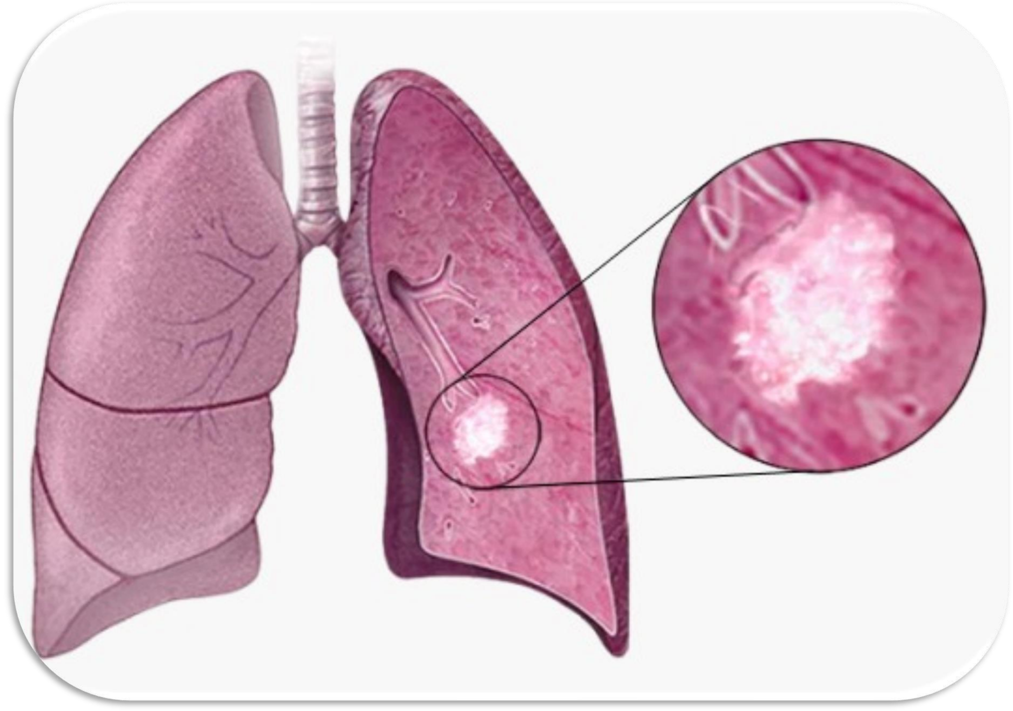
Lung Cancer Detection Using Convolutional Neural Network on CT Scan Images

GROUP # 8

Abstract

Despite advancements, cancer remains a serious health concern. Lung cancer, particularly deadly and often difficult to diagnose early, is a prime example. This study proposes using Convolutional Neural Networks (CNNs) to analyze CT scans for faster, more accurate lung cancer detection. The CNN achieved high training and validation accuracy (96.11% and 82.33%, respectively), suggesting its potential to improve diagnosis and patient outcomes.

