Indonesian Journal of Electrical Engineering and Computer Science
Vol. 37, No. 2, February 2025, pp. 1077~1084
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v37.i2.pp1077-1084

Lung cancer prediction with advanced graph neural networks

Bineesh Moozhippurath, Jayapandian Natarajan
Department of Computer Science and Engineering, School of Engineering and Technology, CHRIST (Deemed to be University), Kengeri Campus, Bangalore, India

Article Info

Article history:

Received Mar 23, 2024
Revised Sep 18, 2024
Accepted Sep 30, 2024

Keywords:

Graph attention network
Graph neural network
Lung cancer
Machine learning
Prediction

ABSTRACT

This research aims to enhance lung cancer prediction using advanced machine learning techniques. The major finding is that integrating graph convolutional networks (GCNs) with graph attention networks (GATs) significantly improves predictive accuracy. The problem addressed is the need for early and accurate detection of lung cancer, leveraging a dataset from Kaggle's "Lung Cancer Prediction Dataset," which includes 309 instances and 16 attributes. The proposed A-GCN with GAT model is meticulously engineered with multiple layers and hidden units, optimized through hyperparameter adjustments, early stopping mechanisms, and Adam optimization techniques. Experimental results demonstrate the model's superior performance, achieving an accuracy of 0.9454, precision of 0.9213, recall of 0.9743, and an F1 score of 0.9482. These findings highlight the model's efficacy in capturing intricate patterns within patient data, facilitating early interventions and personalized treatment plans. This research underscores the potential of graph-based methodologies in medical research, particularly for lung cancer prediction, ultimately aiming to improve patient outcomes and survival rates through proactive healthcare interventions.

Corresponding Author:

Bineesh Moozhippurath
Department of Computer Science and Engineering School of Engineering and Technology
CHRIST (Deemed to be University)
Bangalore, India
Email: bineesh.m@res.christuniversity.in

1. INTRODUCTION

Lung cancer is a serious disease that imposes significant financial burdens on both the healthcare system and patients globally. Its prevalence and impact on public health are evident, as it ranks among the most common and deadly malignancies worldwide. The alarming increase in lung cancer cases underscores the importance of effective prevention, early detection, and cutting-edge treatment techniques. Early identification is crucial for establishing a patient's prognosis and potential outcomes, particularly because lung cancer often exhibits mild or nonexistent symptoms in its early stages [1]. Timely and accurate lung cancer prediction plays a pivotal role in identifying individuals at high risk and facilitating the development of individualized treatment programs, thereby enhancing patient quality of life and survival rates. The utilization of modern technologies for cancer prediction, including computational methods and the abundance of biomedical data, has gained momentum [2]. In the domain of cancer prediction, graph-based deep learning techniques have emerged as promising tools for processing complex biological data, such as genomics, proteomics, and metabolomics [3]. These techniques enable the modeling of intricate dependencies and interactions inherent in biological systems, which often exhibit a graph-like natural structure [4]. In order to improve the precision and efficacy of predictive models, this research targets predicting lung cancer using

graph-based deep learning techniques [5], specifically seeking to develop a robust predictive model capable of capturing hidden linkages within the data using the potential of graph neural networks, notably advanced-graph convolution networks (advanced-GCN) with graph attention networks (GAT) [6]. The combination of these techniques aims to enhance the model's capacity for prediction, which could revolutionize the ability to forecast early lung cancer and, as a result, have a positive effect on patient outcomes. In this article, it gives a comprehensive analysis of the proposed methodology and experimental findings, shedding light on the potential advancements in lung cancer prediction enabled by innovative graph-based deep learning algorithms [7].

Due to the ongoing difficulty in predicting, diagnosing, and treating lung cancer, there has been extensive research and the use of various computational techniques [7]. Here, it gives an in-depth analysis of related research in lung cancer prediction, highlighting various computational approaches that have been used to improve prediction precision and early diagnosis [8]. A fundamental role in predicting lung cancer is played by simple machine learning techniques, which provide efficient tools for data analysis and decision-making [9]. Predictive models are trained using supervised learning algorithms, a fundamental method [10]. In order to categorize lung cancer types or forecast disease progression, methods like support vector machines (SVM) and random forests are frequently used [11]. These methods use information derived from medical imaging or genomic data [12]. Clustering methods like K-means are used for patient categorization in unsupervised learning, which employs unlabeled data [13]. Recursive feature elimination is one of the feature selection techniques that helps find key predictors and enhance model performance [14]. These fundamental methods offer a vital framework for creating forecasting models and drawing vital conclusions from data on lung cancer [15]. Enhancing prediction accuracy through integration with cutting-edge algorithms and multi-modal data is a promising way to improve lung cancer prognosis and treatment planning [16].

