Lung cancer detection using hybrid integration of autoencoder feature extraction and ML techniques

Annemneedi Lakshmanarao[1], Nirmal Gopal[2], Nagagopiraju Vullam[3], Mandapati Sridhar[4], Modalavalasa Krishna Kanth[5], Uma Maheswari Rayudu[6]
[1]Department of IT, Aditya University, Surampalem, India
[2]Data Scientist, AI and Data Science Department, Chennai, India
[3]Department of CSE, Chalapathi Institute of Engineering and Technology, Guntur, India
[4]Department of Computer Applications, R.V.R. and J.C. College of Engineering, Guntur, India
[5]Department of CSE-Cyber Security, Malla Reddy Engineering College, Hyderabad, India
[6]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Guntur, India

Article Info

Article history:

Received Mar 10, 2024
Revised Sep 4, 2024
Accepted Sep 29, 2024

Keywords:

Autoencoder
Feature extraction
Kaggle
Lung cancer
ML techniques

ABSTRACT

Lung cancer posed a significant global health challenge, necessitating innovative approaches for early detection and accurate diagnosis. In this paper, CT scan images for lung cancer with three classes namely benign, malignant, and normal are collected from Kaggle. We initially applied conventional machine learning (ML) algorithms including support vector machine (SVM), random forests (RF), decision trees (DT), logistic regression (LR), naive bayes (NB), and k-nearest neighbor for lung cancer detection. The results with these conventional algorithms are recorded. Later, we proposed a novel hybrid model that integrated diverse machine learning algorithms to further enhance accuracy. Our approach combined the power of autoencoders for feature extraction. Using Autoencoder technique, features from images are extracted and a new feature vector is created. Later, the same conventional ML classifiers applied and achieved enhanced performance. The hybrid model demonstrated remarkable performance in identifying lung cancer cases when compared to individual classifiers. Through extensive experimentation, we showcased the efficacy of our integrated framework, achieving high accuracy, precision, recall and F1-score metrics across multiple classifiers. This hybrid approach represented a significant advancement in lung cancer detection, offering a versatile and robust solution for early diagnosis and personalized treatment strategies in clinical settings.

Corresponding Author:

Annemneedi Lakshmanarao
Department of IT, Aditya University
Surampalem, India
Email: laxman1216@gmail.com

1. INTRODUCTION

Lung cancer places a heavy strain on communities and public health systems. Images from CT scans have become an important resource for identifying and categorizing lung cancer because they provide comprehensive information on pulmonary abnormalities. Due to their precise views of the lungs and associated tissues, CT scans are useful for lung cancer screening. High-resolution CT scans may detect lung cancer nodules and masses. To acquire tissue samples for pathology, CT images help guide biopsy needles. This helps confirm lung cancer diagnosis and identify its features for therapy planning. In this paper, lung cancer CT scan images are collected from Kaggle, categorized them into benign, malignant, and normal classes. Initially, we used traditional machine learning algorithms for detection, then introduced a hybrid model combining various ML algorithms and autoencoders for feature extraction. This approach yielded superior performance in identifying lung cancer cases compared to individual classifiers, showcasing high accuracy and offering a significant leap forward in early diagnosis and treatment strategies.

Raja *et al.* [1] presented a lung CT scan-based lung cancer detection system. Lung cancer detection module used Transfer Learning with AdaDenseNet and Prox-SMOTE to boost accuracy. VGG-16 extracted features and an support vector machine (SVM) classifier classified them in the cardio-vascular diseases (CVD) prediction module. We found that lung cancer and CVD are interdependent, with lung cancer detection accuracy of 98.28% and CVD prediction accuracy of 91%. Sait [2] created a convolutional neural network (CNN) model with DenseNet-121 feature extraction and deep autoencoders for dimensionality reduction. MobileNet V3-Small recognized lung cancer kinds from these characteristics. Quantization-aware training and early halting reduced processing power and enhanced accuracy. Alqahtani *et al.* [3] introduced the improved water strider algorithm with autoencoder. It aimed to detect lung cancer, utilizing median filtering (MF) for noise removal and MobileNetv2 as a feature extractor with IWSA-based hyperparameter optimization. The technique, employing a convolutional autoencoder (CAE), demonstrated superior performance through simulations compared to alternative approaches. Gao *et al.* [4] explored various AI techniques for lung cancer detection. The authors claimed that aI techniques overcomes traditional approaches for lung cancer detection.

