**IEEE Base Paper Title:**

**Prediction of Lung Cancer using Convolution Neural Networks**

**(or)**

**Our Proposed Project Title:**

**DeepLungNet: Accurate Lung Cancer Prediction using InceptionV3-based CNNs on Histopathological Images**

**IEEE BASE PAPER ABSTRACT:**

Lung cancer is one of the deadly diseases whose prediction is required to reduce the death rate. So, Artificial intelligence is used on CT scan images are used for achieving better accuracy in an automated manner. Deep Learning is one of the emerging trends for predicting values. Convolution Neural Networks is one of the deep learning algorithms which implemented to sample produces better outcomes as compared to other machine learning algorithms. In this paper, the data-set has been taken with 1000 images of chest scans for different types of lung cancers such as Adenocarcinoma, Large Cell Carcinoma and Squamous Cell Carcinoma. Multiple machine learning algorithms has been compared and then it has been confirmed that CNN is one of the best among all to check the accuracy of the prediction. The paper includes VGG – 16 implemented on data set having different types of Lung Cancer and thus helping to check the severity and precautions for the same in a distinct manner.

**OUR PROPOSED PROJECT ABSTRACT:**

Lung cancer remains a significant global health concern, demanding accurate and early detection for effective treatment. This project presents a novel approach to predict lung cancer using Convolutional Neural Networks (CNNs), specifically employing the InceptionV3 architecture. The aim of this project is to develop a robust and accurate model to classify the Lung Cancer. The dataset used in the project contains histopathological images of Lungs which is classified into three classes: Lung benign tissue, Lung adenocarcinoma, and Lung squamous cell carcinoma. The project is implemented in Python, leveraging its versatile libraries and frameworks to handle the complexities of deep learning tasks. The dataset utilized for training and validation comprises 15,000 histopathological images, synthetically augmented from 750 original images of lung tissue. The dataset is HIPAA-compliant and validated, ensuring its quality and ethical compliance. Throughout the project, careful attention is paid to achieving high model accuracy. The training process yielded an impressive training accuracy of 94.00%, indicating that the model effectively learned the intricacies of the training data. Moreover, the validation accuracy reached 93.00%, illustrating the model's generalization capability on unseen data. The use of InceptionV3 architecture enables the model to capture complex patterns and features present in the lung tissue images, facilitating accurate classification across the three distinct classes. By incorporating transfer learning, the model benefits from the pre-trained weights of InceptionV3 on large-scale image datasets, thereby accelerating the convergence of training and enhancing its ability to discern features relevant to lung cancer classification. This project's significance lies in its potential to assist medical professionals in accurately diagnosing lung cancer at an early stage, leading to timely interventions and improved patient outcomes. The model's high accuracy and generalization demonstrate its potential practical application in real-world scenarios, complementing the efforts of healthcare practitioners in the fight against lung cancer. Its notable accuracy and capability to handle a dataset with three classes make it a valuable contribution to the field of medical image analysis and lung cancer diagnosis. The project sets the stage for further research and exploration of CNN-based models in other medical imaging applications, potentially revolutionizing the way we detect and treat various diseases in the future.

**EXISTING SYSTEM:**

* The existing system for lung cancer prediction was built on the foundation of the VGG-16 architecture, a widely-used deep learning model known for its deep convolutional layers. The system's primary objective was to classify chest scans into different types of lung cancers, including Adenocarcinoma, Large Cell Carcinoma, and Squamous Cell Carcinoma, based on the features extracted by the VGG-16 model.
* The existing system model VGG-16, is characterized by its deep architecture, consisting of 16 layers, including 13 convolutional layers and 3 fully connected layers. Its convolutional layers are designed to capture hierarchical and complex features from the input images, while the fully connected layers are responsible for making the final predictions.
* In the existing system, to train and evaluate the model, a dataset containing 1000 images of chest scans with different types of lung cancers was utilized. This dataset served as the foundation for the model's learning process, enabling it to discern patterns and characteristics specific to each lung cancer type.
* Throughout the development and training of the existing system, the primary focus was to achieve high prediction accuracy. The system achieved an accuracy of 77.62%, indicating its ability to correctly classify lung cancer types in a considerable portion of the cases.
* While the existing system demonstrated promising results, the continuous advancement in deep learning and convolutional neural networks inspired the need for exploration and improvement.

