Published Online First: October, 2024
https://doi.org/10.21123/bsj.2024.10835
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

Fractal Dimension and Entropy Analysis of Medical Images for KNN-Based Disease Classification

Krishna Mahapatra*, *Selvakumar R*

School of Advanced Sciences, Vellore Institute of Technology, Vellore, 632014, Tamil Nadu, India.

Received 31/01/2024, Revised 07/07/2024, Accepted 09/07/2024, Published Online First 20/10/2024

Abstract

The study of medical image-based disease detection has witnessed a notable surge in interest and success with computational methods. This work proposes a novel framework to detect diseases at earlier stages from medical images, using mathematical model and Machine Learning. Introduce two new quantitative measures for COVID-19 and Tumor disease detection: image uncertainty based on Shannon entropy and image complexity based on fractal dimension. Our result demonstrated that in COVID-19 positive images exhibited skewed pixel distribution due to the hazy regions resulting in lower entropy values for diseased cases compared to healthy ones. The second quantity, fractal dimension was measured by box counting method determines the image's complexity. The outcomes of both techniques were applied to the classification of images using the Machine Learning (ML) model k-NN (k-nearest neighbor). This complete framework provides a new and unique approach to identify and classify diverse types of images with a classification accuracy of 90% for Covid and 70% for Tumor achieved. Our work shows that Entropy and fractal dimensions can distinguish between COVID-19 and healthy patients, making them promise for early diagnosis. This manuscript presents a novel, computationally efficient and explainable methodology for disease classification that provides early-stage disease detection.

Keywords: CT scan, Fractal Dimension, K-Nearest Neighbors, Machine Learning, Magnetic Resonance Imaging, Medical Image Analysis, Shannon Entropy, Statistical Analysis.

Introduction

Convolutional neural networks (CNNs) have achieved remarkable success in detecting diseases including cancer [1], heart disease [2], and diabetes [3], Alzheimer's disease [4], schizophrenia [5], from medical images [6]. Despite being a potential tool for disease detection, CNNs are computationally expensive and time consuming [7]. Moreover, the limited interpretability of CNNs hampers trust in medical applications, hindering decision-making and acceptance by healthcare professionals. Here, this study uses two different methods for detecting two different diseases: Tumors and COVID-19.

The prevalence of brain tumors, which are severe illnesses that impair multiple brain functions, and the use of magnetic resonance imaging (MRI) technology to examine anatomical changes occurring in tumors has increased substantially in recent years [8,9]. MRI plays a vital role in helping to identify the lesion location, assess the extent of tissue involvement, and evaluate the resulting mass effect on the brain, ventricular system, and vasculature [10]. Structural MRI studies have shown that brain tumor patients typically have smaller total brain volumes and larger ventricles than healthy individuals [11]. To analyze MRI scans for research purposes, large groups of patients and healthy control subjects are used [12–14].

Published Online First: October, 2024
https://doi.org/10.21123/bsj.2024.10835
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

Chest computed tomography (CT) scans have the potential to diagnose, identify, and predict the prognosis of coronavirus disease 2019 (COVID-19) [15]. The clinical utility of CT will increase with suitable preventative safety measures, protocol optimization, and a standardized reporting system based on pulmonary findings in this disease [16,17]. However, the results of chest CT scans can be both false-positive and false-negative. Imaging techniques including chest CT findings of COVID-19 and its complications, and chest CT diagnostic accuracy using a mathematical model and machine learning for patients and healthy control subjects have been discussed [18,19].

This project proposes a mathematical model for classifying and detecting diseases (Tumors and COVID-19) based on mathematical modeling and machine learning. Here, two types of MR images, functional MRI (fMRI) and structural MRI (sMRI), are helpful techniques for examining functional and structural abnormalities in images [20–22]. This model computes different mathematical quantities from MR images, such as the moment of inertia, fractal dimensions, and entropy [23–25]. These quantities are then used by machine learning models, such as support vector machines (SVMs) [26], K-nearest neighbors (KNN) [27] and logistic regression (LR) [28], for classification using features of images with certain accuracy [29–31].

Classification is a supervised machine learning procedure in which input data are assigned to predetermined groups or classes [32]. The primary need for applying a classification technique is that all data items must be assigned to specific classes, with each data object being assigned to only a single class. Additionally, the primary emphasis of the k-NN classifier has been on datasets containing exclusively numerical features. One prevalent classification technique that relies on distance metrics is the k-NN method. The conventional k-NN classification algorithm is utilized to identify the k-NN and categorize numerical data records by evaluating the Euclidean distance between the test sample and all training samples [33]. The primary principle of k-NN algorithm is to compute the distances between the tested data samples and the training data samples to identify the nearest neighbors. The tested sample is thereafter allocated to the class of its closest neighbor. The technique can be used for both classification and regression between healthy and patient datasets.

This research proposes a novel image classification method for disease detection that leverages Shannon entropy and the fractal dimension. This work contributes to developing interpretable and computationally efficient tools for disease diagnosis through medical image analysis. Our study's results highlight this approach's potential for advancing the field of disease detection and improving patient outcomes, Table 1 shows the literature survey of existing work and its advantages and limitations.