Jawaid *et al.* [5] used image processing, segmentation, and cancer staging to identify lung cancer in CT images. It simplified diagnostics with 96.5% accuracy and early detection using Android devices and an artificial neural network (ANN). Rao and Arshad *et al.* [6] suggested a neural network for image collecting, preprocessing, pixel enhancement, segmentation, and feature extraction. Machine learning is essential for accurate and cost-effective lung cancer diagnosis and treatment, according to current studies. Gupta and Dawn [7] proposed a CNN-based lung CT scan cancer prediction and classification method. Augmentation corrected class imbalance after picture preprocessing and resizing. Mahum and Salman *et al.* [8] suggested Lung-RetinaNet, an effective RetinaNet-based lung tumor detector. To enhance tumour localization, it used a multi-scale feature fusion module and a dilated and lightweight context approach. The technique outperformed others models. Bharathy *et al.* [9] trained the dataset using support vector classification (SVC), k-nearest neighbor (KNN), decision tree, logistic regression (LR), and random forest (RF). The most accurate algorithm was RF, with 88.5% performance. Kumari *et al.* [10] recommended precedence-based machine learning methods for cervical and breast cancer prediction. After training the fundamental models, ranking-based methods predicted accuracy. Experimental findings on two datasets showed that the suggested framework detected cancer accurately.

Computer vision and deep learning (DL), especially second- and third-neural network convergence, were used in [11]. Each site had to eliminate gabor filter-induced white gaussian scan image noise and phase the respiratory organ using twin tree complex moving ridge rework dual-tree complex wavelet transform and discrete cosine transform (DTCWT) technology. SVM found several quality characteristics in this article. Lakshmanarao *et al.* [12] suggested an ML-based fusion classifier model. Following base-level model training, ranking-based algorithms predicted accuracy. The suggested approach was evaluated on UCI and Kaggle cancer datasets. The suggested approach worked on these datasets. ML was used to create effective lung cancer risk assessment models in [13]. The rotation forest technique performed well in precision, recall, f-measure, accuracy, and AUC. The model performed well in experiments, with an AUC of 99.3% and accuracy of 97.1%. Pandian *et al.* [14] presented a technique to identify aberrant lung tissue development but stressed the requirement for high accuracy to reduce misdiagnosis. Analyzing healthy and malignant lung pictures, databases for different CT scanning system perspectives were created. A neural network classified normal and cancerous pictures using image textural properties. Lalitha [15] introduced a ML-based automated lung cancer detection system that can distinguish benign, aggressive, and normal lung cancers. The suggested lung cancer detection technique outperformed the others with over 95% accuracy. The study aimed to detect early-stage lung cancer using machine learning methods. It sought to assess more cases efficiently, achieving results comparable to or faster than human experts. Nine different algorithms were utilized in the developed model (naive bayes (NB), LR, decision trees (DT), RF, gradient boost (GB), and SVM), and their success was evaluated based on accuracy, sensitivity, and precision metrics. The results demonstrated a maximum accuracy of 91% in cancer detection.

Dirik [16] aimed to develop an automated model for early-stage lung cancer detection using nine machine learning algorithms (NB, LR, DT, RF, GB, and SVM). The model's success was evaluated with accuracy, sensitivity, and precision metrics from the confusion matrix. The results showed that the proposed model detected cancer with a maximum accuracy of 91%. Nazir *et al.* [17], DL and ML accelerated cancer diagnosis and stage categorization. This work validated with PCA and discrete wavelet transform for image fusion using multiresolution rigid registration and image segmentation. Mishra and Gangwar [18], SVM, LR, ANN, and NB were utilized to study and prognose lung cancer. This article used survey data and machine learning algorithms like SVM, NB, KNN etc. to predict lung cancer. Abdullah *et al.* [19] evaluated three classifiers SVM, KNN, and ANN by detecting early-stage lung cancer. Using UCI datasets, the study examined classification algorithms' accuracy using WEKA. SVM has the highest accuracy (95.56%), followed by CNN (92%) and KNN (88%). hybrid methods were used to diagnose lung and breast cancer based on cell size and shape [20]. They introduced breast and lung cancer block diagrams and discussed cancer detection and diagnostic issues and prospective uses.

