**Abstract:**

This study presents an innovative and comprehensive approach for the diagnosis of COVID-19 leveraging chest X-ray (CXR) images, incorporating a diverse array of machine learning techniques. The dataset utilized comprises CXR images categorized into COVID-19 cases with and without pneumonia manifestations, addressing the complexities of differential diagnosis in respiratory infections. At the core of the classification framework lies the K Nearest Neighbors (KNN) classifier, serving as the foundational model. Building upon this foundation, centroid-based feature extraction is employed to enhance the representation and discrimination of image features, facilitating more accurate classification. Furthermore, the study explores the application of AntColony optimization to fine-tune the KNN model's hyperparameters, optimizing its performance for COVID-19 diagnosis. Additionally, a novel Custom Linear Discriminant Analysis (CustomLDA) technique is introduced to extract discriminative features tailored specifically to the nuances of COVID-19 diagnosis from CXR images. The integration of these advanced methodologies demonstrates promising results in accurately distinguishing COVID-19 cases with and without pneumonia manifestations. By showcasing the potential of integrating multiple machine learning techniques, this study contributes to the development of a robust and effective framework for medical image classification in the context of respiratory infections.

**Introduction:**

COVID-19, a highly contagious respiratory illness, has evolved into a global pandemic since its emergence in December 2019, with a significant impact on public health and the global economy. As of October 2, 2020, the World Health Organization (WHO) reported over 34 million confirmed cases, with 1.02 million fatalities and 23.9 million recoveries .The virus, first identified in Wuhan, China, has rapidly spread worldwide, necessitating swift and accurate diagnostic measures to curb its transmission.

Timely detection of COVID-19 is crucial for effective patient management and containment of the virus's spread. While reverse transcription-polymerase chain reaction (RT-PCR) remains the gold standard for diagnosis, its limited sensitivity, especially in the early stages of infection, underscores the need for complementary diagnostic tools . Chest X-ray (CXR) imaging, along with computed tomography (CT) scans, presents a valuable alternative for early detection, particularly in regions where RT-PCR resources are scarce or inaccessible.

Pneumonia, whether viral, bacterial, or fungal in origin, poses a significant health threat globally, contributing to substantial morbidity and mortality. The Centers for Disease Control and Prevention (CDC) reports over one million annual hospitalizations and nearly 50,000 deaths in the United States alone due to pneumonia. Given its diverse etiology and overlapping symptoms with other respiratory conditions, pneumonia diagnosis remains challenging, even for experienced radiologists.

Medical imaging, particularly CXR and CT scans, plays a pivotal role in diagnosing both pneumonia and COVID-19, offering valuable insights into lung pathology. Previous studies have highlighted the efficacy of machine learning algorithms in pneumonia detection, with deep learning techniques demonstrating superior performance. Feature extraction techniques, such as K-nearest neighbors (KNN) has shown promise in COVID-19 classification .

In light of the pressing need for accurate and efficient diagnostic methods, this study proposes a novel approach leveraging machine learning techniques, including KNN classifier enhanced with centroid-based feature extraction, AntColony optimization, and Custom Linear Discriminant Analysis (CustomLDA). By harnessing the power of these advanced methodologies, this research aims to contribute to the early and accurate diagnosis of COVID-19 and pneumonia, facilitating timely intervention and improved patient outcomes.

**Dataset:**

The dataset utilized in this research comprises two primary sources obtained from the Kaggle dataset repository. The first dataset, denoted as CXR images (Pneumonia) ,consists of 5,856 JPEG images capturing anterior-posterior chest X-ray images of pediatric patients aged one to five years. These images were sourced from the Guangzhou Women and Children’s Medical Center. A meticulous screening process was conducted to eliminate unreadable or low-quality scans. Diagnosis labels were meticulously evaluated by two expert radiologists, with a third radiologist validating the labels to ensure accuracy and minimize errors.

The dataset was partitioned into training and testing sets, with an 80-20 split, respectively. Each subset was further categorized into normal and pneumonia classes. The dataset encompasses a total of 4,185 chest X-ray images for training, 1,047 for validation, and 624 for testing. Among these, 390 cases were labeled as pneumonia, while 234 were classified as normal. Notably, all chest X-ray images in the training and testing subsets represent both pneumonia and normal classes.

The second dataset, referred to as the COVID-19 Radiography database, was collaboratively curated by researchers from multiple institutions, including Qatar University, the University of Dhaka in Bangladesh, and collaborators from Pakistan and Malaysia, in collaboration with medical professionals .This dataset comprises a diverse range of chest X-ray images, including 219 COVID-19 cases, 1,341 normal cases, and 1,345 viral pneumonia cases.

It is important to note that in this study, all features available in the dataset were utilized for analysis, and an 80% training and 20% testing split was employed to train and evaluate the machine learning models. A visual representation of sample chest X-ray images from COVID-19, pneumonia, and normal classes is provided below, facilitating a comprehensive understanding of the dataset's characteristics and variability.

**Result:**

the amalgamation of K Nearest Neighbors (KNN) classifier, centroid-based feature extraction, AntColony optimization, and Custom Linear Discriminant Analysis (CustomLDA) presents a sophisticated framework for COVID-19 diagnosis from chest X-ray images. Through meticulous experimentation and analysis, the study achieved a notable accuracy score of 0.8691588785046729 post-implementation of CustomLDA.

This improvement in accuracy underscores the effectiveness of CustomLDA in extracting discriminative features tailored to the nuances of COVID-19 diagnosis from chest X-ray images. By fine-tuning the feature extraction process to focus on relevant characteristics indicative of COVID-19 infection, CustomLDA facilitates more precise classification, thus enhancing diagnostic accuracy.

Moreover, the integration of AntColony optimization further refines the classification process by optimizing the hyperparameters of the KNN classifier. This optimization mechanism ensures that the classifier operates at its peak performance, maximizing accuracy while minimizing computational overhead.

Additionally, the utilization of centroid-based feature extraction enhances the representation and discrimination of image features, providing a solid foundation for classification. By capturing the essence of image characteristics through centroid computation, the model can effectively differentiate between COVID-19 cases with and without pneumonia manifestations.

**Conclusion:**

In summary, the proposed methodology not only demonstrates promising results but also highlights the potential of integrating multiple machine learning techniques for robust and effective medical image classification. However, further research is warranted to validate these findings on larger datasets and explore additional optimization strategies to enhance diagnostic accuracy further. Nonetheless, the presented framework lays a solid foundation for advancing the field of medical image analysis and holds promise for improving the diagnosis and management of respiratory infections such as COVID-19.