The aim of this research [17] is to identify plasma metabolites as possible lung cancer diagnostic biomarkers, especially in the Chinese population. The authors examined a dataset that included 110 people with lung cancer and 43 healthy people using a cutting-edge multidisciplinary strategy that combines metabolomics and machine learning. 61 plasma metabolites were found by the authors using focused metabolomics analysis [18]. The threat posed by lung cancer to the world's health is significant, calling for specialized approaches for early identification and efficient treatment [19]. A crucial strategy in this effort has arisen in recent years with the integration of genomes and molecular data [20]. Understanding lung cancer at both the genetic and molecular levels has been completely transformed by genomics [21], which is the study of an organism's genes and their activities, and molecular data, which contains complex information about molecular interactions [22]. This paradigm shift has made it possible to identify genetic changes, biomarkers, and molecular pathways that are closely connected to the onset, progression, and response to treatment of lung cancer [23]. Utilizing genetic and molecular data makes it easier to create predictive models, enabling lung cancer early detection and accurate prognostication [24]. By enabling prompt interventions and optimizing treatment plans, these models have a tremendous potential to improve patient outcomes and, ultimately, survival rates. The research conducts a critical analysis of lung cancer, a widespread disease with significant clinical, histological, and molecular variation and high mortality rates [25]. The diversity in lung cancer underscores the need for personalized treatment plans. Deep learning algorithms have revolutionized lung cancer prediction by detecting deep patterns and connections in vast, complex datasets, particularly in genetics and medical imaging. They excel at extracting critical information from CT scans and X-rays, aiding in early detection and diagnosis by identifying subtle abnormalities often missed by the human eye. Additionally, deep learning models integrate multi-omics data, such as proteomics and genomics, providing a comprehensive understanding of the molecular causes of lung cancer, which is crucial for tailoring treatment plans and improving outcomes. The research delves into gene-environment interactions affecting cellular homeostasis and cancer development, highlighting the potential of biomarkers as early indicators of disease progression. The advent of high-throughput profiling methods, like mass spectrometry and RNA sequencing, has facilitated data-driven biomarker discovery. While traditional statistical methods have limitations, the combination of machine learning (ML) and deep learning (DL) offers a powerful approach in disease research. By integrating ML/DL techniques and optimizing them across multi-omics datasets, robust prediction models essential for precision medicine are developed. The paper explains recent developments in ML/DL methods, especially in the identification of cancer-related biomarkers and their use in precision medicine. The potential for individualized patient treatment plans based on medical histories and diagnostics has been shown by the wide range of applications of DL techniques, particularly in the diagnosis and treatment of cancer. In order to carry out precise therapeutic interventions for cancer, the review highlights the critical need for DL-based multimodal ensemble techniques that integrate gene-protein-metabolite interactions. However, there are certain drawbacks to DL models, such as small sample sizes, difficulties with interpretability, the demand for processing resources, and the requirement for expert knowledge.

Graph neural networks (GNNs) have emerged as a promising approach for predicting lung cancer by capturing complex relationships within biological systems. Representing molecular and genetic data as graphs—where nodes are biological entities like genes or proteins, and edges represent their interactions—GNNs excel in modeling the intricate molecular networks and pathways critical to lung cancer progression. They can identify key genes or proteins associated with lung cancer, predict cancer cell behavior, and suggest potential treatment targets. The integration of multi-omics data, encompassing genomic, epigenetic, and proteomic information, further enhances prediction accuracy. This study also leverages medical image processing techniques to examine internal anatomy non-invasively, aiding in more precise diagnoses and better treatment planning. Specifically, the research focuses on lung tumor prediction using a hybrid evolutionary approach that combines bacterial foraging optimization with whale optimization. By preprocessing and segmenting 150 CT images, the novel Whale-Bacterial foraging optimization technique is applied to extract optimal features, demonstrating superior performance in distinguishing between cancerous and non-cancerous nodules compared to traditional methods.