Mohamed *et al.* [21] presented a CNN-metaheuristic hybrid. A CNN architecture was constructed and the model's solution vector calculated. This solution vector was sent to ebola optimization search algorithm (EOSA) model to choose the optimal weights and bias for CNN model classification training. Several models were tested utilizing chest X-rays or CT scans to diagnose a disease [22]. The study sought the best DL lung disease prediction methods. The approach was assessed using various metrices. Histopathological lung cancer tissue pictures were utilized to train deep neural networks to identify lung cancer in [23]. Inception V3, random forest, and CNN were tested for automated lung cancer cell identification. Training the CNN to extract key characteristics improved lung cancer detection efficiency and accuracy. Nadkarni and Borkar [24] demonstrated automatic lung cancer detection with CT scans. Median filtering for image preprocessing and mathematical morphological segmentation of the lung region of interest were used to identify lung cancer. Khosravan and Bagci [25] proposed a 3D deep multi-task CNN achieved 91% segmentation accuracy and 92% FP reduction on the LUNA16 dataset.

2. METHOD

The proposed architecture for cancer detection is shown in Figure 1. A hybrid lung cancer detection model was created using traditional ML techniques and autoencoders for feature extraction. The dataset includes Kaggle CT scan pictures classed as benign, malignant, and normal, depicting lung cancer stages.

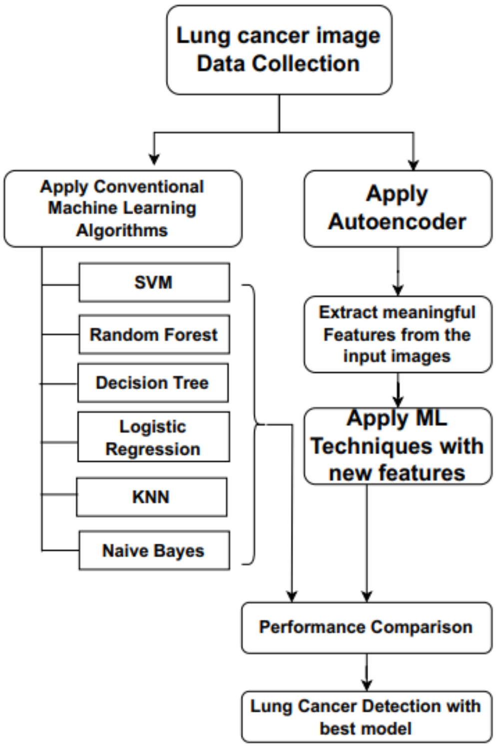

Figure 1. Proposed architecture for lung cancer detection

Initially, a range of conventional ML algorithms, including SVM, RF, DT, LR, NB, and KNN classifiers, was applied to the dataset. The results obtained from these algorithms served as a baseline for comparison and were meticulously recorded for subsequent analysis.

Following the evaluation of the conventional algorithms, autoencoders were introduced for feature extraction from the CT scan images. Autoencoders were employed to capture latent representations of the input data, generating a new feature vector that encapsulated the salient characteristics of the lung cancer images. Subsequently, the extracted features from the autoencoders were integrated with the aforementioned conventional machine learning classifiers. This hybrid model architecture was designed to harness the strengths of both feature extraction through autoencoders and the predictive capabilities of diverse machine learning algorithms.

The hybrid model was then trained using the extracted features obtained from the autoencoders. Following training, the performance of the hybrid model was evaluated using appropriate evaluation metrics. Through extensive experimentation and validation on independent test sets, the efficacy of the integrated framework was demonstrated. The performance metrics of the hybrid model were compared with those obtained from individual classifiers, showcasing its superiority in achieving high accuracy and robustness in lung cancer detection. This hybrid approach represented a significant advancement in early diagnosis and personalized treatment strategies for lung cancer, offering a versatile and effective solution for clinical settings.

2.1. Dataset

A Kaggle dataset [26] was collected for experimentation. It contains three classes of images namely benign, malignant, and normal with 120,561 and 416 images Here, “benign” refers to non-cancerous abnormalities or growths, “malignant” indicates cancerous lesions or tumors, and “normal” represents images showing healthy lung tissue without any abnormalities. Later, this dataset is divided into training and testing sets with 85% and 15% ratio.

2.2. ML classification algorithms

In this work, six ML algorithms were utilized to analyze lung cancer datasets and aid in accurate identification and classification of lung cancer cases. These algorithms include SVM, RF, DT, LR, NB, and KNN. Each algorithm was employed to analyze features extracted from the datasets, such as CT scan images, and to recognize patterns and relationships within the data, contributing to the overall objective of improving cancer diagnosis.

SVM finds the appropriate hyperplane to divide feature classes. random forest trains several DT and produces class mode for categorization. Decision Tree classifies by recursively partitioning feature space by value. LR uses predictor variables to model binary outcomes. NB classifies features using bayes’ theorem and strong independence assumptions. KNN classifies an unseen instance based on the majority class of its closest neighbors in feature space.