Table 1. Literature survey of existing work and their advantages and limitations

Author	Method	Classifier	Entropy/ Fractal Dimension	Advantage
Hasan AM et al. [13]	To exploit the Convolutional Neural Network (CNN) in discriminating breast MRI scans into pathological and healthy.	CNN	Entropy	326 T2W-TSE images and 326 STIR images is 98.77%.
Utkarsh Lal et al.[34]	To conduct a comparative analysis of the various FDs that, as feature extraction measures from EEG, can discriminate Parkinson's patients who are ON and OFF medication from healthy controls using ML architecture.	KNN	Fractal Dimension	highest accuracies, yielding a mean accuracy of 99.65% for PD patients ON medication and 99.45% for PD patients OFF medication, respectively.
Raveenthini M et al.[35]	To develop a generic multiocular CAD system for Diabetic retinopathy and glaucoma diagnosis	CNN based DL	Entropy and fractal features	The proposed ensemble model resulted in accuracy as 98.08%, for three-class classification to categorize samples as

Published Online First: October, 2024
https://doi.org/10.21123/bsj.2024.10835
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

				normal, DR or glaucoma.
Munmun Swain et al.[36]	Two most effective techniques are used in separate operations, FA: Box Count Method (BCM) and Support Vector Machine (SVM).	SVM	Fractal Dimension	The combination of SVM and FD yielded the highest with 98.13% accuracy.
Dheerendranath Battalapalli et al.[37]	To comprehensively investigate the potential of fractal dimension (FD) measures in discriminating brain gliomas by examining tumor and nontumorous.	ML	Fractal Dimension	FD values of enhanced tumor regions yielded high accuracy (93%).
Anisha Rebinth et al.[38]	The Glaucoma Image Classification (GIC) is made by using different entropy features and Maximum Likelihood Classifier (MLC).	Maximum likelihood classifier	Shannon Entropy	The classification accuracy of 96% by using Shannon entropy feature and MLC
Pilar Ortiz-Vilchis et al.[39]	An Entropy-based Measure of Complexity (EMC) to analyses lung CT images of COVID-19.	Statistical Analysis	Entropy	Lung Damage Measure (LDM) increased as much as clinical classification and CO-RADS scores.
Ammini Renjini et al.[40]	For accurate prediction of rhonchi (RB) and Bronchial Breath (BB) signals, Power Spectral Density (PSD) and non-linear measures are fed as input attributes to various machine learning models.	PCA	Fractal Dimension	PCA cover approximately 86.5% of the overall variance of the data class, successfully distinguishing BB and RB signals.
Urvashi Sharma et al.[41]	To find universal threshold values for prediction and provide an optimum block size for encoding used Independent Gradient Edge Predictor16 (RIGED16) and Block Based Arithmetic Encoding (BAAE).	Block Based Arithmetic Encoding (BAAE)	Fractal Analysis	The proposed technique gives a higher coding efficiency rate compared to other techniques.

Methodology

This study employs a four-step approach to classify images based on features derived from Shannon entropy and fractal dimension. First, these features are extracted from both MRI and CT scans of individuals classified as either tumor-positive, COVID-positive, or healthy controls. Statistical analysis was used to evaluate then evaluates the significance and potential discriminative power of these features. Visualization through scatter plots further aids in understanding the data distribution. Finally, the k -nearest neighbor (k-NN) technique leverages these features to classify new images based on their similarity to known classes. Notably, preliminary analysis revealed a significant difference in average entropy values between COVID-19-positive and COVID-19-negative patients, suggesting its potential as a discriminative feature.

Entropy

Entropy, a fundamental concept in thermodynamics, statistical mechanics, and information theory, may provide a clearer understanding of its relationship to chaos, unpredictability, and a lack of information [42]. In particular, Shannon entropy is used in information theory, communication theory, and related areas such as cryptography and image processing [43]. Several survey reports have reviewed the applications of entropy in a variety of areas, such as economics [44], image processing [45], discrete mathematics [46], and signal processing [47,48]. The concept of information theory, Shannon's entropy of an image, involves the average level of "randomness" or "uncertainty" inherent to the variable's possible outcomes [49]. To measure the uncertainty or randomness of the image, then sum the probability of each pixel value in the image

Published Online First: October, 2024
https://doi.org/10.21123/bsj.2024.10835
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

multiplied by the logarithm of that probability [50]. Here the probabilistic events are the pixel intensities of the 2-D histogram between the images (normal and patient images) [51].

Mathematical Explanation:

The continuous probability distribution function $p(x)$, depicted in the histogram plots of Fig. 1.

1. **Histogram Calculation:** First, a histogram of pixel intensities from the medical image is constructed. The frequency of occurrence of each intensity level in the image is represented by this histogram.
2. **Probability Distribution**: Calculate the probability p_i of each intensity level ii by dividing the frequency of that intensity by the total number of pixels in the image. Mathematically: $p_i = \frac{n_i}{N}$, where: n_i is the frequency of intensity level ii in the histogram. N is the total number of pixels in the image.
3. **Shannon Entropy Calculation**: The most significant entropy in applications is Shannon entropy, which is defined as follows: let X be a discrete random variable as Eq. 1

$$H(X) = -\sum_i^{m-1} p_i \,.\, log_b p_i = \int p_i(x).log_b p_i(x).dx \quad (1)$$

where m is the number of gray levels (256 for 8-bit images), p_i is the probability of a pixel having gray level i, $p \in [0,1]$, and b is the base of the logarithm function, which is a natural measure of additivity. In medical imaging, entropy is a measure of complexity or irregularity in MRI structures of Tumors. For such cases, biologically a new type of cell growth is observed that is easily captured in the MR images. Analyzing such images gives rise to low entropy values. Low entropy says that the pixel values are reasonably uniform or predictable, whereas high entropy suggests that the pixel values are highly variable or random. Depending on the nature of the disease, a higher or lower entropy value might be observed, as calculated from the associated images. Next, to measure the complexity of an MR image, the fractal dimension is calculated by dividing the number of boxes needed to cover an image by the size of the boxes. A higher fractal dimension has been associated with disease, which indicates greater complexity and more detail at smaller image scales.

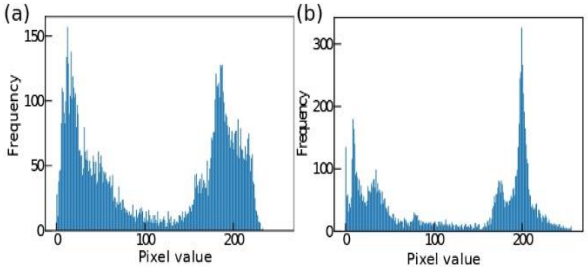

Figure. 1. Pixel value histograms for (a) healthy and (b) COVID-19-affected CT scan images. COVID-19 images have a more no uniform pixel value distribution.

