IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2025

DOI: 10.17148/IJARCCE.2025.14512

An Innovative Diagnostic Framework for Lung Cancer Detection

P. Parameswari[1], S. Sathish Kumar[2]

Principal, Palanisamy Arts College, Erode, Tamil Nadu, India[1]

PhD Research Scholar, Karuppanan Mariappan College, Tirupur, Tamil Nadu, India[2]

Abstract: The disease known as cancer is typified by abnormal cell proliferation that spreads throughout the body. In the lungs, abnormal cell growth leads to lung cancer. The lungs, the body's main respiratory control system, ensure that oxygen reaches every part of the body. It purifies the air and prevents infections and unwanted substances from entering the body. According to our immune system, every organ can battle inflammation and infections. Sometimes, though, they fall short in the fight against these infections, inflammations, and even malignant cells. This will inevitably lead to the development of cancer. Stages 0 through 4 are used to classify lung cancer. Early detection of lung cancer, either stage 0 or stage 1, increases a patient's chances of survival. If the cancer is found in its advanced stages, the chances of survival are quite low. Early identification of breast cancer is therefore essential. Many medical diagnostic methods, including X-rays and lung cancer screening, are available for the prediction of lung cancer. However, there are instances in which these diagnostic methods result in false positives or false negatives, requiring patients to get needless medical care. To avoid these outcomes linked to lung cancer projections, alternative approaches are needed. Even while there are other computerised methods for predicting lung cancer, they are also not very accurate. Therefore, using the lung cancer patient databases from Kaggle, we have created three models—Kernel Optimised Neural Network (KONN), Hierarchical Optimisation Neural Network (HONN), and Neural Adaptive Transformer Optimiser classifier method (NATO)—to predict lung cancer in its early stages. Along with the suggested efforts, the dataset is pre-processed using SMOTE to address the issues of class imbalance. Together with the training time for each suggested model, the performance of these methods is evaluated using the following metrics: accuracy, precision, and recall. When compared to the other two suggested models, such as Kernal Optimised Neural Network and Hierarchical Optimisation Neural Network, the Neural Adaptive Transformer Optimiser classifier approach in this research gives better accuracy and requires less training time.

Keywords: HONN, KONN. Lung Cancer, NATO, SMOTE.

I. INTRODUCTION

There are numerous challenges in detecting lung cancer accurately by a clinical decision support system by analysing the clinical and genetic dataset that impact the performance and efficiency. Firstly, detecting lung cancer with better efficiency by the same model faces multiple challenges. The challenges include imbalanced data distribution, high dimensionality, improper feature learning, and complex classifier structures. The outlier removal and missing values can be detected and eliminated. Class imbalance reduces the classifier model's overall efficacy and increases the need for specialized training. The inclusion of irrelevant features and less important features increases the dimensionality and also degrades the predictive accuracy. Additionally, the classifier complexity increases resource utilization with increased processing time. All these limitations degrade the model's performance and reduce its predictive efficacy. A robust and accurate model is developed in this research to overcome these challenges.

This research focuses on improving disease prediction accuracy and efficiency to enhance patient treatment and reduce healthcare costs. This research aims to create models to improve accuracy with faster and more reliable diagnoses, contributing to better clinical decision-making. The models focus on developing effective feature learning capabilities that gather more profound insights into the data, leading to a robust prediction and classification model. This research is motivated by the potential of making significant contributions to the healthcare field by improving the tools available for early disease detection, thereby enhancing patient treatment outcomes. The objectives are developed to overcome and address the challenges in creating a predictive model for lung cancer classification. That is to develop a hybrid classifier that can effectively handle the class imbalance problems. To improve disease prediction accuracy, enhance the feature learning process and reduce the high dimensionality problem. Improve the classifier accuracy by adapting a novel model to reduce the high complexity and training time.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2025

DOI: 10.17148/IJARCCE.2025.14512

II. LITERATURE SURVEY

Researchers have developed various studies for the detection and prediction of lung cancer using both clinical and genetic data. This innovative approach aims to improve early diagnosis and treatment outcomes by identifying at-risk individuals more accurately. By integrating advanced machine learning techniques with comprehensive datasets, researchers hope to enhance their understanding of lung cancer's complexities and inform personalized medicine strategies. Riaz et al. [1] proposed a hybrid neural network model by using the benefits of MobileNetV2 and UNet architectures. It is applied to an MSD dataset, and results depict the score accuracy value of 0.8793, the recall value of 0.8602 and the precision value of 0.93. However, the model has a slightly increased dimensionality problem. Mohammed et al. [2] suggested using Multi-Cancer Multi-Omics Clinical Dataset Laboratories (MCMOCL) methods to predict different types of cancers, which include federated learning, auto-encoder, and XGBoost techniques to make predictions more accurate and faster. The results outperformed previous methods, achieving an accuracy improvement of 98% and a reduction in processing delay by 61%. However, the model has a high dimensionality problem. Ismail [3] proposed a U-Net Convolutional Network classifier specifically designed for biomedical image segmentation to predict lung cancer. This model showed a higher number of incorrect positive results compared to neural networks, which had 6.78 false positives for each case, while radiologists had between 0.33 and 1.39 false positives per case, but it does have a small increase in complexity.

