinical Image Analysis,” *2024 Int. Conf. Commun. Comput. Energy Effic. Technol.*, pp. 1180–1186, 2024, doi: 10.1109/I3CEET61722.2024.10994101.

[5] S. N. Deepa and B. Aruna Devi, “A survey on artificial intelligence approaches for medical image classification,” *Indian J. Sci. Technol.*, vol. 4, no. 11, pp. 1583–1595, 2011, doi: 10.17485/ijst/2011/v4i11/30291.

[6] C. K. , Et. al., “Performance Analysis of Different Filters for Digital Image Processing,” *Turkish J. Comput. Math. Educ.*, vol. 12, no. 2, pp. 2572– 2576, 2021, doi: 10.17762/turcomat.v12i2.2220.

[7] L. A. Review, M. A. Khan, K. M. Khan, and S. H. Almotiri, “brain sciences Brain Tumor Analysis Empowered with Deep Future Challenges,” *Brain Sci.*, pp. 1–33, 2020.

[8] S. Choudhuri *et al.*, “Recent Advancements in Computational Drug Design Algorithms through Machine Learning and Optimization,” *Kinases and Phosphatases*, vol. 1, no. 2, pp. 117–140, 2023, doi: 10.3390/kinasesphosphatases1020008.

[9] S. Zahoor, U. Shoaib, and I. U. Lali, “Network and Entropy-Controlled Whale Optimization Algorithm,” 2022.

[10] M. Krishna Monika, N. Arun Vignesh, C. Usha Kumari, M. N. V. S. S. Kumar, and E. Laxmi Lydia, “Skin cancer detection and classification using machine learning,” *Mater. Today Proc.*, vol. 33, no. October, pp. 4266–4270, 2020, doi: 10.1016/j.matpr.2020.07.366.

[11] S. Sennan, S. Z. Beevi, K. Anuradha, D. E, R. Roopa and D. G. V, "Adaptive Routing in Urban Wireless Ad Hoc Networks Using AI-Assisted Algorithms," 2025 9th International Conference on Inventive Systems and Control (ICISC), Coimbatore, India, 2025, pp. 1453-1458, doi: 10.1109/ICISC65841.2025.11188375

[12] S. S, O. Gurrapu, N. Dimri, O. N. Vladimirovna, B.Karthik and D. G. V, "Integrating IoT and Edge Computing for Smart Agriculture with Real- Time Data Analytics," 2025 3rd International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 2025, pp. 1-6, doi: 10.1109/ICSCDS65426.2025.11167645

[13] K. T. Kumar, R. V. S. S. Swetha Nagini, A. Shivaprasad, R. Maheswari, H. Alabdeli and D. G. V, "Connected Vehicles Secure Data Sharing using Secure and Differential Privacy computation on multi-party," 2025 International Conference on Computational Innovations and Engineering Sustainability (ICCIES), Coimbatore, Tamilnadu, India, 2025, pp. 1-5, doi: 10.1109/ICCIES63851.2025.11032978

[14] D. Ganesh V, M. Ikram, S. O. Husain, J. Ballabh, T. Mohanraj and K. Aanandha Saravanan, "Cutting Force Prediction in Drilling Operations Using Temporal Convolutional Networks (TCN) and Spindle Speed Data," 2025 International Conference on Automation and Computation (AUTOCOM),Dehradun, India, 2025, pp. 1390-1395, doi: 10.1109/AUTOCOM64127.2025.10956789

[15] D. G. V, T. Erkinjon, L. H. Alzubaidi, K. Vijayakumar, K. R. Singh and A. K. Rahuman, "Revolutionizing Talent Acquisition in Human Resources with Advanced Deep Learning-Based Resume Screening and Candidate Matching Algorithms," 2025 International Conference on Automation and Computation (AUTOCOM), Dehradun, India, 2025, pp. 1289-1293, doi: 10.1109/AUTOCOM64127.2025.10957186

[16] Prashant Johri, Vincent Balu, B Jayaprakash, Aaditya Jain, Chintan Thacker, Anupam Kumari, "Quality of Service-Based Machine Learning in Fog Computing Networks For eHealthcare Services With Data Storage System", PP 1-13, published in Soft Computing (Springer), doi.org/10.1007/s00500-023-09041-8, Aug 2023.

[17] Prashant Johri, Seong Ki Kim, Kumud Dixit, Prakhar Sharma, Barkha Kakkar, Yogesh Kumar, Jana Shafi, Muhammad Fazal Ijaz, "Advanced Deep Transfer Learning Techniques for Efficient Detection of Cotton Plant Diseases" published in Frontiers in Plant Science, Vol 15, DOI 10.3389/fpls.2024.1441117, PP 1-22, Dec 2024.