Fractal Dimension

In mathematical terms, a fractal is a geometric set such that its Hausdorff-Besicovitch dimension strictly exceeds its topological dimension [52]. Benoit Mandelbrot introduced the term "fractal" to refer to non-Euclidean structures that exhibit self-similarity at various sizes.

In Euclidean n-dimensional space, a bounded set A can be considered statistically self-similar if A is the union of K_ϵ nonoverlapping subsets with respect to a scaling factor ϵ, each of which is of the form $\epsilon(A_n)$ where the K_ϵ, and A_n sets are congruent in distribution to A.

Here, using the box-counting technique the fractal dimension can be computed as Eq. 2:

$$\text{Fractal dimension} = \frac{\log(K_\epsilon)}{\log(1/\epsilon)} \quad (2)$$

where K_ϵ is the number of self-similar (invariant) shapes and ϵ is the corresponding scaling factor and the fractal dimension of the shape is estimated as shown in Fig. 2.

Let us consider analyzing a medical image of a chest X-ray of a patient with COVID-19 using the box-counting method.

1. **Box Size Determination**: Choose a range of box sizes ϵ, for example, $\epsilon = 2^K$ for k=0,1, 2,...,K, where K is the maximum scale determined by the image resolution.
2. **Covering the image**: Cover the chest X-ray image with nonoverlapping boxes of size ϵ for each scale k. Count the number of boxes K_t needed to cover the chest X-ray at each scale.

Published Online First: October, 2024
https://doi.org/10.21123/bsj.2024.10835
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

3. **Fractal Dimension Calculation**: Box-counting data are used to estimate the fractal dimension D using the formula mentioned earlier as Eq. 3:

$$D = lim_{\epsilon \to 0} \frac{\log(K_\epsilon)}{\log(1/\epsilon)} \quad 3$$

A quantitative assessment of the structural complexity or irregularity of chest X-ray can be obtained from the fractal dimension that is produced. Higher complexity or irregularity is indicated by a higher fractal dimension, whereas greater uniformity or regularity is suggested by a lower fractal dimension. This information can help with the diagnosis and treatment of several medical conditions as well as the analysis of the anatomy and texture of structures in medical images.

One advantage of this operation is that it increases the values of the dark pixels while decreasing the values of the higher brightness levels of images. It also reduces the light intensity of images with wide pixel value changes. The fractal dimension of a shape can be used to quantify its complexity. For example, a shape with a high fractal dimension is more complex than one with a low fractal dimension. Note that the parameters for the above calculations are 1. Binwidth for pixel histogram plot (for entropy), 2. Box size for fractal dimension.

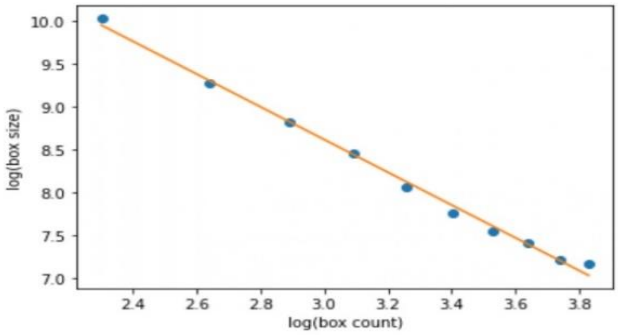

Figure 2. log (box size) vs log (box count) for a specific MRI image of Tumor. The slope of the line connecting the data points was used to determine fractal dimension of the image.

Results and Discussion

The results discussed in the section are obtained for the freely available open-source image data, which have been used, the details of which are given in the "Data availability" section.

Medical Image Analysis

Our analysis explored the potential of Shannon entropy and fractal dimension to distinguish between COVID-19 patients and healthy individuals using CT scan images [53–55]. Fig. 3 reveals distinct patterns in these measures between the diseased patients (shown in red) and healthier patients (blue). Fig. 3(a) shows that COVID-19 patients exhibited significantly lower average entropy (5.66 ± 0.68) than healthy patients (6.47 ± 0.43). Conversely, the average fractal dimension was significantly greater (2.03 ± 0.28) in the COVID-19 patients than in the healthy controls (1.81 ± 0.26), as shown in Fig. 3 (b). These findings suggest that entropy and fractal dimension might capture characteristic changes in lung tissue associated with COVID-19, with lower entropy potentially reflecting increased homogeneity and higher fractal dimension indicating greater structural complexity in diseased lungs.

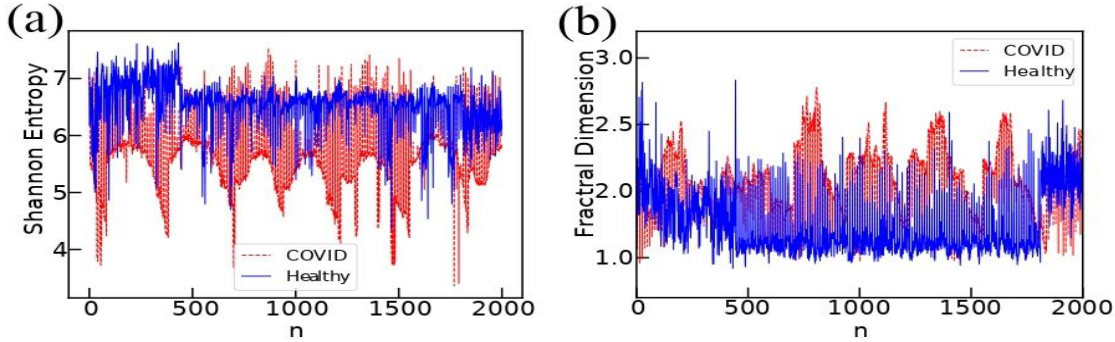

Figure 3. Shannon entropy and fractal dimension for each of the available images for the COVID-19 and non-COVID-19 patients.