2.3. Feature extraction with autoencoder

In this work, an autoencoder neural network is defined for feature extraction. The autoencoder is trained to reconstruct input images while minimizing reconstruction error. Hidden layers of the autoencoder extract salient features from the images, creating a compressed representation. These features are then concatenated to form a new feature vector for each image. The feature vectors are integrated with machine learning classifiers like SVM, random forest, and LR for lung cancer detection. This process enhances classification accuracy by capturing essential characteristics of the images, improving overall performance in identifying lung cancer cases from medical imaging data.

2.4. Hybrid integration with machine learning models

In this step, feature vectors extracted by the autoencoder are integrated with ML models, creating a hybrid framework for lung cancer detection. These vectors encapsulate essential CT scan image characteristics, enhancing ML algorithms’ performance. The hybrid approach combines autoencoder feature extraction with ML classifiers. Each model efficiently utilizes these vectors to classify images into benign, malignant, or normal types. This integration yields superior precision, recall, and accuracy, bolstering early lung cancer detection and diagnosis.

3. RESULTS AND DISCUSSION

3.1. Applying ML classification algorithms with original dataset

In the initial phase, we aimed at evaluate the performance of various ML classifiers in lung cancer detection using five classifiers namely SVM, RF, DT, LR, NB, and KNN. As a preprocessing step, the images are resized to a consistent dimension of 128x128 pixels. This resizing ensures uniformity in the input data. Each pixel in the images, represented by numeric values corresponding to its intensity, is flattened into a one-dimensional array to convert the images into numeric data, known as feature vectors. The flattening process enables the grayscale images to be represented as numeric arrays suitable for processing by ML algorithms. This conversion is essential for subsequent analysis tasks and enables efficient processing by the

ML classifiers. Grayscale images are preferred for their computational efficiency and relevance to lung cancer detection tasks. The dataset is then split into training and testing sets. The results with all these classifiers are tabulated. Table 1 and Figure 2 shows results with ML models.

Table 1. Output with actual image dataset and ML models

Model	Precision	Recall	F1	Accuracy
SVM	0.96	0.97	0.96	0.97
RF	0.96	0.88	0.91	0.95
DT	0.85	0.89	0.87	0.9
LR	0.97	0.95	0.92	0.95
NB	0.63	0.70	0.64	0.68
KNN	0.90	0.91	0.88	0.89

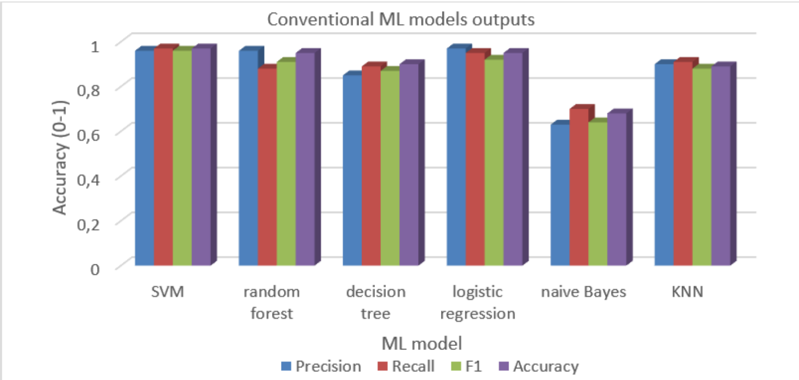

Figure 2. Output with ML classifiers

In evaluating machine learning models for lung cancer detection, several key performance metrics were assessed. SVM and LR demonstrated the highest precision (SVM: 0.96, LR: 0.97), recall (SVM: 0.97, LR: 0.95), and accuracy (SVM: 0.97, LR: 0.95), making them promising choices for accurate lung cancer identification. random forest and decision tree models showed slightly lower performance, with higher rates of false positives (random forest precision: 0.96, recall: 0.88, F1-score: 0.91, accuracy: 0.95; decision tree precision: 0.8, recall: 0.8, F1-score: 0.87, accuracy: 0.9). NB exhibited the lowest overall performance (precision: 0.63, recall: 0.70, F1-score: 0.64, accuracy: 0.68), while KNN showed balanced precision and recall (precision: 0.90, recall: 0.91, F1-score: 0.88, accuracy: 0.89).

