International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

Customized 3D CNN Model-based Lung Cancer Classification from Chest X-ray Images

V. Navyasree[1], M. Meghana[2], N. Vaishnavi[2], N. Bhargavi[2], P. Mounika[2]

[1]Assistant Professor, [2]UG Student, [1,2]Department of Electronics and Communication Engineering
[1,2]Malla Reddy Engineering College for Women, Maisammaguda, Hyderabad, Telangana, India

ABSTRACT

Medical imaging tools are essential in early-stage lung cancer diagnostics and the monitoring of lung cancer during treatment. Various medical imaging modalities, such as chest X-ray, magnetic resonance imaging, positron emission tomography, computed tomography, and molecular imaging techniques, have been extensively studied for lung cancer detection. These techniques have some limitations, including not classifying cancer images automatically, which is unsuitable for patients with other pathologies. It is urgently necessary to develop a sensitive and accurate approach to the early diagnosis of lung cancer. Deep learning is one of the fastest-growing topics in medical imaging, with rapidly emerging applications spanning medical image-based and textural data modalities. With the help of deep learning-based medical imaging tools, clinicians can detect and classify lung nodules more accurately and quickly. Therefore, this work implements the advanced modifications in CNN model for the detection of lung cancer from chest scan images. The proposed CNN model is able to classify the benign and malignant i.e., normal, and cancerous with higher accuracy as compared to state-of-the-art machine learning approach called support vector machine (SVM) classifier. In addition, the obtained quality metrics discloses the superiority of proposed deep CNN model for assisting the expertise in an enhanced diagnosis.

Keywords: X-ray imaging, lung cancer, image classification, support vector machine, deep learning.

1. INTRODUCTION

Lung cancer is the primary cause of cancer death worldwide, with 2.09 million new cases and 1.76 million people dying from lung cancer in 2018 [1]. Four case-controlled studies from Japan reported in the early 2000s that the combined use of chest radiographs and sputum cytology in screening was effective for reducing lung cancer mortality. In contrast, two randomized controlled trials conducted from 1980 to 1990 concluded that screening with chest radiographs was not effective in reducing mortality in lung cancer [2, 3]. Although the efficacy of chest radiographs in lung cancer screening remains controversial, chest radiographs are more cost-effective, easier to access, and deliver lower radiation dose compared with low dose computed tomography (CT). A further disadvantage of chest CT is excessive false positive (FP) results. It has been reported that 96% of nodules detected by low-dose CT screening are FPs, which commonly leads to unnecessary follow-up and invasive examinations. Chest radiography is inferior to chest CT in terms of sensitivity but superior in terms of specificity. Taking these characteristics into consideration, the development of a computer-aided diagnosis (CAD) model for chest radiograph would have value by improving sensitivity while maintaining low FP results [4].

Many computer-aided detection (CAD) systems have been extensively studied for lung cancer detection and classification [5, 6]. Compared to trained radiologists, CAD systems provide better lung nodules and cancer detection performance in medical images. Generally, the CAD-based lung cancer detection system includes four steps: image processing, extraction of the region of interest (ROI), feature selection, and classification. Among these steps, feature selection and classification play the most critical roles in improving the accuracy and sensitivity of the CAD system, which relies on image processing to capture reliable features. However, benign, and malignant nodule classification is a challenge. Therefore, a rapid, cost-effective, and highly sensitive deep learning-based CAD system for lung cancer prediction is urgently needed.

The recent application of convolutional neural networks (CNN), a field of deep learning (DL), has led to dramatic, state-of-the-art improvements in radiology. DL-based models have also shown promise for nodule/mass detection on chest radiographs, which have reported sensitivities in the range of 0.51–0.84 and mean number of FP indications per image (mFPI) of 0.02–0.34. In addition, radiologist performance for detecting nodules was better with these CAD models than without them. In clinical practice, it is often challenging for radiologists to detect nodules and to differentiate between benign and malignant nodules. Normal anatomical structures often appear as if they are nodules, which is why radiologists must pay careful attention to the shape and marginal properties of nodules. As these problems are caused by the conditions rather than the ability of the radiologist, even skilful radiologists can misdiagnose. Therefore, the main purpose of this work was to train and validate a DL-based model capable of detecting lung cancer on chest radiographs, and to evaluate the characteristics of this DL-based model to improve sensitivity while maintaining low FP results.