By calculating the Shannon entropy and fractal dimension for both tumor and nontumor areas of interest (ROIs), a comparison can be made to discern dissimilarities in the level of complexity,

Published Online First: October, 2024
https://doi.org/10.21123/bsj.2024.10835
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

heterogeneity, or structural attributes between the two categories of regions. Due to the scarcity of MR images for tumors, patients did not exhibit any clear patterns as Fig. 4.

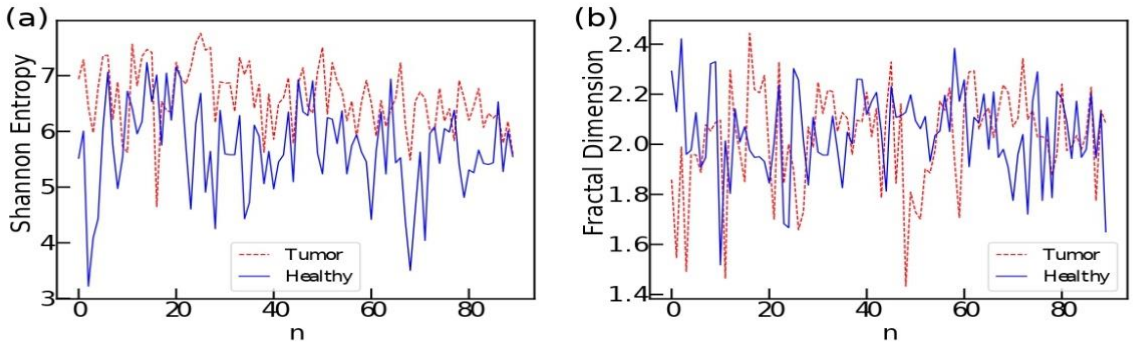

Figure 4. (a) Shannon entropy and (b) fractal dimension of Tumor data

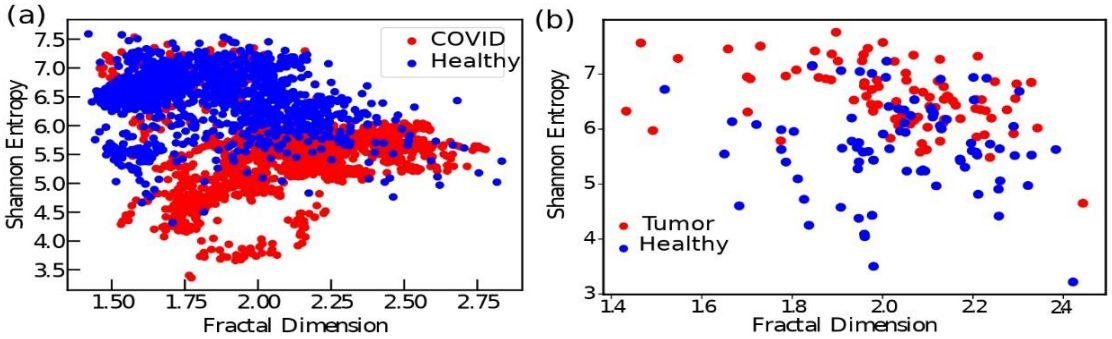

Figure 5. Shannon entropy VS fractal dimension of (a) COVID-19 (left) and (b) Tumor data images (right).

Additionally, this observes that the CT scan images corresponding to COVID-19 patients, on average, showed higher Shannon entropy (S) values, and lower fractal dimension values (d). On the other hand, healthy patients have lower entropy values and greater fractal dimensions [56–58]. This correlation inherently gives rise to the classification scheme between the diseased and healthy patients in the S–d plane as Fig. 5.

Therefore, on average, plotting S_i vs d_i will give rise to 2 separable clusters. These clusters can then be separated using various machine learning techniques, one of which has been used in this work, namely the k-nearest neighbor method. As explained earlier, diseased patients (shown in red) and healthy controls (shown in blue) were included. Two distinct clusters are visible for the COVID-19 figure, whereas due to unavailability of data the Tumor figure conceals any inherent clustering.

Theoretical Analysis

Consider a grayscale image where each pixel indicates the intensity of light. When noise is added (for example, random pixel value changes), the image becomes more unpredictable and disorganized. The unpredictability or randomness can be quantified using entropy. A uniform or homogeneous probability distribution gives rise to higher entropy values than nonhomogeneous distributions. For SARS-Cov2 patients, CT images show ground-glass opacities or cloudy areas in peripheral regions [59]. This phenomenon manifests itself in the image's increased number of darker pixel values. Therefore, the distribution of the pixel values becomes skewed toward the left of the spectra (meaning toward lower values). However, for the images of healthy patients, the distribution remains quite homogeneous through the spectra of pixel values, which results in larger values of entropy. On the other hand, the skewed pixel distribution of diseased patients shows lower entropy values. Mathematically, it can represent decreasing entropy as follows: $\boldsymbol{E_{shannon}(Healthy) > E_{shannon}(Disease)}$

where $E_{shannon}(Healthy)$ is the entropy of healthy, and $E_{shannon}(Disease)$ is the entropy of diseased. Note that this behaviour of entropy is entirely dependent on the pixel distribution and, hence, on the specific type of

Published Online First: October, 2024
https://doi.org/10.21123/bsj.2024.10835
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

disease and how it impacts the system. On the other hand, noise disrupts the self-similarity of patterns or textures, making the image appear less structured. This can be quantified using the fractal dimension. Fractal Dimension Decreasing implies that the structures within the image are becoming less complex or less self-similar. Mathematically, it can represent fractal dimension decreasing as: **$FD_{Healthy} < FD_{Disease}$**

where $FD_{Healthy}$ and $FD_{Disease}$ are the fractal dimensions indicating an increase in the fractal dimension. Therefore, entropy is a tool used in image processing to measure an image's informational density. Note that the proportionality between the Shannon entropy and the fractal dimensions, as indicated by some authors [60,61], is not observed for our case partially due to the nonuniformity of the sizes of the accessible dataset used.