3.2. Apply Autoencoder for feature extraction

An autoencoder architecture is defined for feature extraction from CT scan images. The architecture consists of an encoder and a decoder component. In the encoder section, convolutional layers with ReLU activation functions are applied to capture essential features of the input images. Max pooling layers are utilized to down sample the feature maps, reducing dimensionality while retaining relevant information. Encoder comprises of convolutional layers followed by max-pooling layers. It starts with a convolutional layer with 16 filters and a kernel size of (3, 3), applying ReLU activation functions to capture low-level features from the input CT scan images. Subsequently, a max-pooling layer down samples the feature maps by a factor of 2, reducing their spatial dimensions while preserving essential information. Another convolutional layer with 8 filters and the same kernel size is applied, further extracting higher-level features, followed by another max-pooling layer for down sampling. This sequence of convolutional and max-pooling layers forms the encoder section, progressively reducing the spatial dimensions of the input images and extracting increasingly abstract features, ultimately leading to a compressed representation of the input data at the bottleneck layer.

The decoder architecture mirrors the encoder's structure but in reverse order, aimed at reconstructing the input images obtained at the bottleneck layer. It begins with a convolutional layer that employs 8 filters and the same kernel size (3, 3) as the encoder's last convolutional layer. This layer initiates the reconstruction process by extracting features from the compressed representation. Subsequently, an up-sampling layer increases the spatial dimensions of the feature maps to match those of the original input images. Another convolutional layer with 16 filters and the same kernel size continues the reconstruction process by further refining the features extracted from the bottleneck layer. Another up-sampling layer is then applied to restore the spatial dimensions further. Finally, the output layer consists of a convolutional layer with 3 filters and a SoftMax activation function. This layer reconstructs the input images by mapping the extracted features back to pixel intensities, completing the decoding process. Overall, the decoder architecture aims to reverse the encoding process, gradually reconstructing the original input images from the compressed representation obtained at the bottleneck layer.

The autoencoder is compiled with the Adam optimizer. Subsequently, the model is trained on the provided images using a specified number of epochs and steps per epoch, with validation performed on the same dataset generator. The training done with 25 epochs. Figure 3 shows the epoch wise loss in training and validation phases. From Figure 3, it is observed that the loss value is decreasing in both training and testing phases, indicating that it can able to extract useful features.

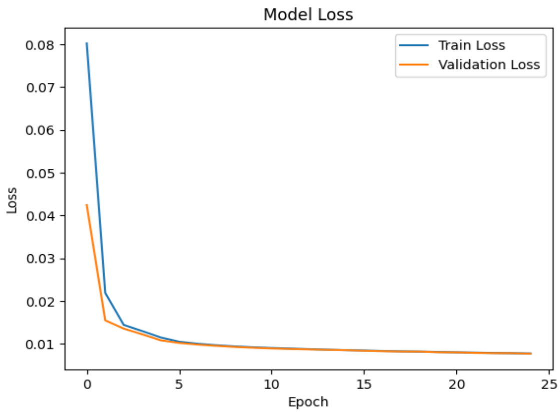

Figure 3. Epoch wise training and validation loss with autoencoder

3.3. Image reconstruction with trained autoencoder

The reconstructed images are generated using the trained autoencoder model. Figure 4 shows the original and reconstructed images of first 8 samples. This visualization enables a qualitative assessment of the performance of the autoencoder in reconstructing the input images.

By comparing the original and reconstructed images, we can gauge how accurately the autoencoder captures the essential features of the input data during the training process. The reconstructed images may appear slightly blurred. This is a common characteristic of autoencoder-based reconstructions, especially when the model is trained with a limited amount of data with specific constraints.