2. LITERATURE SURVEY

The development of malignant cells in the lungs is known as lung cancer. Overall men and women's mortality rates have increased as a result of growing cancer incidence. Lung cancer is a disease wherein the cells in the lungs quickly multiply. Lung cancer cannot be eradicated, but it can be reduced [7]. The number of people affected with lung cancer is precisely equal to the number of people who smoke continuously. Lung cancer treatment was evaluated using classification approaches such as Naive Bayes, SVM, Decision Tree, and Logistic Regression. Pradhan et al. [8] conduct an empirical evaluation of multiple machine learning methods that can be used to identify lung cancer using IoT devices. A survey of roughly 65 papers employing machine learning techniques to forecast various diseases was conducted in this study. The study focuses on a variety of machine learning methods for detecting a variety of diseases in order to identify a gap in prospective lung cancer detection in medical IoT. Deep residual learning is used by Bhatia et al. [9] to identify lung cancer from CT scans. With the UNet and ResNet algorithms, we propose a series of pre-processing approaches for emphasising cancer-prone lung regions and retrieving characteristics. The extracted features are fed through several classifiers, namely Adaboost and Random Forest, and the individual predictions are ensembled to calculate the likelihood of a CT scan becoming cancerous.

Without learning inadequate human information, Shin et al. [10, 11] use deep learning to investigate the characteristics of cell exosomes and determine the similarities in human plasma extracellular vesicles. The deep learning classifier was tested with exosome SERS data from regular and lung cancer cell lines and was able to categorise them with 95% efficiency. The deep learning algorithm projected that 90.7% of patients' plasma exosomes were more similar to lung cancer cell extracellular vesicles than the mean of healthy controls in 43 patients, encompassing stage I and II cancer patients. In the ability to forecast lung ADC subtypes, researchers looked at four clinical factors: age, sex, tumour size, and smoking status, as well as 40 radiomic markers. The LIFEx software was used to extract radiomic characteristics from PET scans of segmented cancers. The clinical and radio mic variables were ranked, and a subset of meaningful features was chosen based on Gini coefficient scores for histopathological class relationships [12]. In the estimation of survival, a deep learning network with a tumour cell and metastatic staging system was used to examine the dependability of individual therapy suggestions

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

supplied by the deep learning preservation neural network. The C statistics were employed to evaluate the performance of the model. The computational intelligence survival neural network model's longevity forecasts and treatment strategies were made easier with the use of a customer interface [13].

A lung cancer detection model that utilizes image analysis and machine intelligence to identify the occurrence of lung cancer in CT scans and blood tests has been developed. Despite the fact that CT scan findings are more efficient than mammograms, patient CT scan pictures are divided into normal and abnormal categories [14, 15]. Even in the same tumour stage, non-small-cell cancer patients have a wide range of clinical performance and results. This research investigates deep learning applications such as medical imaging, which could help with patient stratification by automating the measurement of radiographic properties.

3. PROPOSED METHODOLOGY

A deep CNN model for lung cancer classification from CT scan images is a powerful approach that leverages the capabilities of deep learning to automatically learn and extract relevant features from raw image data. Here is an overview of how a deep CNN model can be used for classifying CT scan images into normal and malignant categories:

1. Data Collection and Preprocessing:

- A dataset of CT scan images is collected, where each image is labeled as normal (non-cancerous) or malignant (cancerous).
- Preprocessing steps are applied to the images, including resizing them to a consistent resolution, normalizing pixel values, and enhancing image quality.

2. Dataset Splitting: The dataset is divided into two subsets: training, and testing sets. This division allows for training the model, tuning hyperparameters, and evaluating its performance independently.

3. Architecture Design:

- The deep CNN architecture is designed to learn hierarchical features from the CT scan images. A typical architecture consists of multiple convolutional layers followed by pooling layers and fully connected layers.
- Convolutional layers use learnable filters to detect features such as edges, textures, and shapes at different spatial scales.
- Pooling layers reduce the spatial dimensions of the feature maps while retaining the most important information.
- Fully connected layers combine the extracted features and make the final classification decision.

4. Model Training:

- The CNN model is trained using the training dataset. During training, the model learns to optimize its internal parameters (weights and biases) to minimize a specified loss function (e.g., binary cross-entropy).
- Backpropagation and gradient descent techniques are used to update the model's parameters iteratively.

5. Hyperparameter Tuning: Hyperparameters, such as learning rate, batch size, and the number of layers and neurons, are tuned using the validation set to optimize the model's performance.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362
IJARST

7. Regularization Techniques: Regularization methods like dropout and batch normalization may be incorporated to prevent overfitting and enhance model robustness.