The outline of the proposed research entails the integration of graph neural networks (GNNs) to enhance precision and early diagnosis in lung cancer prediction and treatment. This approach underscores the significance of proactive patient care and personalized treatment plans in improving outcomes and survival rates. The initial hypothesis postulates that by integrating GNNs, particularly the A-GCN with the GAT model tailored for lung cancer prediction, increased accuracy and efficacy can be achieved, ultimately leading to enhanced patient outcomes and proactive healthcare interventions.

2. METHOD

The lung cancer prediction dataset, sourced from Kaggle's "Lung Cancer Prediction Dataset," comprises 309 instances with 16 attributes, including demographic and lifestyle factors such as gender, age, smoking habits, alcohol consumption, and peer pressure. Data preprocessing is a critical step aimed at enhancing the suitability of the dataset for machine learning algorithms. Duplicate entries were identified and removed to maintain data integrity, and label encoding was applied to categorical features like "GENDER" using scikit-learn's LabelEncoder. To address class imbalance, RandomOverSampler from imbalanced-learn was utilized to ensure a balanced training set. To harness the relational information inherent in the dataset, an adjacency matrix was constructed to represent patient-patient relationships. The matrix values were set to 1 if both patients had lung cancer and 0 otherwise. Transforming the adjacency matrix into edge indices allowed for the utilization of graph-based neural network architectures, enhancing the analysis and prediction capabilities. The final preprocessing steps involved converting the dataset into PyTorch tensors and organizing it into data objects. These objects, namely "train_data" and "test_data," house the training and testing data along with their respective edge indices. Figure 1 illustrates the general structure of the prediction model.

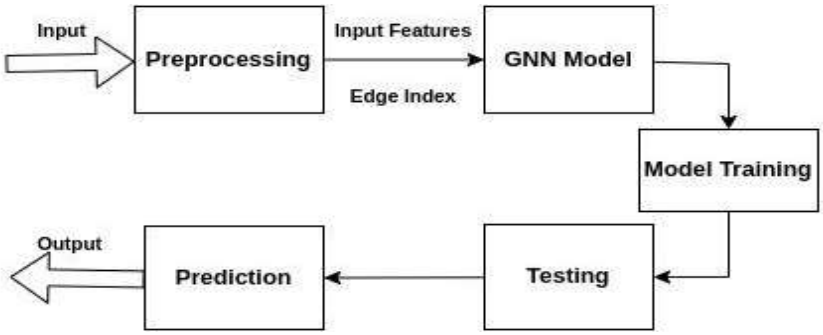

Figure 1. Structure of prediction model

A crucial stage in our research method is model integration, which unites disparate predictive models into a cohesive framework for improved lung cancer prediction. Ensemble learning techniques, such as stacking or averaging, were employed to integrate the outputs of individual models and reduce biases. Rigorous validation and cross-validation procedures were conducted to fine-tune the ensemble and achieve an ideal balance between predictive power and model variety. Figure 2 is the proposed advanced GCN with GAT architecture.