**DISADVANTAGES OF EXISTING SYSTEM:**

* Less Accuracy: The existing system model achieved 77.62% of accuracy in predicting lung cancer types which is relatives less accurate. The existing system exhibits a relatively lower accuracy rate compared to more other. This can result in false positives or false negatives, leading to potential misidentifications or missed leading of cancer predictions which endanger human life.
* Limited Model Depth: The VGG-16 architecture is relatively shallow compared to more modern and sophisticated architectures like InceptionV3. Its 16-layer design may restrict the model's ability to capture intricate and subtle features present in the histopathological images of lung tissue, potentially leading to suboptimal performance in distinguishing between different cancer types.
* High Computational Complexity: VGG-16's deep architecture requires a substantial amount of computational resources for both training and inference. This high computational complexity results in longer training times and increased hardware demands, making it less practical for real-time or resource-constrained applications.
* Overfitting Risk: With a smaller dataset of 1000 images, the existing system may be more susceptible to overfitting, where the model becomes too specialized in the training data and fails to generalize well on unseen data. This limitation could compromise the system's ability to accurately predict lung cancer types on new and diverse patient cases.
* Limited Dataset Size: The dataset used for training the existing system consists of only 1000 images. While this dataset served as a starting point, it might not fully capture the variations and complexities present in lung cancer cases, leading to potential biases and reduced robustness in real-world scenarios.
* Less Transfer Learning Benefit: Transfer learning, a powerful technique that utilizes pre-trained models on large-scale datasets, is less effective with VGG-16 due to its relatively older architecture. This limitation might slow down the training process and result in lower convergence rates compared to more modern architectures like InceptionV3.
* Limited Feature Representation: VGG-16's deep architecture might not fully capture and represent the subtle textures and patterns specific to lung cancer tissue in the histopathological images. This limitation could potentially affect the model's ability to distinguish between closely related cancer types accurately.
* Overall, while the existing system using the VGG-16 architecture showed promising results at the time of its development, these disadvantages highlight the need for more advanced and modern deep learning architectures, larger datasets, and improved methodologies to achieve higher accuracy and reliability in lung cancer prediction.

**PROPOSED SYSTEM:**

* The proposed system aims to revolutionize lung cancer prediction by leveraging state-of-the-art deep learning techniques, specifically employing the InceptionV3 architecture. This advanced system is designed to overcome the limitations of the existing VGG-16-based approach and enhance accuracy and robustness in classifying histopathological images of lung tissue into three distinct classes: Lung benign tissue, Lung adenocarcinoma, and Lung squamous cell carcinoma.
* The core of the proposed system is the InceptionV3 architecture, which is a deep convolutional neural network renowned for its exceptional performance in image recognition tasks. The model's inception modules, featuring multiple parallel convolutional operations, enable it to capture a wide range of image features at different scales, facilitating better feature representation and more accurate classification.
* The proposed system is implemented using Python, a versatile and widely-used programming language known for its rich libraries and frameworks suitable for deep learning tasks. Python's flexibility allows seamless integration with deep learning libraries like TensorFlow or PyTorch, empowering efficient model development and training.
* The proposed system utilizes a carefully curated and augmented dataset containing 15,000 histopathological images of lung tissue. The images are generated from an initial dataset of 750 images of lung tissue (250 benign lung tissue, 250 lung adenocarcinomas, and 250 lung squamous cell carcinomas) using the Augmentor package. The dataset's diversity and size provide the model with a comprehensive learning experience, contributing to its accuracy and generalization capability.
* During the training phase, the proposed system achieves an impressive training accuracy of 94.00%, indicating its ability to learn and adapt to the complex patterns present in the training data. The validation accuracy attains 93.00%, signifying the model's capacity to generalize and accurately predict lung cancer types on previously unseen data.
* The proposed system leverages transfer learning, initializing the InceptionV3 model with pre-trained weights on large-scale image datasets. This technique jumpstarts the training process, allowing the model to inherit knowledge about general image features from the pre-training phase. The subsequent fine-tuning process tailors the model to specialize in lung cancer classification, effectively utilizing the dataset's information to achieve high accuracy.
* The proposed system classifies histopathological images into three classes, representing different types of lung tissue: Lung benign tissue, Lung adenocarcinoma, and Lung squamous cell carcinoma. This three-class classification enables accurate identification and differentiation of diverse lung cancer cases, empowering healthcare professionals with essential information for diagnosis and treatment planning.
* The proposed system's accuracy and efficiency offer significant clinical relevance, enabling medical practitioners to make informed decisions regarding lung cancer diagnosis and treatment. By providing accurate predictions, the system enhances early detection and intervention, potentially leading to improved patient outcomes and better management of lung cancer cases.
* The proposed system represents a substantial advancement in lung cancer prediction, effectively addressing the limitations of the earlier VGG-16-based approach. By adopting the InceptionV3 architecture and augmenting the dataset, the system achieves higher accuracy and robustness, positioning itself as a valuable tool in the fight against lung cancer.