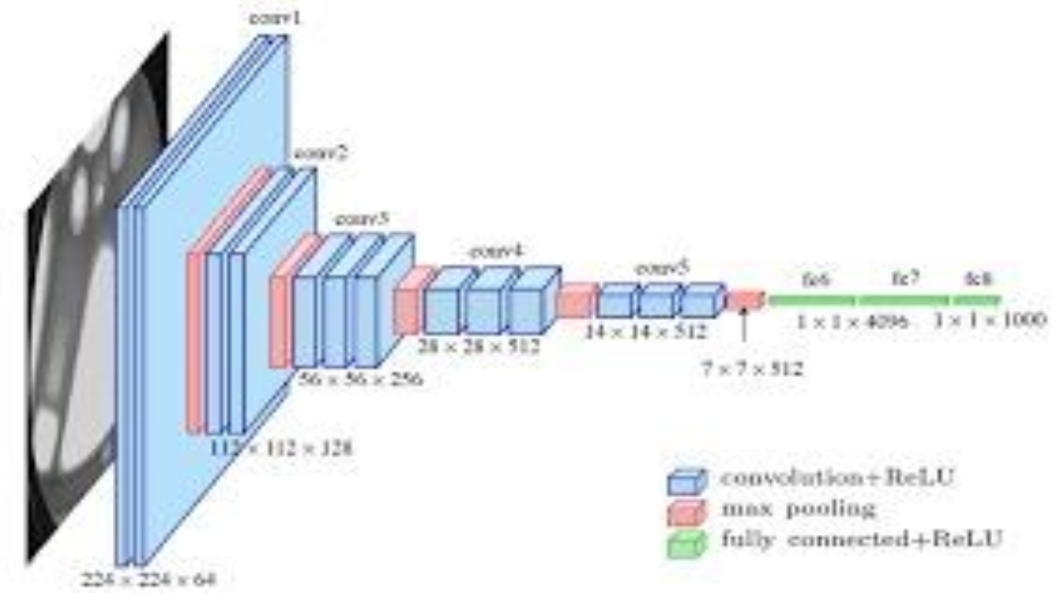
OBJECTIVES



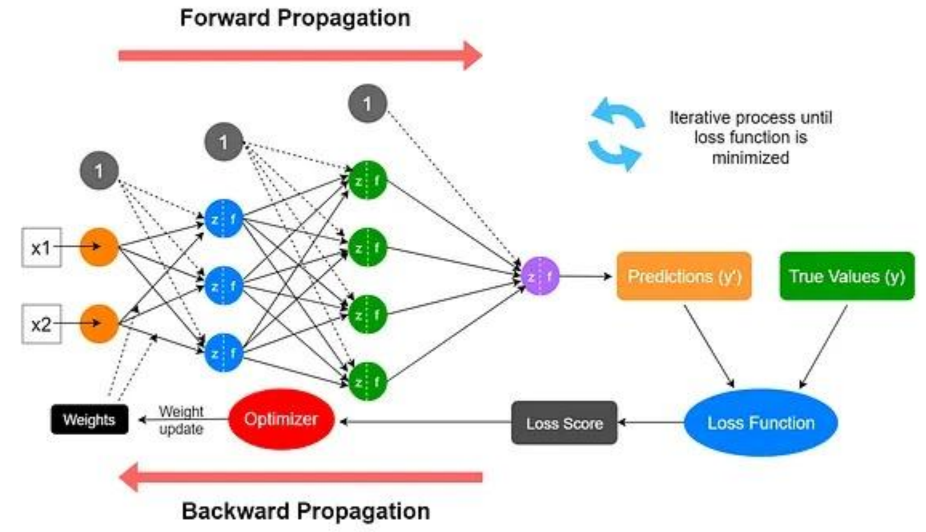
- Improve early diagnosis of lung cancer, a critical factor in successful treatment.
- Leverage deep learning, a powerful subset of AI, for lung Cancer Detection
- Develop a high-accuracy deep learning model for lung nodule detection in chest CT scans.
- Reduce false positives in lung cancer screening by improving efficiency and reducing unnecessary biopsies.
- Improve radiologist through automated lung nodule detection.
- Enable earlier lung cancer diagnosis by identifying suspicious nodules To assist radiologists.
- Evaluate the cost-effectiveness of integrating the deep learning model into lung cancer screening programs.
- Contribute to the development of AI-powered tools for improving public health outcomes in lung cancer detection and treatment.

METHODOLOGY

This work presents a methodology for lung cancer detection in CT scans utilizing a Convolutional Neural Network (CNN). To extract valuable features from the images, a pre-trained VGG-16 model, known for its strong image recognition capabilities, was employed. We leverage k-fold cross-validation, a technique that splits the data into multiple folds for training and validation, to ensure model robustness and prevent overfitting.