It seems that in the case of some diseases, the opposite scenarios can also appear due to the following:

a. For healthy patients, there was a lower entropy and a greater fractal dimension. This suggests that the image has clarity, the pixel values are predictable, and there is considerable self-similarity among the structures.
b. For diseased patients, entropy may increase when there are artifacts, noise, or pathology (such as a tumor). Higher entropy levels arise from the randomness that the noise and abnormalities add to the image.

These relations inherently give rise to a classification technique to classify the images, the details of which are given in the next section. This technique was used for both sets of data: CT scan of COVID-19 and MR images of tumors.

Statistical Analysis

Here, statistical tests (e.g., t-test) were used to assess whether there were significant differences in entropy between healthy controls and individuals with Tumors or COVID-19. To compare and check for significant differences in entropy between MRI scans and CT scans, the t-test can be used. Levene's test was used to determine whether the variances of two or more classes were statistically different. This test is used to determine the homogeneity of variances between the classes through the following hypotheses:

Null Hypothesis (H0): The variances of the two classes are equal (homogeneous). *H*0: $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2$

Alternative Hypothesis (Ha): The variances of the two classes are not equal (heterogeneous). *Ha:* $\sigma_i^2 \neq \sigma_j^2$ for at least one pair (i, j).

Typically, the test statistic calculated by Levene's test is derived from the absolute deviations of the observations from their group mean values. The test statistic can be calculated as follows as Eq. 4:

$$W = \frac{(N-k)}{(k-1)} \times \frac{\sum_{i=1}^{k} N_i (\overline{Z}_{i.} - \overline{Z}_{..})^2}{\sum_{i=1}^{k} \sum_{j=1}^{N_i} |\overline{Z}_{i.} - \overline{Z}_{..}|^2} \quad (4)$$

The test statistic W follows an F-distribution with k−1 and N−k degrees of freedom.
N is the total number of observations.
k is the number of groups.
N_i is the number of observations in the i[th] group.
$\overline{Z}_{i.}$ is the mean of the absolute deviations from the median in the i[th] group.
$\overline{Z}_{..}$ is the overall mean of the absolute deviations from the median.
$\overline{Z}_{ij}$ is the j[th] observation in the i[th] group.
The value of W is the critical value of the F-distribution at a chosen significance level (e.g., 0.05) i.e the p-value. If the calculated W exceeds the critical value, it will reject the null hypothesis, indicating that there is evidence to suggest that the variances of the classes are not equal.

p-Value for Homogeneity of Variances

The p-value is a probability value used in hypothesis testing to determine whether the null hypothesis should be accepted or rejected. If the p-value is small (typically smaller than a chosen significance level, such as 0.05), the observed data are unlikely to have occurred under the null hypothesis. In that case, the null hypothesis is rejected, and the alternative hypothesis is accepted. If the p-value is large, the null hypothesis may fail to be rejected. The p-value, is calculated via the the F-Distribution, also known as the Fisher-Snedecor distribution, which is a continuous probability distribution used in statistics. It is used most frequently when comparing variances or determining the importance of variance differences. The p-value for an F-distribution is calculated as Eq. 5:

$$P(p\text{−}value) = 1 - F(critical, d_1, d_2) \quad (5)$$

Published Online First: October, 2024
https://doi.org/10.21123/bsj.2024.10835
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

where F (critical, d_1, d_2) is the cumulative distribution function (CDF) of the F-distribution at the critical value. d_1 is the numerator degree of freedom, and d_2 is the denominator degree of freedom. Levene's test, was used to assess the homogeneity of the variances among the classes. For healthy control and Tumor, p-value = 0.016 (< 0.05). For healthy control and COVID-19 patients, the p-value is ≈ 0, (< 0.05). This rejects the null hypothesis. There is a significant difference between the classes, which indicates that the variances are significantly different (heterogeneous).

k-Nearest Neighbour Technique (k-NN)

Distance-based algorithms are extensively employed in the context of data classification challenges. The classification process relies on the computation of distances between the test sample and the training samples for the outcome. Here, the variable k denotes the quantity of closest neighboring data points that are considered for classification or regression tasks. When the value of k = 1, the new data item is allocated to the class of its nearest neighbor. The neighbors are selected from a set of training data items for which the correct classification is known. The Euclidean distance function is commonly used as the primary metric in the k-NN algorithm.

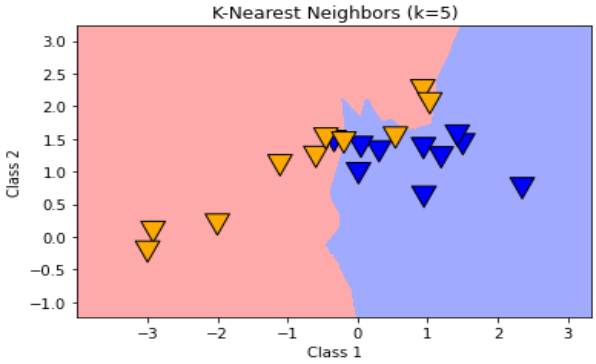

Figure 6. k-Nearest neighbor classification

	Predicted	
Actual	Positive	Negative
Positive	True Positive (TP)	False Negative (FN)
Negative	False Positive (FP)	True Negative (TN)

Figure 7. A confusion matrix for binary classification

In k-NN classification, the value of "k = 5" indicates that the algorithm will take into account the five nearest data points Fig. 6. As previously mentioned, the k-NN classifier can classify new data objects based only on their distance from labeled examples.

The confusion matrix is the predominant method employed for summarizing the efficacy of a classification system. Fig. 7 displays the confusion matrix pertaining to the binary classification scenario.