**ADVANTAGES OF PROPOSED SYSTEM:**

* Higher Accuracy: The utilization of the InceptionV3 architecture in the proposed system leads to a remarkable increase in accuracy compared to the existing system. Achieving a training accuracy of 94.00% and a validation accuracy of 93.00%, the proposed system demonstrates its ability to make highly precise predictions, improving the reliability of lung cancer diagnosis.
* Robust Generalization: The proposed system exhibits superior generalization capabilities, enabling it to accurately classify lung cancer types on previously unseen data. By effectively capturing diverse and intricate features from the augmented dataset, the model becomes more resilient to variations in lung tissue images, enhancing its applicability in real-world clinical settings.
* Augmented Dataset: With the use of the Augmentor package, the proposed system expands the dataset from 1000 images in the existing system to 15,000 images. This augmented dataset provides the model with a more comprehensive representation of lung tissue variations, reducing the risk of overfitting and contributing to its improved accuracy.
* Transfer Learning Efficiency: By employing transfer learning with pre-trained weights on the InceptionV3 model, the proposed system significantly accelerates the training process. This efficiency allows the model to benefit from the knowledge acquired on large-scale image datasets, leading to faster convergence and better overall performance.
* Three-Class Classification: The proposed system classifies histopathological images into three distinct classes: Lung benign tissue, Lung adenocarcinoma, and Lung squamous cell carcinoma. This three-class classification enhances the system's clinical utility, enabling precise identification and differentiation of various lung cancer types, aiding medical practitioners in making well-informed decisions.
* Practical Clinical Application: With its high accuracy and robustness, the proposed system holds immense practical value in clinical settings. Healthcare professionals can utilize the system as a valuable tool for early lung cancer detection, facilitating timely interventions and potentially improving patient outcomes.
* Ethical and Compliant Data: The proposed system maintains ethical standards by using a dataset that adheres to HIPAA compliance and validation procedures. This ensures patient privacy and data protection while supporting responsible and reliable research in the medical domain.
* Advanced Deep Learning Techniques: By embracing the InceptionV3 architecture, the proposed system demonstrates the integration of cutting-edge deep learning techniques in medical image analysis. This showcases the potential of modern artificial intelligence approaches in revolutionizing lung cancer prediction and diagnosis.
* Scalability: The proposed system's design allows for scalability, enabling its extension to accommodate larger datasets or include additional lung cancer subtypes. This adaptability contributes to its longevity and relevance in an ever-evolving healthcare landscape.
* Overall, the proposed system's advantages highlight its superiority over the existing system, establishing it as an innovative and promising solution for accurate lung cancer prediction using convolutional neural networks. With higher accuracy, robust generalization, and practical clinical relevance, the system presents a significant step forward in leveraging deep learning for early detection and management of lung cancer cases.

**SYSTEM ARCHITECTURE:**

Predicted Results: Lung Cancer Classification (Benign, Adenocarcinoma, & Squamous cell carcinoma)

Histopathological Lung Image Dataset

InceptionV3 Architecture

Performance Analysis and Graph

**SYSTEM REQUIREMENTS:**

**HARDWARE REQUIREMENTS:**

* System : Pentium i3 Processor.
* Hard Disk : 500 GB.
* Monitor : 15’’ LED
* Input Devices : Keyboard, Mouse
* Ram : 6 GB

**SOFTWARE REQUIREMENTS:**

* Operating system : Windows 10 Pro.
* Coding Language : Python 3.10.9
* Web Framework : Flask

**REFERENCE:**

Aanchal Vij; Kuldeep Singh Kaswan, “Prediction of Lung Cancer using Convolution Neural Networks” 2023 International Conference on Artificial Intelligence and Smart Communication (AISC), IEEE CONFERENCE, 2023.