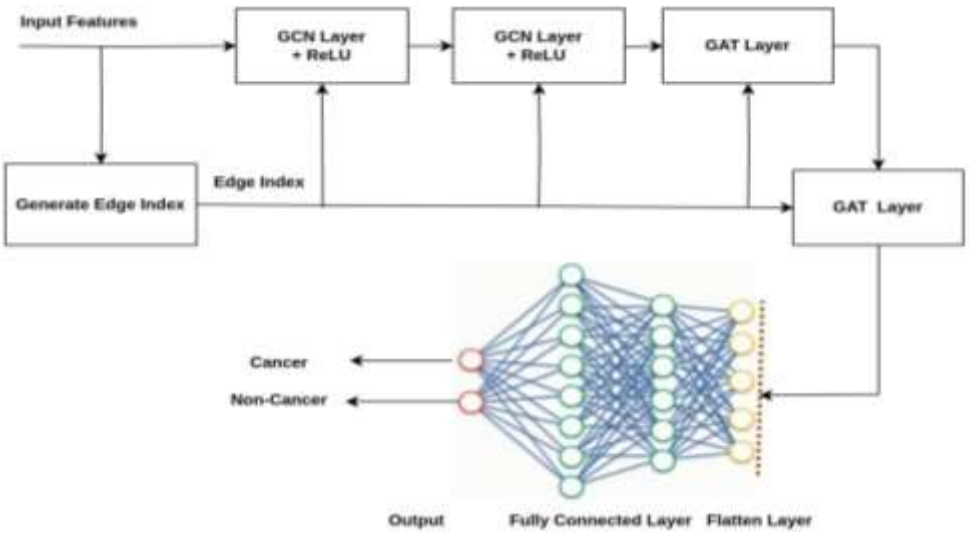

Figure 2. Proposed advanced GCN-GAT architecture

Graph convolutional networks (GCNs) and graph attention networks (GATs) are the two models in the ensemble that show distinct advantages in identifying patterns in the patient dataset. The proposed approach aims to enhance the graph convolutional network (GCN) structure and combine it with GAT model. This comprehensive methodological approach seeks to provide a highly advanced and dependable tool for the early detection and treatment of lung cancer. The proposed model structure is elucidated in the form of pseudocode in the subsequent section.

Algorithm 1. GCN-GAT model definition

```
Initialize Model Parameters:
        1.1. Set in_features: Number of input features.
        1.2. Set hidden_dim: Dimension of the hidden layer.
        1.3. Set out_features: Number of output features.
        1.4. Set heads: Number of attention heads in GAT (default: 4).
        1.5. Set concat: Concatenate attention heads (default: True).
Define Model:
        2.1. Create a class GCN_GAT that inherits from nn.Module.
        2.2. Initialize the model parameters: in_features, hidden_dim, out_features, heads,
        and concat.
Define Layers:
        3.1. Define the first Graph Convolutional layer (GCN): gc1.
        3.2. Define the second Graph Convolutional layer (GCN): gc2.
        3.3. Define the Graph Attention layer (GAT): gat_layer.
        3.4. Define an additional Graph Attention layer (GAT): gat_layer1.
        3.5. Define the fully connected layer: fc.
Forward Pass:
        4.1. Define the forward method to perform the forward pass.
        4.2. Apply ReLU activation to the output of the first GCN layer (gc1).
        4.3. Apply ReLU activation to the output of the second GCN layer (gc2).
        4.4. Apply dropout for regularization.
        4.5. Apply the first Graph Attention layer (GAT) (gat_layer).
        4.6. Apply the second Graph Attention layer (GAT) (gat_layer1).
        4.7. Flatten the tensor before passing it to the fully connected layer.
        4.8. Apply the fully connected layer (fc).
        4.9. Apply the sigmoid activation for the output.
Return Output:
        5.1. Return the sigmoid-activated output representing class probabilities.
```

Interestingly, a specific collection of six metabolic variables demonstrated exceptional discriminating power, effectively differentiating between patients with stage I lung cancer and healthy controls (AUC=0.989, sensitivity=98.1%, specificity=100.0%). Furthermore, the top five metabolic biomarkers identified by the FCBF algorithm demonstrated potential as screening biomarkers for lung cancer early detection. It was proposed that the Naive Bayes algorithm might be a helpful tool for early lung tumor prediction. This multidisciplinary approach not only validates blood-based screening but also offers an accurate, rapid, and comprehensive diagnostic tool for early lung cancer diagnosis. Beyond lung cancer, this

interdisciplinary methodology's potential applications open doors for improving early cancer diagnosis in a wider context. This study provided a ground-breaking multidisciplinary approach that successfully combines metabolomics and machine learning to identify early lung cancer diagnostic indicators. The discovered metabolic indicators exhibit impressive diagnostic power, especially for the early diagnosis of lung tumors. It is necessary to create more precise prediction methods because the performance of current models for early-stage lung cancer detection depending on germline variations has been found to be restricted. This study used a multicenter case-control study to make a powerful predictive method for early lung cancer diagnosis using the powerful machine learning technique of extreme gradient boosting (XGBoost). The findings showed a strong correlation between certain single nucleotide polymorphisms (SNPs), including TYMS rs3819102 and BAG6 rs1077393, and the chance of developing lung cancer. Additionally, the XGBoost model's predictive performance was greatly enhanced by the addition of SNPs and epidemiological data, excelling particularly in the prediction of squamous cell cancer (SCC). This cutting-edge multimodal strategy demonstrates a promising route for improving early lung cancer diagnosis and justifies additional investigation for applications of precision medicine in cancer diagnostics.

