ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR1516

Lung Cancer Detection using Ensemble Techniques

[1]Piyush Choudhari; [2]Yash Soniminde; [3]Anubhav Sharma; [4]Prisha Shah; [5]Amish Faye; [6]Nita J. Mahale
Department of Artificial Intelligence And Data Science, DY Patil College Of Engineering, Akurdi, Pune, Maharashtra, India

Abstract:- This paper implements a system for enhancing the detection of lung cancer through an ensemble approach, which amalgamates the predictive outputs generated by three distinct convolutional neural networks (CNNs): ResNet50, EfficientNet, and InceptionNet. Leveraging the diverse architectural features and learning capabilities of these CNNs, the ensemble method aims to synergistically fuse their individual predictions to achieve heightened accuracy and robustness in identifying potential lung cancer manifestations.

Keywords:- *Lung Cancer Detection; CNN; Ensemble Techniques; Resnet50; VGG16; Inceptionnet.*

I. INTRODUCTION

This paper introduces a methodology to enhance lung cancer detection by integrating predictions from ResNet50, EfficientNet, and InceptionNet convolutional neural networks. Leveraging the architectural features of these models, the ensemble approach averages their outputs, aiming for heightened accuracy and robustness in identifying potential lung cancer manifestations. Through evaluation, this study demonstrates the accuracy of the proposed ensemble method to be 90.2%, offering a promising avenue for advancing clinical diagnosis and patient outcomes in health management. A system has been proposed to streamline the operational efficiency of organizations, researchers, and medical professionals by implementing automated processes. This system entails the development of application programming interfaces (APIs) to facilitate seamless interaction with the model and databases. Its core functionality involves the classification of CT-Scans in large-scale batches, followed by the systematic storage of the processed data within the database infrastructure.

Goal and Objectives

- Implementing a model to classify CT-Scan images of lungs as cancerous or non cancerous.
- Delivering the model to the end user in a cost effective and quick way.
- Diagnosing patients and detecting early signs of lung cancer to encourage early intervention.

II. MODEL TRAINING

- The data collection process started with grouping the CT-scan images of lungs, which formed the most critical part in the training of the following models. The source of the images, the well-established Kaggle dataset, the Chest CT-Scan Images Dataset, represents a thorough collection of medical imaging data of 1400+ images. Since the dataset is large, it was necessary to properly group them, which is why it was systematically divided into three distinctive directories: train, test, and validation. Particularly, the train directory included 70% of the images, whereas the test and validation directories contained 20% and 10%, accordingly, for the sake of later robust model evaluation.
- After the data collection was complete, a detailed data cleaning phase was performed. Thus, every image was classified into one of four classes: normal, adenocarcinoma, large cell carcinoma, and squamous cell carcinoma. Such type of categorization was essential for the following stages of training and validation. Moreover, each inappropriate or unusable image was found and eliminated from the dataset to ensure its quality and purity.
- With a clean and well-organized dataset in place, the next step was data preprocessing. This phase involved a series of essential transformations to prepare the images for model training. Initially, the images were relabeled and indexed according to their respective classes to facilitate efficient data handling. Subsequently, they underwent rescaling and resizing to a standardized dimension of 224 by 224 pixels, ensuring uniformity across the dataset. To further enhance model generalization and robustness, various data augmentation techniques were applied, including horizontal flipping, contrast adjustment, and grayscale conversion.
- Following data preprocessing, the focus shifted to model loading and initialization. Pre-trained convolutional neural networks (CNNs) such as ResNet50, EfficientNet, and InceptionNet were selected for their well-established architectures and superior performance in image classification tasks. These models were loaded along with their pre-trained weights, allowing them to leverage the knowledge gained from extensive training on large-scale image datasets.
- The subsequent training phase involved feeding the preprocessed images into the input layers of the CNNs. The models were trained with non-trainable weights, allowing them to learn and extract meaningful features from the input data over multiple epochs. A total of 15 epochs were chosen to balance between model convergence and computational efficiency.
- Upon completion of individual model training, an ensemble method was employed to combine the predictive outputs of the three CNNs. This ensemble model leveraged the collective intelligence and diverse perspectives of multiple CNN architectures, leading to improved prediction accuracy and robustness.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR1516

- Model evaluation was conducted using test data. Precision, recall, F1 score were calculated and rate of type 1 errors and type 2 errors were noted.
- Model was saved to disk. APIs were created to interact with the model and the database.