Chandran et al. [4] proposed the Observational Medical Outcomes Partnership Common Data Model, which enables standardization of data from disparate observational databases into a common format to facilitate standardized analytics and efficiencies. The outcome to be predicted was the risk of new lung cancer starting 1 day to 3 years (1,095 days) after the index. However, these methods are computationally expensive. Mohammad et al. [5] developed an enhanced CNN model that excels with 100% accuracy, 99.2% precision, 98.0% recall, and an F1 score of 98.4%, indicating its strong ability to accurately detect cases while minimizing errors. However, current challenges such as poor image quality, physician stress, inadequate information, and ineffective communication often hinder early detection, leading to escalated medical costs and further deterioration of patient health. Sandeep et al. [6] developed the VGG16 model, which produced excellent performance in classification for image classification tasks, with an accuracy rate of 91.2%. However, the model has limitations in handling the class imbalance problem. Satya Prakash et al. [7] proposed the hybridization of the K-Nearest Neighbor model and the Bernoulli Naive Bayes model for classification, which gives an accuracy value of 92.86%. However, the model has overfitting issues with training and complexity due to the ensemble model. Yusupha et al. [8] proposed a comprehensive Machine Learning (ML) model designed to predict lung cancer at an early stage, utilizing a dataset sourced from Kaggle. This model gives an accuracy value of 97.9%, a recall value of 98.2%, and a precision value of 98.4 %, demonstrating its transformative potential for early lung cancer detection. However, the model lacked data pre-processing techniques. Therefore, effective methods must be designed to overcome and solve the existing constraints.

III. RESEARCH CONTRIBUTIONS

This research is divided into three phases for early detection and accurate classification of lung cancer, each focusing on specific model development and optimization aspects. Figure 1 shows the block diagram of the contributions.

Proposed Model 1: Kernel Optimized Neural Network Algorithm-based SMOTE prediction model

This proposed model utilizes the SMOTE method as a pre-processing tool and a hybrid classifier for classification. The pre-processing methods include handling missing values and cleaning data to improve the data quality. This process increases the dataset's integrity and improves the accuracy of the prediction models. The features are extracted, selected, and given to the hybrid model to ensure higher accuracy for classification. The hybrid classifier is developed by integrating Divide and Conquer Kernel-based Support Vector Machine (DCKSVM) with Radial Basis Function Neural Network (RBFNN) techniques to improve the performance of the classification model to solve the class imbalance problem. DCKSVM is applied since it has been proven as a good classifier, and the results from the DCKSVM are tuned using RBFNN. The DCKSVM model is used as the initial classification, and RBFNN is used to refine and validate the results. The proposed hybrid algorithm is effective for predicting lung cancer.

Proposed Model 2: Disease Detection using Hierarchical Optimization Neural Network (HONN) Classifier

This proposed research aims to focus on reduced redundancy and improved feature quality by grouping related features of lung cancer. The model used pre-processing steps with the SMOTE Method. The optimal features are selected using Feature aggregation with Label encoding to reduce the higher feature dimensionality problem and also to reduce redundancy. This technique also aims to identify and rank features based on their correlation with the target variable. Label encoding is used in this research to convert categorical data into numerical values, making it compatible with machine learning algorithms.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2025

DOI: 10.17148/IJARCCE.2025.14512

After feature selection, the proposed Hierarchical Optimization Neural Network classifier is applied to enhance the neural network by optimizing its hyperparameters, leading to better accuracy and faster convergence. It is also used to solve the class imbalance problem and high dimensionality problem and helps to improve disease accuracy. This process increases performance and accuracy in predicting lung cancer classification tasks.