The research was carefully crafted to explore how advanced graph-based neural networks perform in predicting different levels of lung cancer severity. Utilizing the lung cancer dataset comprising 309 records, the experiment employed random oversampling techniques to ensure data robustness. The training dataset, comprising 380 instances (80%), and the testing dataset, with 96 instances (20%), were carefully selected. To encode the data, a label encoding approach was adopted, and models were constructed with 2 layers each for GCNs and GATs, along with an embedding dimension (dim) set to 64. Training iterations spanned 100 epochs, utilizing an initial learning rate of 0.01 and a dropout rate of 0.25. The experimental setup leveraged Python version 3.10.12 and PyTorch 2.0 software, executed on a Tesla T4 GPU server with 15 GB of GPU memory, operating under Ubuntu 20.04 LTS.

3. RESULTS AND DISCUSSION

The results of the AGCN-GAT model were further reinforced by integrating other models to establish a robust comparison baseline for a thorough analysis. In comparing the performance of various models in predicting lung cancer risk, the proposed advanced GCN emerges as a standout performer. The advanced GCN demonstrates superior predictive capabilities with an accuracy of 0.9454, precision of 0.9213, recall of 0.9743, and an F1 score of 0.9482, outperforming models like linear regression, KNN, random forest, and neural networks. Notably, while linear regression achieved high precision, it faltered in recall, indicating potential limitations in capturing all relevant instances. Similarly, KNN exhibited comparatively lower accuracy and precision, suggesting a less robust predictive performance. Random forest and neural network models demonstrated commendable overall performance, yet their accuracy and precision fell short of the advanced GCN. Table 1 is deliberate the experimental setup details.

Table 1. Experimental setup parameters

Parameters	Values
Dataset	Lung Cancer Dataset
Dataset size	309 records
Data Oversampling	Random Oversampling
Training & Testing dataset size	380 (80%) & 96 (20%)
Number of GCN layers & Attention layers	2
Embedding Dimension (dim)	64
Number of training iterations	100
Initial learning rate & Dropout rate	0.01 & 0.25
Software	Python 3.10.12, PyTorch 2.0

The advanced GCN model's exceptional performance underscores its effectiveness in capturing intricate patterns and relationships within the lung cancer prediction dataset, showcasing its potential for enhancing early detection efforts and personalized risk assessments. These results highlight the promise of advanced graph-based methodologies in medical research, particularly in the realm of lung cancer prediction, opening avenues for further exploration and development of predictive models for proactive healthcare interventions. Table 2 is a deliberate performance analysis of the existing and proposed models. Figure 3 is deliberate comparison between accuracy and precision levels. Figure 4 shows the deliberate recall value in comparing the performance of various models in predicting lung cancer risk; Figure 5 shows the proposed model’s precision, recall, and F1 score performance.

Table 2. Performance analysis of existing and proposed model

Model	Accuracy	Precision	Recall	F1 Score
Linear regression	0.91	0.90	1.00	0.95
KNN	0.79	0.79	1.00	0.88
Random forest	0.89	0.90	0.98	0.93
Neural network	0.89	0.91	0.95	0.93
Advanced GCN-GAT	0.95	0.92	0.97	0.95