Results of k-Nearest Neighbors

Table 1, and Table 2 show the results obtained by applying k-NN to heterogeneous datasets with k = 5, 10, and 15. The k-NN algorithm is utilized for the classification of data into two distinct classes (COVID-19 and non-COVID-19 also Tumor and non-Tumor). This is achieved by identifying the k-NN of a given data point within the feature space and subsequently assigning the class label. The available dataset was randomly divided into 20% (80%) test data (train data) using the python package scikit-learn, to perform the k-NN analysis. The high values of accuracy for the COVID-19 cases indicate that our technique indeed gives rise to a classification scheme that is efficient and effective. Here, the precision is greater than the accuracy, which suggests that the model is relatively better at avoiding false positives.

Table 2. The results obtained for COVID-19 by k-NN

K Values	Accuracy	Precision	Recall
k = 5	91%	93%	89%
k = 10	92%	93%	91%
k = 15	92%	93%	91%

Table 3. The results obtained for Tumor by k-NN

K Values	Accuracy	Precision	Recall
k = 5	69%	80%	60%
k = 10	69%	61%	57%
k = 15	75%	78%	75%

Published Online First: October, 2024
https://doi.org/10.21123/bsj.2024.10835
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

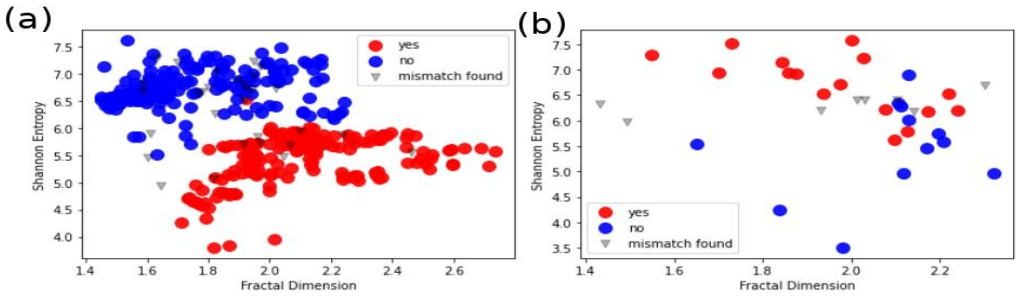

Figure 8. Classification using k-NN for (a) COVID-19 and (b) Tumor

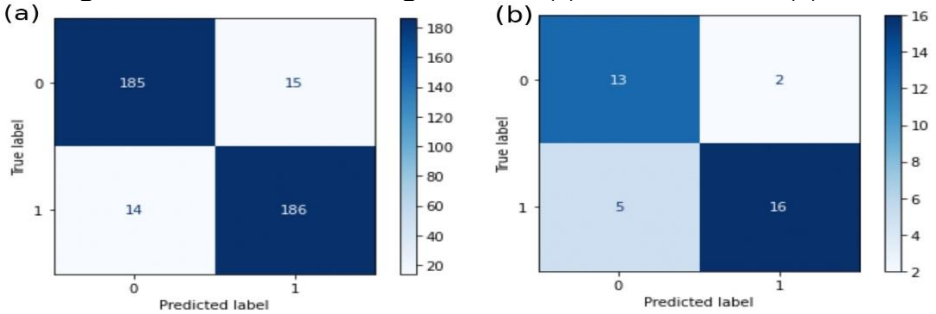

Figure 9. Confusion matrix for (a) COVID-19 (Left) and (b) Tumor (Right)

Fig. 8 shows that for the COVID-19 cases, most test data points are correctly predicted by the k-NN technique. Therefore, the number of mismatches, i.e., incorrect predictions, is less than the number of correct predictions, which results in high accuracy values, as shown in Table 2. Similar calculations for the tumor cases show that the percentage or ratio of the incorrect predictions to correct predictions increases, leading to slightly lower accuracy, as shown in Table 3. The confusion matrix is also plotted in Fig. 9. The diagonal elements indicate the true positive (TP) and true negative (TN) values, and the off-diagonal elements indicate the false positive (FP) and false negative (FN) values. The high values of TP and TN compared to those of FP and FN also show the high accuracy with which our model predicts correct results. Here also added the confusion matrix plot for the tumor case for completeness. Limitations: Although our model is computationally more effective than the contemporary methods it still suffers from the following limitations:

1. The method depends explicitly on the specific nature of the disease and how it affects the medical image. For example: showed that for COVID-19 positive patients, entropy is greater than that for non-COVID-19 patients, whereas for tumors the opposite trend is observed.
2. The correlation between entropy and fractal dimension shown in this draft can also change depending on the specific disease.

Conclusion

In this study, a novel image classification method using Shannon entropy and fractal dimension was developed to distinguish between diseased and healthy patients with two types of disease: COVID-19 and brain tumors. The behavior of these quantities varies depending on the type of disease and can be used for classification. Specifically, it was found that the entropy of COVID-19 images decreases due to changes in the pixel distribution as the fractal dimension increases. These values are plotted in the S-d plane, and two separable clusters are observed. Then, the k-NN method, a supervised machine learning technique, was used to classify the images. Our method achieved a high classification accuracy of approximately 90% for COVID-19 and 70% for brain tumors. The lower accuracy for brain tumor classification is likely due to the paucity of data. It believed that with a larger dataset, our method could

Published Online First: October, 2024
https://doi.org/10.21123/bsj.2024.10835
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

achieve even higher accuracy. Overall, this novel model provides an interpretable and computationally efficient alternative for disease classification. Our methodology can potentially be used for the early detection of various diseases such as COVID-19, Tumor. In future work, the same technique could be used for different diseases, such as Alzheimer's disease. can also use different ML techniques to improve accuracy.

Availability of data and materials

Source from the open-access Kaggle website: https://www.kaggle.com/datasets/hasimdev/ brain-mri-dataset, https://www.kaggle.com/datasets/drsurabhithorat/covid-19-ct-scan-dataset, the dataset was used for this study. Details the brain MR dataset consists of MR images used has 2 categories: Tumor and healthy controls. The COVID-19 dataset consists of CT scan images and has two categories: Non-COVID-19 and COVID-19. There are 5427 images for COVID-19, 2628 for non-COVID-19 patients. For tumors 155 and 98 for healthy controls. No special preprocessing technique was used for the analyses performed in the draft. This available dataset was randomly divided into 20% (80%) of the test data (training data) for the k-NN classifier.

