

**Combination of enhanced K-means clustering and KNN classification
techniques in Medical Diagnosis and Treatment Suggestions**
ARTIFICIAL INTELLIGENCE

DESCRIPTION:

Introduction:

COVID-19 is a widespread disease, causing thousands of deaths daily. Early diagnosis is essential to stop infection spread and save the life. The large number of COVID-19 patients is rendering health care systems in many countries overwhelmed. Hence, a trusted automated technique for identifying and quantifying the infected lung regions and severity prediction is needed to provide right treatment at the right time.

Objective

To develop automated tool using enhanced clustering and classification techniques to detect affected region, predict severity and provide optimised treatment suggestions.

Field of Invention:

The present invention relates to a method and process of medical diagnosis that helps in identifying the infected regions in CT Scan images at early stage using double iterated enhanced k-means clustering and KNN classification technique to detect infected region more accurately and provide treatment suggestions based on historical patient data.

Background of the Invention:

Medical diagnosis system is used to help the healthcare professionals to provide proper diagnosis and treatment. Huge amounts of patient's medical data like X-ray, CT scan, blood test reports are to be analysed manually to provide treatment recommendations. Also, it takes years of training for a human to assimilate these data and enable them to provide a diagnosis and recommend treatment for each new case they see.

To improve the survival of patients, there is a need to diagnose the disease and start treatment at the early stage. Now, based on my research on artificial intelligence and machine learning technologies, I identified a process to optimise the algorithms and diagnose disease, identify its severity and suggest optimised treatment required at the early stage.

Summary of Invention:

The newly discovered approach of automated medical diagnosis system will utilize combination of K-means clustering and K Nearest Neighbor (KNN) classification techniques to detect infected region quickly and efficiently. Based on the result the system can identify its severity and provide treatment suggestions based on history of patient's record. This classification can then be used as an aid for medical professionals when assessing a patient's clinical needs.

I have found that enhancing K-means clustering algorithm by iterating its process (finding centroid in iteration with minimum variance between each datapoint in a cluster), we can detect the infected region more efficiently and by applying K Nearest Neighbor (KNN) to the result we can classify and predict disease severity and optimise the treatment required based on historical patient records.

Enhanced Kmeans Clustering:

Enhanced K-means clustering is applied to historical patient data and k number of clusters and centroids are formed. The entire data set is clustered based on K-means ++ algorithm by finding sum of squared distance between each data point and the complete set process is iterated multiple times to identify the smallest variance distance between each data point and centroid in each cluster. That centroid value in the specific iteration which has shortest variance is identified. For the k number of clusters there will be k centroids. Accurate clusters and centroids values can be predicted using this double iterated K-means algorithm.

Classification using K Nearest Neighbor (KNN) algorithm:

KNN algorithm is a non-parametric approach used for the problem for classification. The approach customs the information about its neighbour points for the classification of output labels. KNN algorithm is used widely for both pattern recognition and classification applications. KNN has the capability to predict the target class with more precise in simpler manner.

The KNN algorithm used in our problem, considers the output as a target class. The problem is solved or classified by the majority voting of its neighbours, where the value of K is taken as a small and real valued positive integer. The KNN algorithm is more sensitive to the local part of the input data which makes it more unique to the classification problem.

The pseudocode for KNN algorithm is follows,

1. Calculate the Euclidean distance $d(x, x_i)$ where $i = 1, 2 \dots n$ between the points.
2. Sort the above obtained n distances in ascending order.
3. Consider k as a real-valued positive integer and the first k distances are taken from the above arrangement.
4. Now by using these k distances, estimate the k points.
5. For $k \geq 0$, k_i is the number of points correspond to the i -th category amid the k points.
6. Check the condition of $k_i > k_j$ if and if the value of i not equal to j , the keep x in the category i .

Determine the value of nearest centroid of the input data by calculating the distance between the input data and the centroids obtained using K-means ++ clustering algorithm. Calculate the distance between the new input data and the data points in the identified cluster. Sort the distance and determine k nearest neighbours based on minimum distance values. Analyse the category of those neighbors and assign the category for the test data based on majority vote. Return the value of the predicted category.

Processing CT scan Images:

Currently, there is an urgent need for efficient tools to assess the diagnosis of COVID-19 patients. Identified the feasible solutions for detecting and labelling infected tissues on CT lung images of such patients. Enhanced K++ means is used in identifying infected tissue regions in CT lung images and KNN is used in classifying the infected region based on severity and provide treatment suggestions.

Data processing:

Enhanced K-means clustering and classification techniques can be used to identify root cause factors that leads to severity in covid patients. Centroid values calculated for k number of clusters in covid patient's historical data with severity and map to its corresponding value for other factors like ALT (Alanine Aminotransferase), Myalgia, Haemoglobin, Gender, Temperature, Na⁺, K⁺, Lymphocyte count, Creatinine, Age, White blood count, Abnormality

in above factors is calculated with respect to centroid value for each cluster and identified as leading factor to severity.