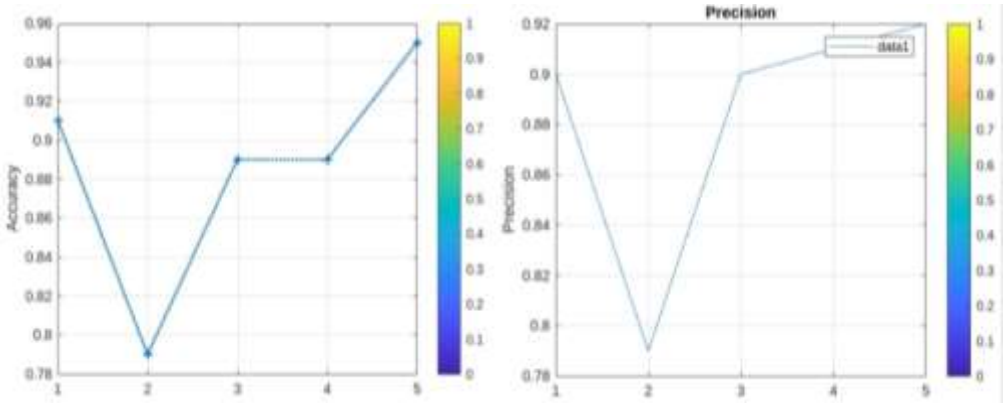

Figure 3. Accuracy and precision level comparison

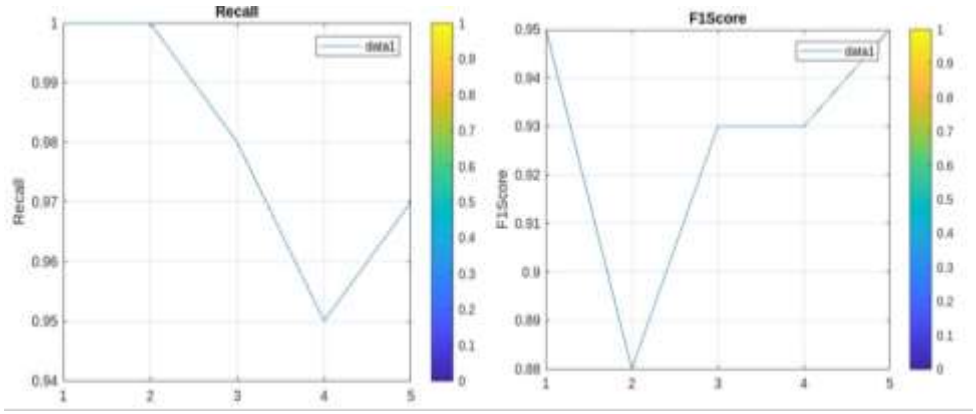

Figure 4. Recall and F1 score comparison

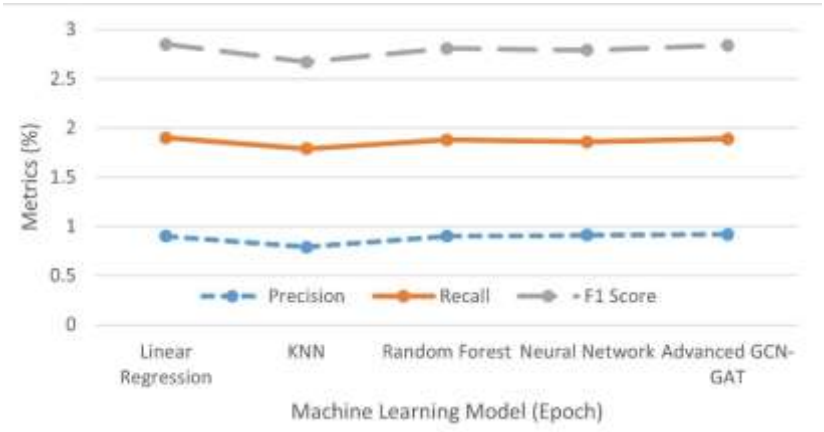

Figure 5. Proposed AGCN-GAT model performance

The advanced GCN model's exceptional performance underscores its effectiveness in capturing intricate patterns and relationships within the lung cancer prediction dataset, showcasing its potential for enhancing early detection efforts and personalized risk assessments. These results highlight the promise of advanced graph-based methodologies in medical research, particularly in the realm of lung cancer prediction, opening avenues for further exploration and development of predictive models for proactive healthcare interventions. Future research should focus on expanding datasets to include diverse patient populations, integrating multi-omics data for improved predictive power, developing interpretability frameworks for GNNs, and conducting longitudinal studies to assess real-world efficacy. Key experiments should validate the model on external datasets, optimize GNN architectures and hyperparameters, integrate clinical text data with structured data, and perform ablation studies to understand the AGCN-GAT model's components. Despite promising results, limitations include reliance on a small dataset, significant computational resource requirements, and challenges in interpretability. Addressing these limitations involves enhancing model interpretability, efficiency, and validation across larger, more diverse datasets.