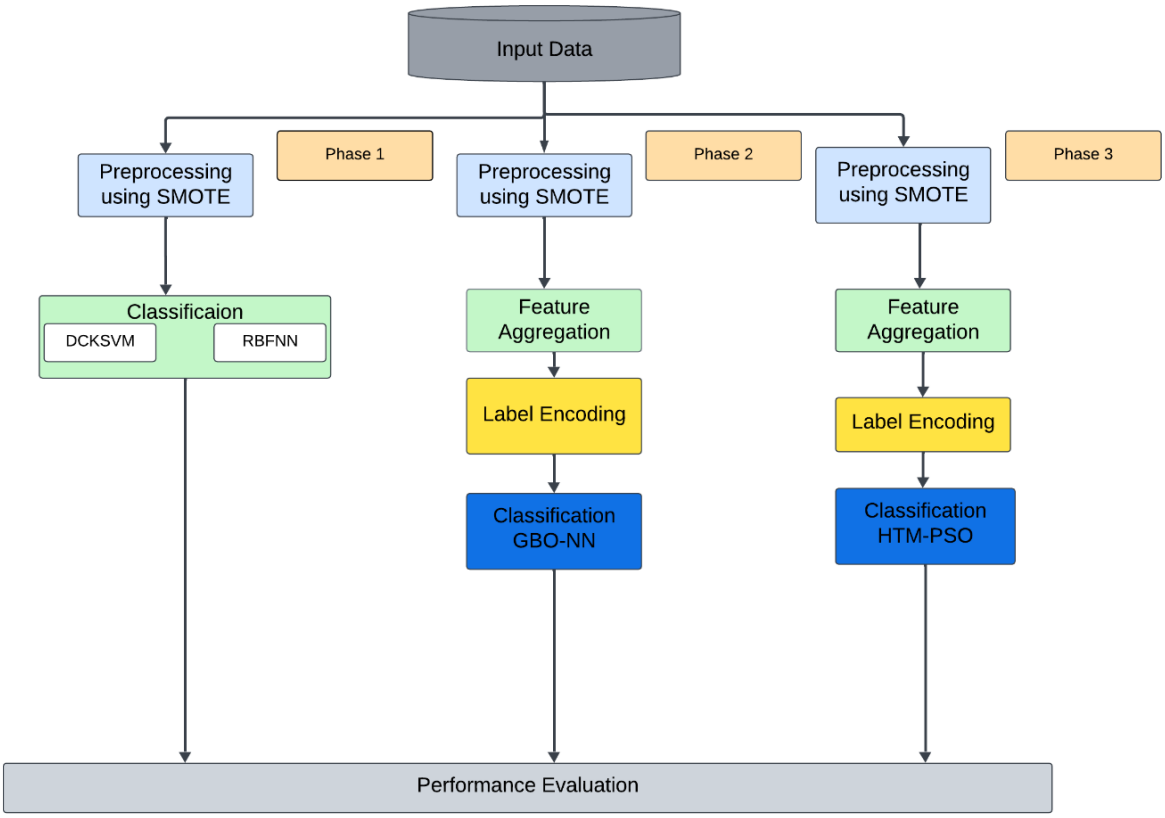

Figure 1. Block Diagram of the Proposed Research Contributions

Proposed Model 3: Disease Detection using Neural Adaptive Transformer Optimizer (NATO) Classifier

This proposed approach tackles the ML classifiers by incorporating the transformer model. It addressed feature relevance and sequence dependencies and improved optimization stability and reduced complexity. This proposed algorithm also shallow feature learning process and model complexity problems too. This proposed model consists of three stages: data pre-processing for removing outliers, missing value imputation and normalization using SMOTE method, feature selection using feature aggregation with label encoding and classification by Neural Adaptive Transformer Optimizer Model Classifier. In the pre-processing steps, the outliers are identified and removed using the SMOTE method and the removed outlier values are considered to be missing values. These missing values are imputed by applying the SMOTE method. After handling the missing values, the data are normalized. Each feature is ranked with a specific target variable, and the features are extracted when a high absolute correlation value is present. The extracted features are selected using the Feature Aggregation along with the Label Encoding, which removes the worst solution and retains the best solution to solve the shallow feature learning problem. The Neural Adaptive Transformer Optimizer Model classifier is developed by combining a deep architecture to execute Machine learning methods for better classification tasks. The Neural Adaptive Transformer Optimizer Model includes a weighted model to handle imbalanced class distribution within data and to attain generalization. The regularized model is applied to reduce the over-fitting problem and to reduce training time. The Neural Adaptive Transformer Optimizer Model is a hybrid algorithm that combines the hybrid transformer model with Particle Swarm Optimization to focus on important features, improving interpretability, accuracy, and reducing training time. The Neural Adaptive Transformer Optimizer Model algorithm produces a powerful and efficient method for fine-tuning and produces a more accurate and robust model for detecting lung cancer with less processing time when compared to the other two proposed algorithms.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2025

DOI: 10.17148/IJARCCE.2025.14512

IV. RESEARCH CONTRIBUTIONS

The proposed research for lung cancer detection is implemented using the Python language on a PyCharm tool, with the Intel Core i5 processor system configuration, Windows 10 OS with 8GB RAM, and a 512GB SSD. In this research, the lung cancer database is available on Kaggle, which is used for the early diagnosis of lung cancer. This dataset has 309 patient records, which comprise both cancerous and non-cancerous. The non-cancerous category is the 90% data, where 270 records belong to this category, and only 10%, which is 39 records, belong to the cancerous patient's category.
Each record in this dataset has different features, and the values of these features are either numerical or text values. Among these features, 16 features are selected and will be used for analysis and prediction of the proposed algorithms in the lung cancer dataset.