Input data is mapped to the particular cluster and classification K Nearest Neighbor (KNN) algorithm is applied to each data points in the cluster. the data points are sorted and mapped to the factors leading to the severity, the most repeated factors within the given range of input data are identified as the factors that lead to severity. Treatment optimisation is based on the historical data of covid patients. This will help health care professional to identify the group of people at risk and helping them to give right treatment at the right time.

Clustering and classification techniques are used to identify the combinations of clinical characteristics of COVID-19 that predict patients at risk for more severe illness. The predictive models learn from historical data to help predict who will develop acute respiratory distress syndrome (ARDS), a severe outcome in COVID-19.

Usage in other medical applications:

Enhanced K-means Clustering method is used to group similar patterns into clusters whose members are more like each other (calculating sum of squared distance) than to members of other clusters. Then classification algorithm KNN can be used to classify the input data to the nearest cluster and sort the data points in that specific cluster and identify the occurrence of most repeated values for prediction.

Combination of enhanced K-means clustering and KNN classification algorithm can be used in following applications in Health care

Optometry:

Detection of macular degeneration and other eye related issues and severity can be predicted using and other eye related issues can be detected at early stage using clustering and KNN algorithms.

Dermatology

Clustering and classification technique in artificial intelligence and machine learning helps in processing the images. Dermatology is an imaging abundant speciality and the development of deep learning has been strongly tied to image processing. Therefore, there is a natural fit

between the dermatology and deep learning. There are 3 main imaging types in dermatology: contextual images, macro images, micro images.

Enhanced K-means clustering and KNN classification techniques helps in detection and classification of skin cancer from lesion images. Intended system can also provide the treatment recommendations to help patients to get right treatment at the right time.

Radiology

Combination of K-means and KNN techniques is used to detect and diagnose diseases within patients through Computerized Tomography (CT) and Magnetic Resonance (MR) Imaging. Combination of K-means and KNN techniques has been able to serve well for detecting abnormalities and monitoring change over time; two key factors in oncological health. double iterated enhanced K-means algorithm is used to detect a wide range of disease diagnosis and treatment recommendation. Combination of K-means and KNN processing in radiology will cut down on needed interaction time and allow doctors to see more patients and provide good treatment at the right time. the history of medical imaging shows a trend toward rapid advancement in both capability and reliability of new systems.

Disease Diagnosis

Enhanced k-means Clustering and KNN classification techniques can more accurately diagnose, classify diabetes and CVD.

Combination of K-means and KNN techniques has been able to substantially aid doctors in patient diagnosis through the manipulation of mass Electronic Health Records (EHR's). Medical conditions have grown more complex, and with a vast history of electronic medical records building, the likelihood of case duplication is high. Although someone today with a rare illness is less likely to be the only person to have suffered from any given disease, the inability to access cases from similarly symptomatic origins is a major roadblock for physicians. The implementation of intended system also helps to identify important symptoms and help the physicians ask the most appropriate questions and helps the patient receive the most accurate diagnosis and treatment possible.

Electronic health records

Electronic health records (EHR) are crucial to the digitalization and information spread of the healthcare industry. Combination of K-means and KNN algorithms can evaluate an individual patient's record and predict a risk for a disease based on their previous information and family history. This system takes in large amounts of data and creates a set of rules that connect specific observations to concluded diagnoses. Thus, the algorithm can take in a new patient's data and try to predict the likeliness that they will have a certain condition or disease. Since the algorithms can evaluate a patient's information based on collective data, they can find any outstanding issues to bring to a physician's attention and save time.

Drug Interactions

Clustering and classification algorithms used to identify drug-drug interactions in medical literature. Drug-drug interactions pose a threat to those taking multiple medications simultaneously, and the danger increases with the number of medications being taken. To address the difficulty of tracking all known or suspected drug-drug interactions, K-means and KNN algorithms can be used to extract information on interacting drugs and their possible effects from medical literature.

Conclusion:

Combination of enhanced K-means clustering and KNN classification algorithms can be used to detect abnormalities in CT scan images (Image Processing) and can also be used in medical data processing of the patients more efficiently and predict disease severity and provide required treatment suggestions. This intended system helps to provide right treatment at the right time and save life.

CLAIMS:

A method for increasing accuracy and disease classification comprising following steps:

1. Adding iterations to enhance K-Means++ to identify the smallest variance which will increase the accuracy.
2. Combining K-Means++ and K Nearest Neighbor (KNN) Classification for diagnosis, severity prediction and treatment suggestions.