4. CONCLUSION

This research introduces an innovative approach to lung cancer prediction, aiming to enhance accuracy and efficacy in the early detection of lung cancer. By integrating an advanced GCN with GAT architecture, the proposed model exhibits remarkable performance improvements over existing models, with a significant increase in accuracy by 6% and precision by 2%, compared to the best-performing baseline model. This advancement holds significant promise for improving patient outcomes and facilitating proactive healthcare interventions. The successful integration of GCN and GAT methodologies demonstrates the potential for graph-based approaches to tackle complex medical prediction tasks, extending beyond lung cancer to other diseases with intricate data structures. These improvements suggest a crucial role for graph-based models in personalized medicine, enabling earlier and more accurate risk identification and tailored interventions. This study underscores the transformative potential of advanced machine learning techniques in medical research, setting the stage for future innovations in oncology and beyond.

REFERENCES

[1] C. M. Rudin, E. Brambilla, C. Faivre-Finn, and J. Sage, "Small-cell lung cancer," *Nature Reviews Disease Primers*, vol. 7, no.1, pp. 1-12, 2021, doi:10.1038/s41572-020-00235-0.

[2] A. Sharma and R. Rani, "A systematic review of applications of machine learning in cancer prediction and diagnosis," *Archives of Computational Methods in Engineering*, vol. 28, no.7, pp.4875-4896, 2021, doi:10.1007/s11831-021-09556-z.

[3] D. A.-Aristizabal, M. A. Armin, S. Denman, C. Fookes, and L. Petersson, "A survey on graph-based deep learning for computational histopathology," *Computerized Medical Imaging and Graphics*, vol. 95, no.1, pp. 102027-1002039, 2022, doi: 10.1016/j.compmedimag.2021.102027.

[4] S. Tagore, N. Chowdhury, and R. K. De, "Analyzing methods for path mining with applications in metabolomics," *Gene*, vol. 534, no. 2, pp. 125-138, 2014, doi: 10.1016/j.gene.2013.10.056.

[5] W. Zhang, J. Chien, J. Yong, and R. Kuang, "Network-based machine learning and graph theory algorithms for precision oncology," *NPJ precision oncology*, vol. 1, no. 1, pp. 25-35, 2017, doi: 10.1038/s41698-017-0029-7.

[6] P. Wang and Z. Zhao, "Improving context and syntactic dependency for aspect-based sentiment analysis using a fused graph attention network," *Evolutionary Intelligence*, vol. 17, no.1, pp. 589-598, 2023, doi: 10.1007/s12065-023-00845-z.

[7] X. Luo, X. Zang, L. Yang, J. Huang, F. Liang, J. Rodriguez-Canales, and G. Xiao, "Comprehensive computational pathological image analysis predicts lung cancer prognosis," *Journal of Thoracic Oncology*, vol. 12, no. 3, pp. 501-509, 2017, doi: 10.1016/j.jtho.2016.10.017.

[8] S. Gupta and M. K. Gupta, "A comparative analysis of deep learning approaches for predicting breast cancer survivability," *Archives of Computational Methods in Engineering*, vol. 29, no. 5, pp. 2959-2975, 2022, doi: 10.1007/s11831-021-09679-3.

[9] F. J. Shaikh, and D. S. Rao, "Prediction of cancer disease using machine learning approach," *Materials Today: Proceedings*, vol. 50, no. 1, pp. 40-47, 2022, doi: 10.1016/j.matpr.2021.03.625.

[10] Y. Lin, K. Zhou, and J. Li, "Prediction of slope stability using four supervised learning methods," *IEEE Access*, vol. 6, no. 1, pp. 31169-31179, 2018, doi:10.1109/access.2018.2843787.

[11] A. Kurani, P. Doshi, A. Vakharia, and M. Shah, "A comprehensive comparative study of artificial neural network (ANN) and support vector machines (SVM) on stock forecasting," *Annals of Data Science*, vol. 10, no. 1, pp. 183-208, 2023, doi: 10.1007/s40745-021-00344-x.

[12] M. L. Giger, "Machine learning in medical imaging," *Journal of the American College of Radiology*, vol. 15, no. 3, pp. 512-520, 2018, doi: 10.1016/j.jacr.2017.12.028.

[13] C. M. Eckhardt *et al.*, "Unsupervised machine learning methods and emerging applications in healthcare," *Knee Surgery, Sports Traumatology, Arthroscopy*, vol. 31, no. 2, pp. 376-381, 2023, doi: 10.1007/s00167-022-07233-7.

[14] X. Huang, L. Zhang, B. Wang, F. Li, and Z. Zhang, "Feature clustering-based support vector machine recursive feature elimination for gene selection," *Applied Intelligence*, vol. 48, no.1, pp. 594-607, 2018, doi: 10.1007/s10489-017-0992-2.