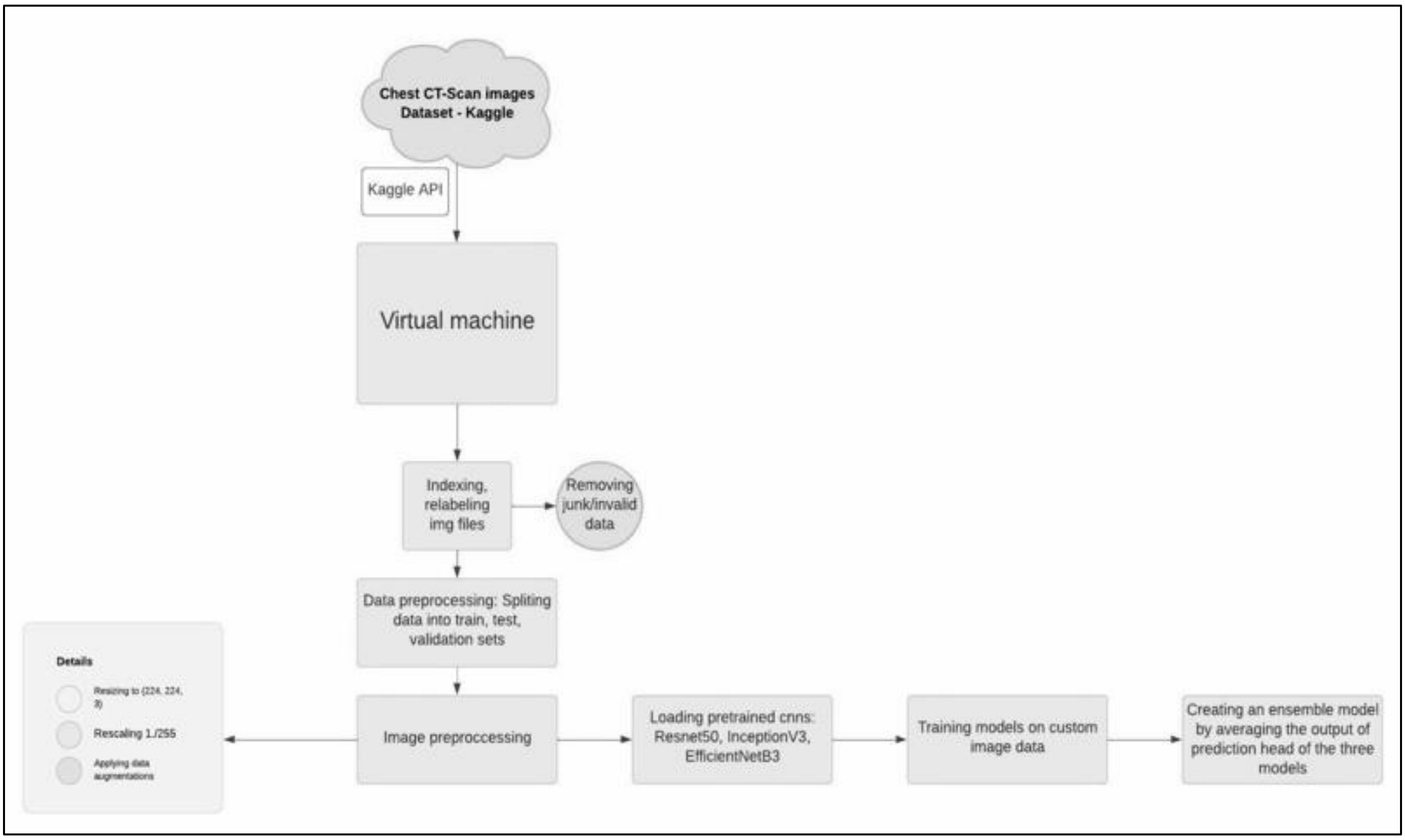

Fig 1: Model

III. RESULTS

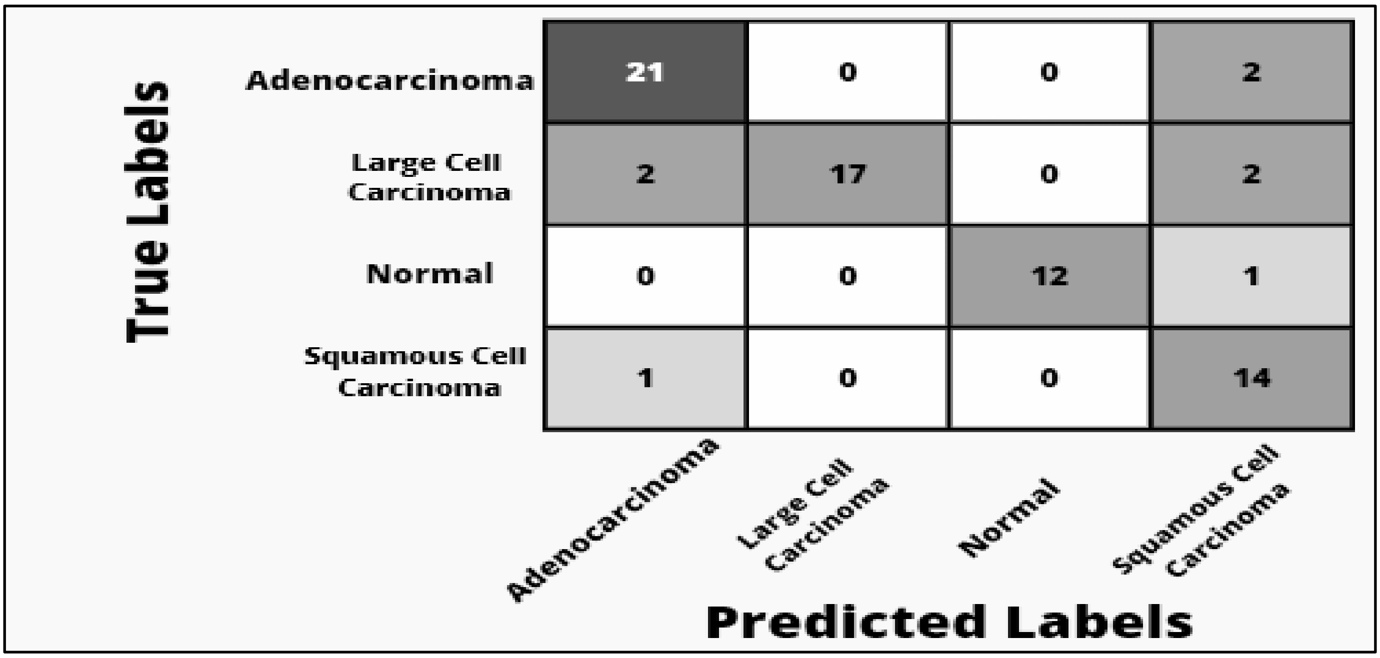

Fig 2: Predicted Lables

- Overall Accuracy = 90.2%
- Overall Sensitivity = 89.5%

IV. MODEL TRAINING WORKFLOW

➢ *APIs*

- Using the implemented APIs, we have also created a way for data to be exchanged seamlessly between multiple processes, using this to effectively interface with both the model and the database.
- As an outcome of this interphase, we have also set up an endpoint at which a user can send his or her CT-scan images in binary format. After submission, the system triggers the classification and sends back the results in a well-structured JSON format. These include all the classification outcomes and other related details of the respective scans and the patients offering them for classification. After that stage, the classified data is stored systematically in the database, complete with its names and other related details .
- APIs are built with interoperability in mind, facilitating seamless integration with existing software systems and workflows commonly utilized in medical and research settings.
- Users receive prompt feedback on their submitted CT-scans, enabling them to make informed decisions rapidly based on the classification outcomes.

V. CONCLUSION

In conclusion, this paper presents a methodology for enhancing lung cancer detection through the integration of predictions from ResNet50, EfficientNet, and InceptionNet convolutional neural networks. By leveraging the architectural features of these models and employing an ensemble approach to average their outputs, we were able to achieve good accuracy and stability in identifying potential lung cancer diagnosis.