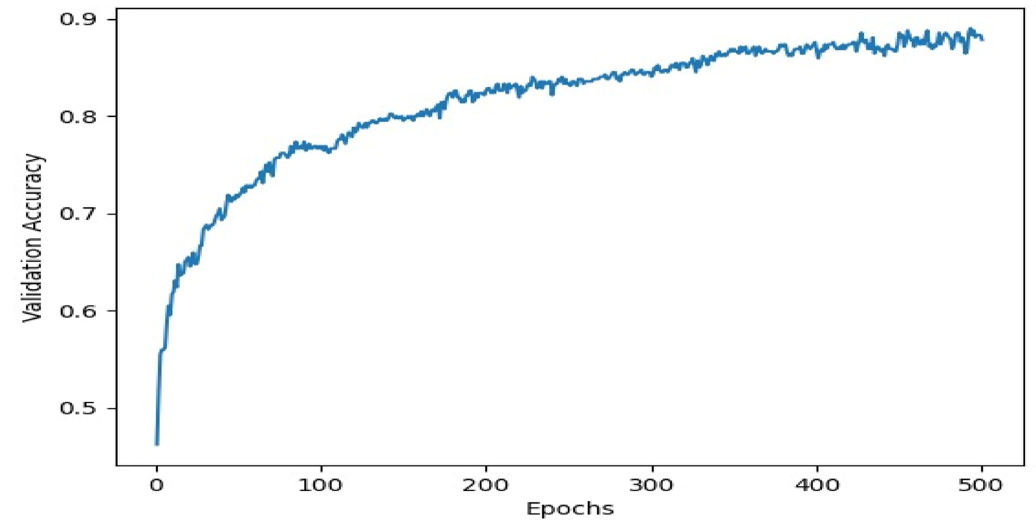
Following feature extraction by VGG-16, a custom neural network was designed using TensorFlow for training and testing. This network comprises a series of layers: first, a Flatten layer transforms the extracted features into a one-dimensional vector. A Dropout layer with a 50% dropout rate is then implemented to prevent overfitting. This is followed by a Dense layer with 32 nodes and a ReLU activation function, which introduces non-linearity for improved learning. Finally, a single-node Dense layer with a sigmoid activation function outputs the predicted class (cancerous or healthy).



To enhance the model's ability to generalize to real-world scenarios, data augmentation techniques were implemented. These techniques artificially increased the dataset size and diversity by zooming, shearing, and adjusting image brightness. Additionally, data normalization was performed to prevent overfitting during training. The training process utilized a dataset containing 234,300 images, categorized as positive (cancerous) or negative (healthy), for training, testing, and validation purposes.

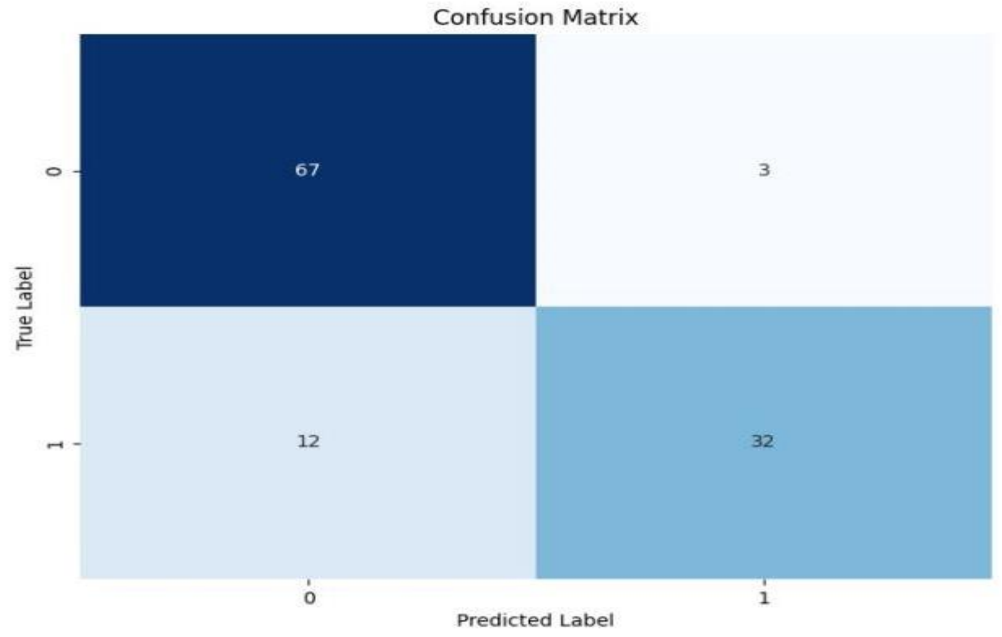
RESULT

It achieved a high training accuracy of 96.11%, indicating the model effectively learned from the training data. However, validation accuracy, which reflects performance on unseen data, was 82.33%. This highlights the importance of techniques like data augmentation to improve generalizability. Test accuracy, evaluated on a separate testing set, reached 85.08%.



Additionally a confusion matrix (Figure 3)

provides a detailed breakdown of the model's performance for each image category (positive/negative for cancer) in the validation set. This matrix allows us to identify potential areas for improvement, such as reducing false negatives (missed cancer cases).