8. Evaluation:

- The model's performance is evaluated using the testing dataset. Common evaluation metrics include accuracy, precision, recall, F1-score, and the confusion matrix.
- The model's ability to correctly classify lung cancer CT scan images into normal and malignant categories is assessed.

The deep CNN models have demonstrated remarkable success in medical image classification tasks, including lung cancer classification. Their ability to automatically learn relevant features from raw image data makes them a valuable tool for improving diagnostic accuracy and efficiency in healthcare.

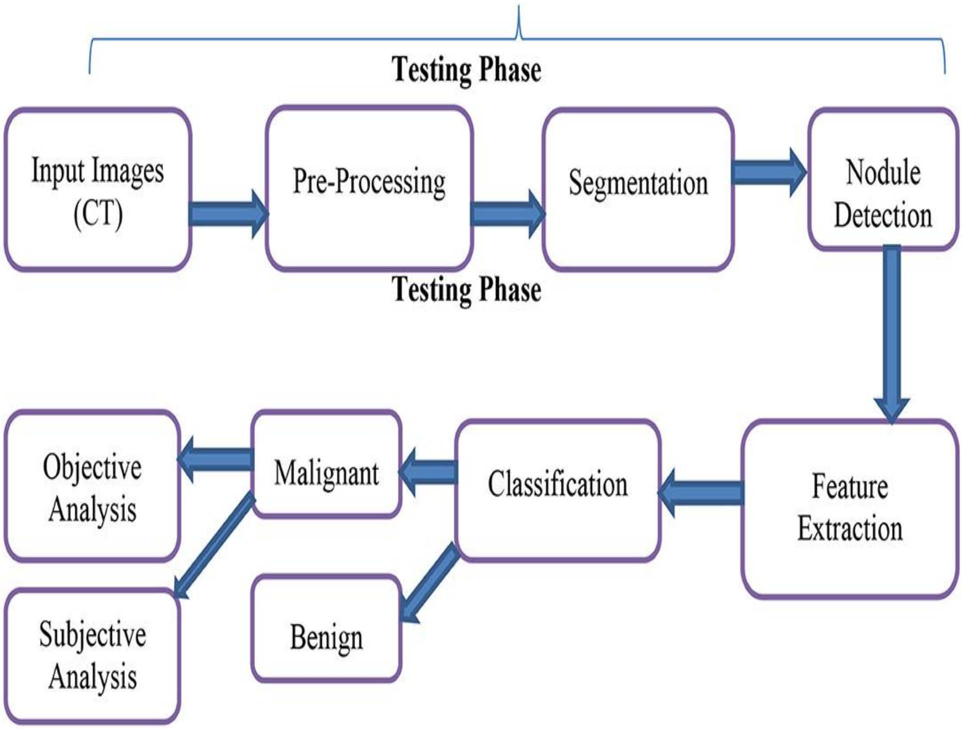

Figure 1: Overall design of proposed CNN model for lung cancer classification from CT scan images.

4. RESULTS AND DISCUSSION

This project uses a total of 138 CT scan images out of which 80% i.e., 110 images are used for training and 20% i.e., 28 images are used for testing the SVM, and CNN models for detecting the normal, and cancer from test CT scan images. It uses the Tkinter library for creating a graphical user interface (GUI) to perform lung cancer detection and classification tasks using Support Vector Machine (SVM) and Convolutional Neural Network (CNN) models.

International Journal For Advanced Research
In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362
IJARST