Furthermore, a proposed system is outlined to streamline operational efficiency for organizations, researchers, and medical professionals. This system is designed to automate processes through the development of application programming interfaces (APIs), enabling seamless interaction with the model and databases. Its core functionality revolves around the classification of CT-Scans in large-scale batches, followed by the systematic storage of processed data within the database infrastructure. By integrating efficient workflow management, this system presents a significant step forward in optimizing healthcare processes and facilitating timely and accurate diagnosis.

REFERENCES

[1]. Nageswaran S, Arunkumar G, Bisht AK, Mewada S, Kumar JNVRS, Jawarneh M, Asenso E. Lung Cancer Classification and Prediction Using Machine Learning and Image Processing. Biomed Res Int. 2022 Aug 22;2022:1755460. doi: 10.1155/2022/1755460. Retraction in: Biomed Res Int. 2024 Jan 9;2024:9851527. PMID: 36046454; PMCID: PMC9424001.

[2]. B. S, P. R and A. B, "Lung Cancer Detection using Machine Learning," 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, 2022, pp. 539-543, doi: 10.1109/ICAAIC53929.2022.9793061.

[3]. https://my.clevelandclinic.org/health/diseases/4375-lung-cancer

[4]. https://www.cancer.org/cancer/types/lung-cancer/about/what-is.html

[5]. https://www.cdc.gov/cancer/lung/basic_info/index.html