Model performance was further analyzed using the F1 score (81%) ,precision (91.42%) and recall (72.72%).

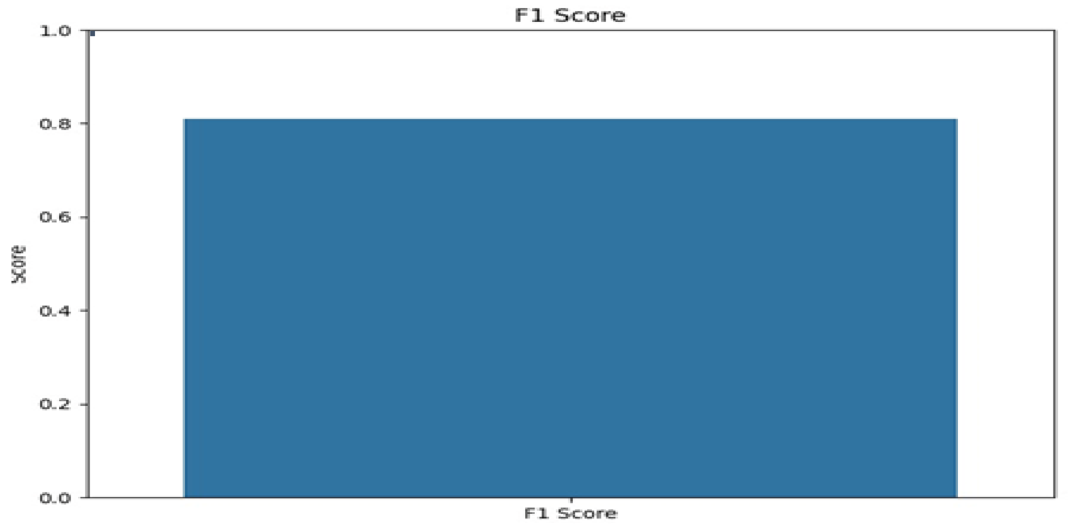
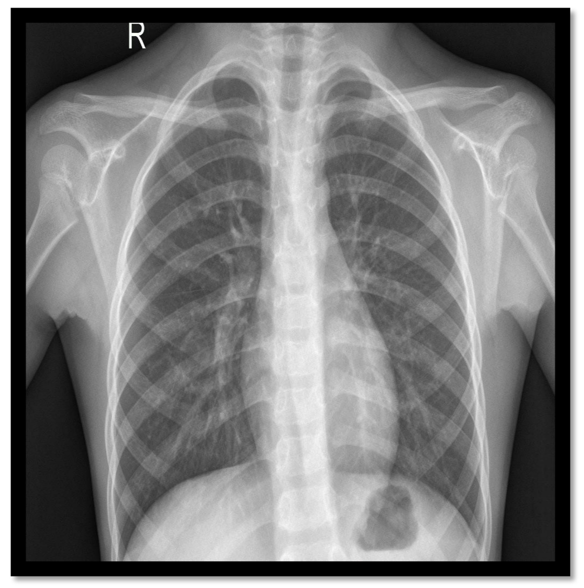


Figure 2(b) visually represents this score

CONCLUSION

The model aimed to classify images into two categories: positive (cancerous) and negative (healthy) cell carcinoma. The CNN architecture achieved promising results, reaching a training accuracy of 96.11% and a validation accuracy of 82.33%. To further assess the model's effectiveness, we calculated precision (91.42%), indicating a high percentage of true positives among predicted cancer cases. However, recall (72.72%) highlighted the need for improvement in identifying all cancer instances.



Not cancerous

Cancerous

A confusion matrix was also generated to visualize the model's performance across different categories, aiding in pinpointing areas for potential improvement. These findings suggest the CNN's capability for lung cancer detection, but further research could focus on enhancing its ability to accurately identify all cancer cases.

REFERANCE

- Lung Cancer Diagnosis and Treatment Using AI and Mobile Applications https://doi.org/10.3991/ijim.v14i17.16607 P. Rajesh , A. Murugan, B. Muruganantham S. Ganesh Kumar SRM Institute of Science and Technology, Tamil Nadu, India dexterrajesh@gmail.com.
- Deep Learning with Python Book by Francois Chollet
- Hands-computer-vision-tensorflow-2 by Benjamin Planche, Eliot Andres