The performance evaluation and comparison results of the proposed methods are analyzed in terms of accuracy and processing time for the lung cancer dataset, as given in Table 1, which represents the classification results for the proposed Kernel Optimized Neural Network, Hierarchical Optimization Neural Network, and Neural Adaptive Transformer Optimizer models. The Neural Adaptive Transformer Optimizer Model exhibited higher accuracy with less training time than other proposed models.

Table 1. Comparison Results for Proposed Models

Models	**Lung Cancer Dataset**			
	Accuracy (%)	**Precision (%)**	**Recall (%)**	**Processing time (s)**
Kernal Optimized Neural Network (KONN)	92.28	91.15	89.25	142.21
Hierarchical Optimization Neural Network (HONN)	94.36	99.73	99.70	141.12
Neural Adaptive Transformer Optimizer (NATO)	96.68	98.28	97.79	138.16

The results are shown in the Table. Figure 1 depicts the values of accuracy, precision, and recall with the processing time. The proposed Neural Adaptive Transformer Optimizer classifier method has achieved the highest accuracy value of 96.68%, with 138.16 seconds taken as a processing time for the lung cancer dataset than the other two proposed models. The proposed Kernel Optimized Neural Network and Hierarchical Optimization Neural Network methods have also performed well and outperformed the existing methods.

V. CONCLUSION AND FUTURE ENHANCEMENT

The research has focused on addressing the critical challenges associated with accurately detecting and predicting lung cancer using advanced computational techniques. The research developed and implemented three novel models, namely the kernel optimized neural network classifier, the hierarchical optimization neural network, and the neural adaptive transformer optimizer, significantly enhancing the predictive accuracy and efficiency in detecting these diseases. The proposed models demonstrated effectiveness in handling classification tasks, addressed the challenges of class imbalance and high dimensionality in complex disease datasets, and focused on shallow feature learning and over-fitting problems with less processing time. These proposed models improved prediction accuracy by refining the classification results, ensuring a more robust model for disease detection. These contributions advanced medical diagnostics and provided timely healthcare decision-making while improving patient outcomes.

REFERENCES

[1] Riaz Z.,Bangul Khan, Saad Abdullah. (2023). Lung tumor image segmentation from computer tomography images using MobileNetV2 and transfer learning, Bioengineering, 10 (98)

[2] Mohammed M. A., Abdullah R. L., Karrar H. A. (2023). Federated auto-encoder and XGBoost schemes for multi-omics cancer detection in distributed fog computing paradigm, *Chemometrics* and *Intelligent Laboratory Systems*, 241(15), 66-73.

[3] Ismail MBS (2021). Lung cancer detection and classification using machine learning algorithm, *Turkish Journal* of *Computer* and *Mathematics Education*, 12(13), 7048–7054.

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2025

DOI: 10.17148/IJARCCE.2025.14512

[4] Chandran U., Reps J., Yang R., Vachani A (2023). Machine learning and real-world data to predict lung cancer risk in routine care, *Cancer Epidemiology, Biomarkers & Prevention*, 32(3), 337–343.

[5] Mohammad Q. Shatnawi, Qusai Abuein, Romesaa Al-Quraan (2025). Deep learning-based approach to diagnose lung cancer using CT-scan images, Intelligence-Based Machine, 11, 100188.

[6] Sandeep Kumar, Jagendra Singh, Vinayakumar Ravi, Prabhishek Singh , Alanoud Al Mazroa, Manoj Diwakar and Indrajeet Gupta (2024). Deep Learning and MRI Biomarkers for Precise Lung Cancer Cell Detection and Diagnosis, The Open Bioinformatics Journal, 17, e18750362335415

[7] Satya Prakash Maurya, Pushpendra Singh Sisodia, Rahul Mishra, Devesh Pratap singh (2024). Performance of machine learning algorithms for lung cancer prediction: a comparative approach, Scientific Reports , 14, 18562.

[8] Yusupha Sinjanka, Veerpal Kaur, Usman Ibrahim Musa, Karandeep Kaur (2024), ML-based early detection of lung cancer: an integrated and in-depth analytical framework, 4(92).