[15] A. B. Pawar *et al.*, "Implementation of blockchain technology using extended CNN for lung cancer prediction," *Measurement: Sensors*, vol. 24, no.1, pp. 100530-100545, 2022, doi: 10.1016/j.measen.2022.100530.

[16] G. Hou *et al.*, "Deep learning approach for predicting lymph node metastasis in non-small cell lung cancer by fusing image–gene data," *Engineering Applications of Artificial Intelligence*, vol. 122, no.1, pp. 106140-106153, 2023, doi: 10.1016/j.engappai.2023.106140.

[17] Y. Xie *et al.*, "Early lung cancer diagnostic biomarker discovery by machine learning methods," *Translational oncology*, vol. 14, no. 1, pp: 100907-100918, 2021, doi: 10.1016/j.tranon.2020.100907.
[18] Y. Zhao *et al.*, "Metabolic profiles characterizing different phenotypes of polycystic ovary syndrome: plasma metabolomics analysis," *BMC medicine*, vol. 10, no.1, pp. 1-12, 2012, doi: 10.1186/1741-7015-10-153.
[19] A. Kerpel-Fronius *et al*., "Screening for lung cancer in individuals who never smoked: an international association for the study of lung cancer early detection and screening committee report," *Journal of Thoracic Oncology*, vol. 17, no. 1, pp. 56-66, 2022, doi: 10.1016/j.jtho.2021.07.031.
[20] M. U. Ahsan, Q. Liu, J. E. Perdomo, L. Fang, and K. Wang, "A survey of algorithms for the detection of genomic structural variants from long-read sequencing data," *Nature Methods*, vol. 20, no. 8, pp. 1143-1158, 2023, doi: 10.1038/s41592-023-01932-w.
[21] A. Ravi *et al.*, "Genomic and transcriptomic analysis of checkpoint blockade response in advanced non-small cell lung cancer," *Nature genetics*, vol. 55, no. 5, pp. 807-819, 2023, doi: 10.1038/s41588-023-01355-5.
[22] X. Hu, Z. Zeng, J. Zhang, D. Wu, H. Li, and F. Geng, "Molecular dynamics simulation of the interaction of food proteins with small molecules," *Food Chemistry*, vol. 405, no.1, pp. 134824-134838, 2023, doi: 10.1016/j.foodchem.2022.134824.
[23] J. Huang *et al.*, "Identification of a disulfidptosis-related genes signature for prognostic implication in lung adenocarcinoma," *Computers in Biology and Medicine*, vol. 165, no.1, pp. 107402-107423, 2023, doi: 10.1016/j.compbiomed.2023.107402.
[24] L. Manganaro *et al.*, "Consensus clustering methodology to improve molecular stratification of non-small cell lung cancer," *Scientific Reports*, vol. 13, no. 1, pp. 7759-7772, 2023, doi: 10.1038/s41598-023-33954-x.
[25] M. S. Hussain *et al.*, "Long non-coding RNAs in lung cancer: Unraveling the molecular modulators of MAPK signaling," *Pathology-Research and Practice*, vol. 249, no. 1, pp. 154738-154750, 2023, doi: 10.1016/j.prp.2023.154738.

BIOGRAPHIES OF AUTHORS

Bineesh Moozhippurath is currently pursuing a Ph.D. in the department of computer science and engineering at Christ University, Bangalore, focusing on cancer prediction using metabolomics & machine learning. He completed his M.E. (computer and communication) from Anna University, Tamil Nadu in 2011. Bineesh holds a bachelor's degree in Information Technology (B.Tech.) from Cochin University of Science and Technology, Kerala in 2006. His research interests include machine learning, graph neural networks, and metabolomics. He can be contacted at email: bineesh.m@res.christuniversity.in.

Jayapandian Natarajan is currently working as associate professor in the department of computer science and engineering at Christ University, Bangalore. He has received his Ph.D. from Anna University, Chennai. He is active life member of ISTE. He is currently doing his research in machine learning and cloud computing. In his 15 years of teaching experience and one year of industry experience. His research interests are grid computing and cloud computing. He has published in 4 book chapters, 35 International Journal articles, 100 international and National Conferences. He can be contacted at email: jayapandian.n@christuniversity.in.