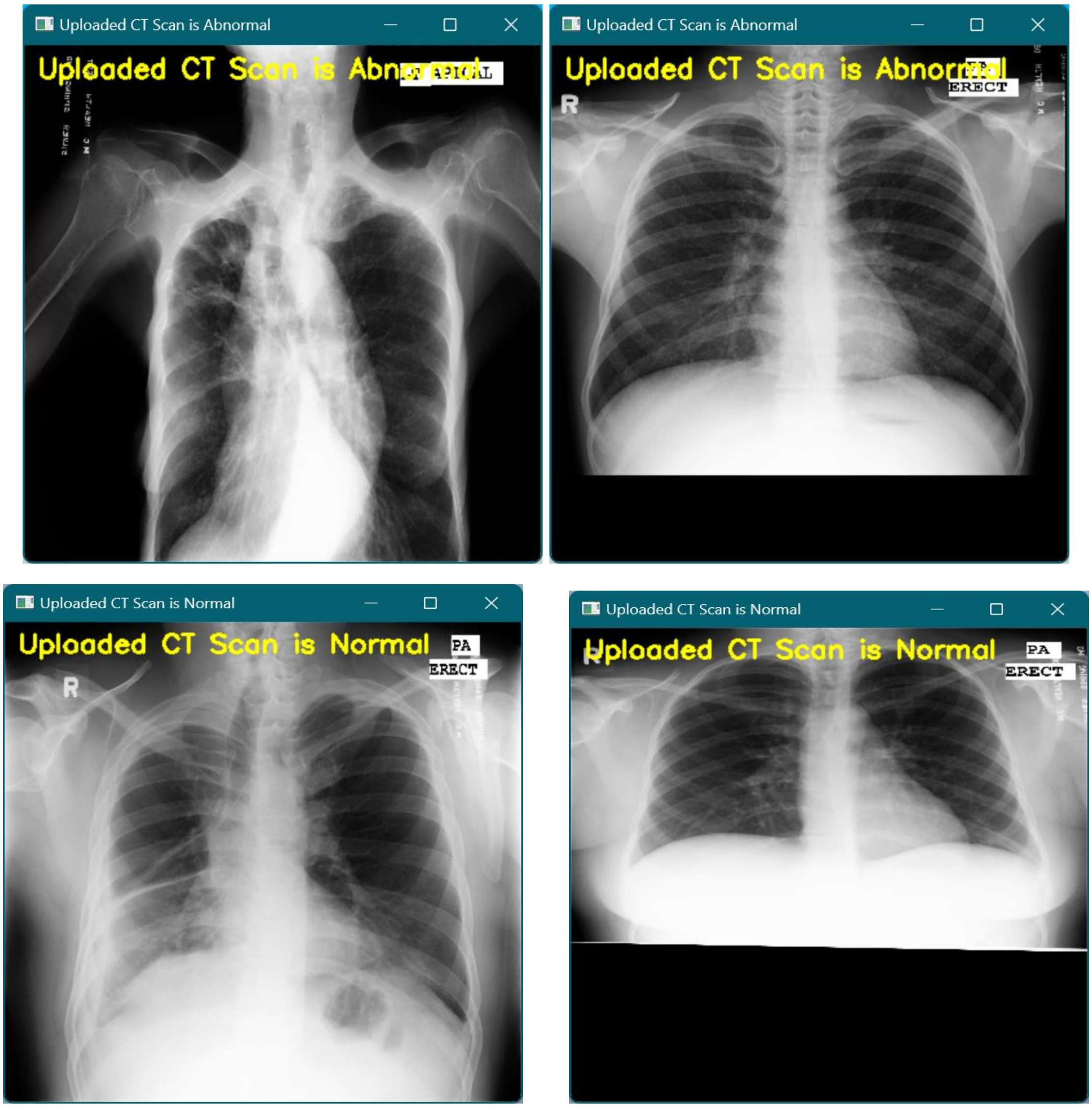

Figure 2: Sample prediction outcome obtained using proposed 3D CNN model.

Figure 2 presents an example of a prediction outcome generated by the 3D CNN model. It includes an image indicating whether a CT scan is normal or abnormal based on the model's prediction. Figure 3 shows a comparative analysis of the accuracy achieved by the SVM and 3D CNN models for lung cancer detection, where the proposed 3D CNN achieved enhanced accuracy i.e., 97.82% while the SVM obtained 78.57% of accuracy.

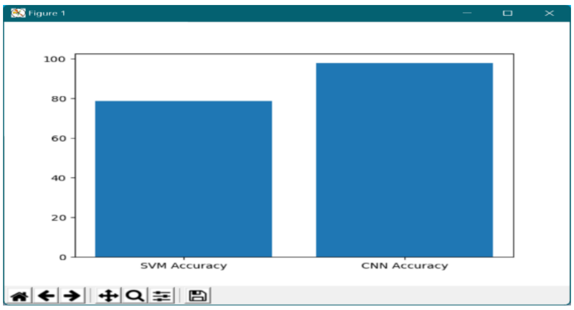

Figure 3: Performance comparison of accuracy for lung cancer detection system using SVM, and proposed 3D CNN model.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
IJARST
ISSN: 2457-0362

5. CONCLUSIONS

This work implements an advanced CNN modifications for the detection of lung cancer from CT scan images. The higher accuracy of the 3D CNN can be attributed to its ability to automatically learn relevant features from the images through multiple convolutional layers, which is especially advantageous for complex and high-dimensional data like CT scans. On the other hand, SVM relies on handcrafted features and may struggle with capturing complex patterns in the data. Based on the obtained results, it is clear that the proposed 3D CNN model significantly outperformed the SVM classifier in the task of detecting lung cancer from CT images. The 3D CNN achieved an accuracy of 97.82%, whereas the SVM achieved an accuracy of 78.57%. This indicates that the CNN model is much more effective at classifying CT scan images as either normal or abnormal, which is crucial for the early detection of lung cancer.