Authors' Declaration

- Conflicts of Interest: None.
- We hereby confirm that all the Figures and Tables in the manuscript are ours. Furthermore, any Figures and images, that are not ours, have been included with the necessary permission for re-publication, which is attached to the manuscript.
- Authors sign on ethical consideration's approval.
- No animal studies are present in the manuscript.
- Ethical Clearance: The project was approved by the local ethical committee at Vellore Institute of Technology, Vellore.

Authors' Contribution Statement

Study and design of introduction were performed by K. M. and R. S., Material preparation, data collection, and analysis were performed by K. M. and verified by R. S., the manuscript text was written by K. M. and reviewed by R. S., the figures were prepared by K.M., Editing and approval of the final manuscript were performed by K. M. and R. S.

References

1. Ibrahim A, Mohamed HK, Maher A, Zhang B. A Survey on Human Cancer Categorization Based on Deep Learning. Front Artif Intell .2022 Jun 27; 5: 884749. https://doi.org/10.3389/frai.2022.884749
2. Dormer JD, Halicek M, Ma L, Reilly CM, Fei B, Schreibmann E. Convolutional neural networks for the detection of diseased hearts using CT images and left atrium patches. In SPIE-Intl Soc Optical Eng. 2018 Feb; 10575: 107. https://doi.org/10.1117/12.2293548
3. Swapna G, Soman KP, Vinayakumar R. Automated detection of diabetes using CNN and CNN-LSTM network and heart rate signals. Procedia Comput Sci. Elsevier B.V. 2018; 132: 1253–62. https://doi.org/10.1016/j.procs.2018.05.041
4. Ebrahimi A, Luo S, Disease Neuroimaging Initiative for the A. Convolutional neural networks for Alzheimer's disease detection on MRI images. J Med Imaging . 2021 Apr 29; 8(02): 024503. https://doi.org/10.1117/1.jmi.8.2.024503
5. Khare SK, Bajaj V, Acharya UR. SPWVD-CNN for Automated Detection of Schizophrenia Patients Using EEG Signals. IEEE Trans Instrum Meas. 2021 Apr 2; 70: 1-9. https://doi.org/10.1109/TIM.2021.3070608
6. Vaz JM, Balaji S. Convolutional neural networks (CNNs): concepts and applications in pharmacogenomics. Mol Divers. 2021 Aug 1; 25(3): 1569–84. https://doi.org/10.1007/s11030-021-10225-3
7. He K, Sun J. Convolutional Neural Networks at Constrained Time Cost, InProceedings of the IEEE conference on computer vision and pattern recognition 2015; 5353-5360: https://doi.org/10.1109/CVPR.2015.7299173

Published Online First: October, 2024
https://doi.org/10.21123/bsj.2024.10835
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