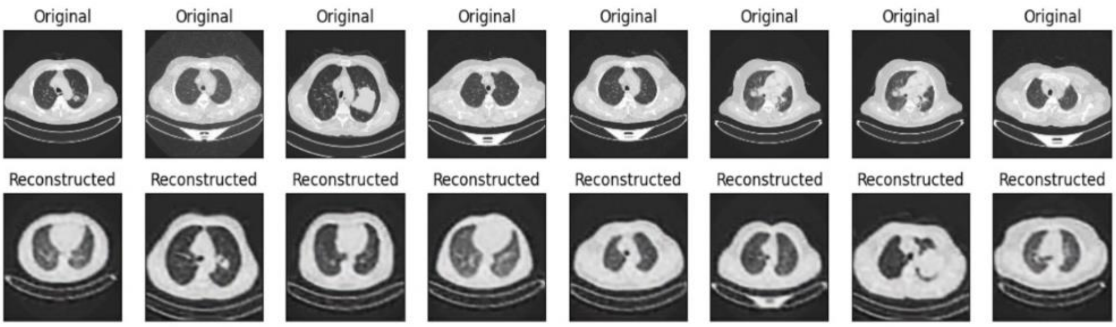

Figure 4. Reconstructed images generated by the trained autoencoder model

3.4. Creating new feature vector with output of autoencoder

To further leverage the trained autoencoder for feature extraction, a separate encoder model is constructed. This model is designed to extract features from the encoder section of the autoencoder architecture. The feature extraction process from the encoder section of the autoencoder yields 8100 features for each input image. These features represent a condensed representation of the original input images, capturing their essential characteristics in a lower-dimensional feature space. These extracted features are then utilized in subsequent steps, such as integration with ML classifiers, for lung cancer detection and classification.

3.5. Applying ML classification algorithms with new feature dataset

In this phase, ML classifiers applied with new feature vector generated by autoencoder. The algorithms namely SVM, RF, DTC, LR, NB and KNN are applied. The performance of all models increases in terms of all matrices. Table 2 and Figure 5 shows the results of ML models applied with autoencoder extracted feature vector.

Table 2. Output with new feature vector and ML models

Model	Precision	Recall	F1	Accuracy
SVM	0.96	0.97	0.96	0.97
RF	0.96	0.88	0.91	0.95
DT	0.85	0.89	0.87	0.9
LR	0.97	0.95	0.92	0.95
NB	0.63	0.70	0.64	0.68
KNN	0.90	0.91	0.88	0.89

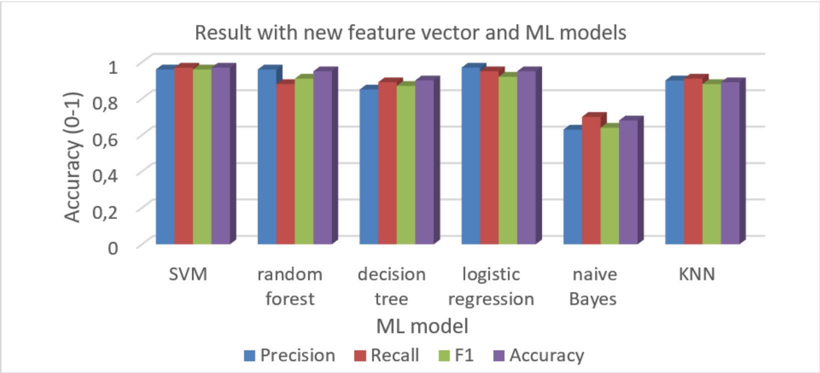

Figure 5. Output with ML models applied with autoencoder extracted dataset

4. CONCLUSION

This paper presented the critical issue of lung cancer detection, recognizing its global significance and the urgent need for innovative diagnostic approaches. Utilizing CT scan images categorized into benign, malignant, and normal classes, the research initially employed conventional ML algorithms, including SVM, random forests, DT, LR, NB, and KNN, to detect lung cancer. Subsequently, a novel hybrid model was introduced, integrating diverse machine learning algorithms with autoencoder-based feature extraction. The autoencoder technique was employed to extract essential features from the images, creating a new feature vector. The same conventional ML classifiers were then applied to this new dataset, resulting in significantly enhanced performance. Notably, the hybrid model achieved remarkable results, with all metrics exceeding 0.90 across multiple classifiers. SVM, LR, and KNN classifiers demonstrated precision and recall values of 0.97 and 0.95, respectively, highlighting the effectiveness of the integrated framework in accurately detecting lung cancer cases. These findings underscore a substantial advancement in lung cancer detection, showcasing the efficacy of the integrated framework in achieving high accuracy and enabling early diagnosis.

REFERENCES

[1] G. Raja *et al.*, "Synergistic analysis of lung cancer's impact on cardiovascular disease using ML-based techniques," *IEEE Journal of Biomedical and Health Informatics*, pp. 1–8, 2024, doi: 10.1109/JBHI.2024.3365176.
[2] A. R. W. Sait, "Lung cancer detection model using deep learning technique," *Applied Sciences (Switzerland)*, vol. 13, no. 22, p. 12510, Nov. 2023, doi: 10.3390/app132212510.
[3] H. Alqahtani, E. Alabdulkreem, F. A. Alotaibi, M. M. Alnfiai, C. Singla, and A. S. Salama, "Improved water strider algorithm with convolutional autoencoder for lung and colon cancer detection on histopathological images," *IEEE Access*, vol. 12, pp. 949–956, 2024, doi: 10.1109/ACCESS.2023.3346894.