REFERENCES

[1] Y. Xu, A. Hosny, R. Zeleznik, C. Parmar, T. Coroller, I. Franco, H.J. Aerts, Deep learning predicts lung cancer treatment response from serial medical imaging, Clin. Cancer Res. 25 (11) (2019) 3266–3275.

[2] M.I. Faisal, S. Bashir, Z.S. Khan, F.H. Khan, An evaluation of machine learning classifiers and ensembles for early stage prediction of lung cancer, December, in: 2018 3rd International Conference on Emerging Trends in Engineering, Sciences and Technology (ICEEST), IEEE, 2018, pp. 1–4.

[3] N. Coudray, P.S. Ocampo, T. Sakellaropoulos, N. Narula, M. Snuderl, D. Feny¨o, A. Tsirigos, Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning, Nat. Med. 24 (10) (2018) 1559–1567.

[4] D.M. Ibrahim, N.M. Elshennawy, A.M. Sarhan, Deep-chest: multi-classification deep learning model for diagnosing COVID-19, pneumonia, and lung cancer chest diseases, Comput. Biol. Med. 132 (2021), 104348.

[5] Avanzo, J. Stancanello, G. Pirrone, G. Sartor, Radiomics and deep learning in lung cancer, Strahlenther. Onkol. 196 (10) (2020) 879–887.

[6] S.H. Hyun, M.S. Ahn, Y.W. Koh, S.J. Lee, A machine-learning approach using PET- based radiomics to predict the histological subtypes of lung cancer, Clin. Nucl. Med. 44 (12) (2019) 956–960.

[7] P. R. Radhika, R.A. Nair, G. Veena, A comparative study of lung cancer detection using machine learning algorithms, 2019, February, in: IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), IEEE, 2019, pp. 1–4.

[8] K. Pradhan, P. Chawla, Medical Internet of things using machine learning algorithms for lung cancer detection, J. Manag. Anal. 7 (4) (2020) 591–623.

[9] S. Bhatia, Y. Sinha, L. Goel, Lung cancer detection: a deep learning approach, in: Soft Computing for Problem Solving, Springer, Singapore, 2019, pp. 699–705.

[10] H. Shin, S. Oh, S. Hong, M. Kang, D. Kang, Y.G. Ji, Y. Choi, Early-stage lung cancer diagnosis by deep learning-based spectroscopic analysis of circulating exosomes, ACS Nano 14 (5) (2020) 5435–5444.

[11] V. Rajasekar, B. Predi´c, M. Saracevic, M. Elhoseny, D. Karabasevic, D. Stanujkic, P. Jayapaul, Enhanced multimodal biometric recognition approach for smart cities based on an optimized fuzzy genetic algorithm, Sci. Rep. 12 (1) (2022) 1–11.

[12] S.H. Hyun, M.S. Ahn, Y.W. Koh, S.J. Lee, A machine-learning approach using PET- based radiomics to predict the histological subtypes of lung cancer, Clin. Nucl. Med. 44 (12) (2019) 956–960.

[13] Y. She, Z. Jin, J. Wu, J. Deng, L. Zhang, H. Su, C. Chen, Development and validation of a deep learning model for non–small cell lung cancer survival, JAMA Netw. Open 3 (6) (2020) e205842, e205842.

[14] W. Rahane, H. Dalvi, Y. Magar, A. Kalane, S. Jondhale, Lung cancer detection using image processing and machine learning healthcare, 2018, March, in: International Conference on Current Trends towards Converging Technologies (ICCTCT), IEEE, 2018, pp. 1–5.

[15] V. Rajasekar, S. Krishnamoorthi, M. Saraˇcevi´c, D. Pepic, M. Zajmovic, H. Zogic, Ensemble machine learning methods to predict the balancing of ayurvedic constituents in the human body: ensemble machine learning methods to predict, Comput. Sci. 23 (1) (2022), https://doi.org/10.7494/csci.2022.23.1.4315.