[4] Q. Gao, L. Yang, M. Lu, R. Jin, H. Ye, and T. Ma, "The artificial intelligence and machine learning in lung cancer immunotherapy," *Journal of Hematology and Oncology*, vol. 16, no. 1, p. 55, May 2023, doi: 10.1186/s13045-023-01456-y.

[5] A. Jawaid, J. Hafeez, S. Khan, and A. Ur Rehman, "Lung cancer detection using artificial neural network on android," in *2023 Global Conference on Wireless and Optical Technologies (GCWOT)*, IEEE, Jan. 2023, pp. 1–8. doi: 10.1109/GCWOT57803.2023.10064658.

[6] B. D. Rao and M. Arshad, "Early detection of lung cancer using machine learning technique," in *2023 International Conference on Computer Communication and Informatics (ICCCI)*, IEEE, Jan. 2023, pp. 1–5. doi: 10.1109/ICCCI56745.2023.10128389.

[7] D. Gupta and S. Dawn, "Detection and staging of lung cancer from CT scan images by deep learning," in *2023 International Conference on Disruptive Technologies, ICDT 2023*, IEEE, May 2023, pp. 274–278. doi: 10.1109/ICDT57929.2023.10151194.

[8] R. Mahum and A. S. Al-Salman, "Lung-RetinaNet: lung cancer detection using a RetinaNet with multi-scale feature fusion and context module," *IEEE Access*, vol. 11, pp. 53850–53861, 2023, doi: 10.1109/ACCESS.2023.3281259.

[9] S. Bharathy, R. Pavithra, and B. Akshaya, "Lung cancer detection using machine learning," in *Proceedings - International Conference on Applied Artificial Intelligence and Computing, ICAAIC 2022*, IEEE, May 2022, pp. 539–543. doi: 10.1109/ICAAIC53929.2022.9793061.

[10] G. Kumari, U. V. Ramesh, P. P. Ramamohan, K. S. Pavani, and A. Lakshmanarao, "Cancer detection with ensemble learning model from novel precedence based algorithms," in *Proceedings of International Conference on Contemporary Computing and Informatics, IC3I 2023*, IEEE, Sep. 2023, pp. 2187–2190. doi: 10.1109/IC3I59117.2023.10397629.

[11] S. K. Bhatt and S. Srinivasan, "Lung cancer detection using AI and different techniques of machine learning," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 8, pp. 630–638, 2024.

[12] A. Lakshmanarao, A. Srisaila, and T. S. R. Kiran, "Cancer prediction using novel ranking algorithms and machine learning," in *Communications in Computer and Information Science*, vol. 1393, 2021, pp. 62–70. doi: 10.1007/978-981-16-3660-8_6.

[13] E. Dritsas and M. Trigka, "Lung cancer risk prediction with machine learning models," *Big Data and Cognitive Computing*, vol. 6, no. 4, p. 139, Nov. 2022, doi: 10.3390/bdcc6040139.

[14] R. Pandian, V. Vedanarayanan, D. N. S. R. Kumar, and R. Rajakumar, "Detection and classification of lung cancer using CNN and Google net," *Measurement: Sensors*, vol. 24, p. 100588, Dec. 2022, doi: 10.1016/j.measen.2022.100588.

[15] S. Lalitha, "An automated lung cancer detection system based on machine learning algorithm," *Journal of Intelligent & Fuzzy Systems*, vol. 40, no. 4, pp. 6355–6364, Apr. 2021, doi: 10.3233/JIFS-189476.

[16] M. Dirik, "Machine learning-based lung cancer diagnosis," *Turkish Journal of Engineering*, vol. 7, no. 4, pp. 322–330, Oct. 2023, doi: 10.31127/tuje.1180931.

[17] I. Nazir, I. U. Haq, S. A. Alqahtani, M. M. Jadoon, and M. Dahshan, "Machine learning-based lung cancer detection using multiview image registration and fusion," *Journal of Sensors*, vol. 2023, no. 1, Jan. 2023, doi: 10.1155/2023/6683438.

[18] A. Mishra and S. Gangwar, "Lung cancer detection and classification using machine learning algorithms," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 11, no. 6 S, pp. 277–282, Jun. 2023, doi: 10.17762/ijritcc.v11i6s.6920.

[19] D. M. Abdullah, A. M. Abdulazeez, and A. B. Sallow, "Lung cancer prediction and classification based on correlation selection method using machine learning techniques," *Qubahan Academic Journal*, vol. 1, no. 2, pp. 141–149, May 2021, doi: 10.48161/qaj.v1n2a58.

[20] D. N. Rao, "Machine learning research on breast and lung cancer detection," *International Journal of Innovation in Engineering*, vol. 3, no. 1, pp. 48–54, Mar. 2023, doi: 10.59615/ijie.3.1.48.

[21] T. I. A. Mohamed, O. N. Oyelade, and A. E. Ezugwu, "Automatic detection and classification of lung cancer CT scans based on deep learning and ebola optimization search algorithm," *PLoS ONE*, vol. 18, no. 8 AUGUST, p. e0285796, Aug. 2023, doi: 10.1371/journal.pone.0285796.

[22] B. S. Deepapriya *et al.*, "Performance evaluation of deep learning techniques for lung cancer prediction," *Soft Computing*, vol. 27, no. 13, pp. 9191–9198, Jul. 2023, doi: 10.1007/s00500-023-08313-7.

[23] A. A. Abd Al-Ameer, G. A. Hussien, and H. A. Al Ameri, "Lung cancer detection using image processing and deep learning," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 28, no. 2, pp. 987–993, Nov. 2022, doi: 10.11591/ijeecs.v28.i2.pp987-993.

[24] N. S. Nadkarni and S. Borkar, "Detection of lung cancer in CT images using image processing," in *Proceedings of the International Conference on Trends in Electronics and Informatics, ICOEI 2019*, IEEE, Apr. 2019, pp. 863–866. doi: 10.1109/icoei.2019.8862577.

[25] N. Khosravan and U. Bagci, "Semi-supervised multi-task learning for lung cancer diagnosis," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, IEEE, Jul. 2018, pp. 710–713. doi: 10.1109/EMBC.2018.8512294.

[26] Hamdalla F. Al-Yasriy, "The IQ-OTH/NCCD lung cancer dataset." [Online]. Available: https://www.kaggle.com/datasets/hamdallak/the-iqothnccd-lung-cancer-dataset

BIOGRAPHIES OF AUTHORS

Dr. Annemneedi Lakshmanarao currently working as associate professor in Aditya University, Surampalem. He completed his B.Tech. in CSIT and M. Tech in Software Engineering. He Completed Ph.D. in Andhra University, Vishakhapatnam. His areas of interest are machine learning, cyber security, deep learning. He is a life member of Computer Society of India (CSI) and ISTE. He can be contacted at email: laxman1216@gmail.com.

Nirmal Gopal who graduated with a bachelor's degree in Computer Science and Engineering from Chennai Institute of Technology. He has contributed two conference papers presented at International Conferences and has a profound interest in the dynamic realms in artificial intelligence, machine learning and data science. He is currently working as an AWS developer. He can be contacted at email: gnirmal2001@gmail.com.

Dr. Nagagopiraju Vullam currently working as professor in Department of CSE, Chalapathi Institute of Engineering and Technology, Guntur, Andhra Pradesh, Guntur Dist. He published papers in various conferences and journals. His areas of interest are machine learning, data analytics through R, artificial intelligence. He can be contacted at email: gopi.raju524@gmail.com.

Dr. Mandapati Sridhar Dr. Sridhar Mandapati, obtained his master"s degree in Computers Applications from S.V University, Tirupathi. He received his Ph.D. in Computer Science and Engineering at Acharya Nagarjuna University, Guntur. He is currently working as Professor in the Department of Computer Applications, R.V.R. and J.C College of Engineering, Guntur. He has 24 years of teaching experience. He has more than 50 International and National publications to his credit. His research interests include data mining and big data analytics, machine learning and cloud computing. He can be contacted at email: mandapati12@gmail.com.

Modalavalasa Krishna Kanth currently working as assistant professor, Department of Computer Science and Engineering (Cyber Security), Malla Reddy Engineering college,Medchal, Telangana. He obtained his M.Tech. in Computer Science and Engineering from JNTU Kakinada, and has good publications in his research area. He can be contacted at email: mkkanth3887@mrec.ac.in.

Uma Maheswari Rayudu is currently working an assistant professor at KL University, Vaddeswaram. She completed her M.Sc. in Application Oriented Mathematics and completed M.Tech in Computer science and Engineering. She has 13 years of experience in teaching. She has extensive experience in teaching subjects like engineering mathematics, probability and statistics, data warehousing, mining, machine learning, business analytics, and big data. She can be contacted at email: mumamaheswari@kluniversity.in.