8. Naser EF, Zeki SM. Using Fuzzy Clustering to Detect the Tumor Area in Stomach Medical Images. Baghdad Sci J. 2021 Apr 30; 18(4): 1294–302. https://doi.org/10.21123/BSJ.2021.18.4.1294
9. 9. Hanif F, Muzaffar K, Perveen K, Malhi SM, Simjee SU. Glioblastoma multiforme: A review of its epidemiology and pathogenesis through clinical presentation and treatment, Asian Pac J Cancer Prev. 2017; 18(1): 3–9.https://doi.org/10.22034/APJCP.2017.18.1.3
10. Villanueva-Meyer JE, Mabray MC, Cha S. Current clinical brain tumor imaging. Clin Neurosurg. 2017 Sep 1; 81(3): 397–415. https://doi.org/10.1093/neuros/nyx103
11. Sanai N, Berger MS. Mapping the Horizon: Techniques to Optimize Tumor Resection Before and During Surgery. Clin Neurosurg. 2008 Jan 1; 55: 14-9.
12. Kirschen MP, Jaworska A, Illes J. Subjects' expectations in neuroimaging research. J Magn Reason Imaging. 2006 Feb; 23(2): 205–9. https://doi.org/10.1002/jmri.20499
13. Hasan AM, Qasim AF, Jalab HA, Ibrahim RW. Breast Cancer MRI Classification Based on Fractional Entropy Image Enhancement and Deep Feature Extraction. Baghdad Sci J. 2023; 20(1): 221–34. https://doi.org/10.21123/bsj.2022.6782
14. Nabaa A. RN. Assessment of image quality of cervical spine complications using Three Magnetic Resonance Imaging Sequences. Baghdad Sci J.2023; 20(3): 1155–63. https://doi.org/10.21123/bsj.2023.8244
15. Rong G, Zheng Y, Chen Y, Zhang Y, Zhu P, Sawan M. COVID-19 Diagnostic Methods and Detection Techniques. In: Encyclopedia of Sensors and Biosensors, First Edition. Elsevier. 2022; 1-4: 17–32. https://doi.org/10.1016/B978-0-12-822548-6.00080-7
16. Zaki SM, Jaber MM, Kashmoola MA. Diagnosing COVID-19 Infection in Chest X-Ray Images Using Neural Network. Baghdad Sci J. 2022; 19(6): 1356–61. https://doi.org/10.21123/bsj.2022.5965
17. Bhalla AS, Das A, Naranje P, Irodi A, Raj V, Goyal A. Imaging protocols for CT chest: A recommendation. Indian J Radiol Imaging. 2019 Jul; 29(03): 236–46. https://doi.org/10.4103/ijri.ijri3419
18. 18. Kwee TC, Kwee RM. Chest ct in covid-19: What the radiologist needs to know. Radiographics. 2020 Nov 1; 40(7): 1848–65. https://doi.org/10.1148/rg.2020200159
19. Dumakude A, Ezugwu AE. Automated COVID-19 detection with convolutional neural networks. Sci Rep. 2023 Dec 1; 13(1): 10607. https://doi.org/10.1038/s41598-023-37743-4
20. Shi D, Zhang H, Wang G, Yao X, Li Y, Wang S, et al. Neuroimaging biomarkers for detecting schizophrenia: A resting-state functional MRI-based radiomics analysis. Heliyon. 2022 Dec 1; 8(12) pages =e12276, https://doi.org/10.1016/j.heliyon.2022.e12276
21. Shi D, Li Y, Zhang H, Yao X, Wang S, Wang G, et al. Machine Learning of Schizophrenia Detection with Structural and Functional Neuroimaging. Dis Markers. 2021; 2021(1): 9963824. https://doi.org/10.1155/2021/9963824
22. Ghazwani H, Nadeem MF, Ishfaq F, Koam ANA. On Entropy of Some Fractal Structures. Fractal Fract. 2023 May 1; 7(5): pages=378, https://doi.org/10.3390/fractalfract7050378
23. Hayashi T, Cimr D, Fujita H, Cimler R. Image entropy equalization: A novel preprocessing technique for image recognition tasks. Inf Sci (N Y). 2023 Nov 1; 647: 119539. https://doi.org/10.1016/j.ins.2023.119539
24. Zmeskal O, Dzik P, Vesely M. Entropy of fractal systems. Computers and Mathematics with Applications. 2013 Aug; 66(2): 135–46. https://doi.org/10.1016/j.camwa.2013.01.017
25. Čukić M, Pokrajac D, Stokić M, Simić S, Radivojević V, Ljubisavljević M. EEG machine learning with Higuchi's fractal dimension and Sample Entropy as features for successful detection of depression, arXiv preprint arXiv: 1803.05985. 2018 Mar 15. https://doi.org/10.1007/s11571-020-09581-x
26. Uddin S, Haque I, Lu H, Moni MA, Gide E. Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction. Sci Rep. 2022 Dec 1; 12(1): 6256. https://doi.org/10.1038/s41598-022-10358-x
27. Zhang T, Han L, Chen W, Shahabi H. Hybrid integration approach of entropy with logistic regression and support vector machine for landslide susceptibility modeling. Entropy. 2018 Nov 1; 20(11): 884. https://doi.org/10.3390/e20110884
28. Sheth V, Tripathi U, Sharma A. A Comparative Analysis of Machine Learning Algorithms for Classification Purpose. Procedia Comput Sci. Elsevier B.V. 2022 Jan 1; 215: 422-31. https://doi.org/10.1016/j.procs.2022.12.044
29. Vahid M. Ehsan Soleimanfar Mehrzad Navabakhsh Identification of Patterns and Factors Affecting the Health of Employees Based on Datamining of Occupational Examinations with the Purpose of Promoting Occupational Health. Iran J Health Educ Health Promot. 2019; 7(3): 295-305. https://doi.org/10.30699/ijhehp.7.3.295
30. Sarker IH. Machine Learning: Algorithms, Real-World Applications and Research Directions, SN Comput Sci. Springer.2021 May; 2(3):160. https://doi.org/10.1007/s42979-021-00592-x
31. Maity T, Balachandran AK, Krishnamurthy LP, Nagar KL, Upadhyayula RS, Sengupta S, et al. Data-Driven Approaches to Predict Dendrimer Cytotoxicity. ACS Omega 2024; 23: 24899–24906. https://doi.org/10.1021/acsomega.4c01775

Published Online First: October, 2024
https://doi.org/10.21123/bsj.2024.10835
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

32. Syriopoulos PK, Kalampalikis NG, Kotsiantis SB, Vrahatis MN. kNN Classification: a review. Ann Math Artif Intell. 2023 Sep; 1: 1-33; https://doi.org/10.1007/s10472-023-09882-x
33. Ali N, Neagu D, Trundle P. Evaluation of k-nearest neighbour classifier performance for heterogeneous data sets. SN Appl Sci. 2019 Dec 1; 1(12): 1-5. https://doi.org/10.1007/s42452-019-1356-9
34. Lal U, Chikkankod AV, Longo L. Fractal dimensions and machine learning for detection of Parkinson's disease in resting-state electroencephalography. Neural Comput Appl. 2024 May 1; 36(15): 8257–80. https://doi.org/10.1007/s00521-024-09521-4
35. M R, R L. Multiocular disease detection using a generic framework based on handcrafted and deep learned feature analysis. Intell Syst Appl. 2023 Feb 1; 17: 200184. https://doi.org/10.1016/j.iswa.2023.200184
36. Swain M, Kisan S, Chatterjee JM, Supramaniam M, Mohanty SN, Jhanjhi NZ, et al. Hybridized Machine Learning based Fractal Analysis Techniques for Breast Cancer Classification. Int J Adv Comput Sci Appl. 2020; 11(10): 179-84.https://doi.org/10.14569/IJACSA.2020.0111024
37. Battalapalli D, Vidyadharan S, Prabhakar Rao BVVSN, Yogeeswari P, Kesavadas C, Rajagopalan V. Fractal dimension: analyzing its potential as a neuroimaging biomarker for brain tumor diagnosis using machine learning. Front Physiol. 2023 Jul 17; 14 :1201617. https://doi.org/10.3389/fphys.2023.1201617
38. Rebinth A, Kumar SM, Kumanan T, Varaprasad G. Glaucoma Image Classification Using Entropy Feature and Maximum Likelihood Classifier. In: J Phys Conf Ser. 2021 Jul 1; 1964(4) :042075. https://doi.org/10.1088/1742-6596/1964/4/042075
39. Ortiz-Vilchis P, Ramirez-Arellano A. An Entropy-Based Measure of Complexity: An Application in Lung-Damage. Entropy. 2022 Aug 14; 24(8): 1119. https://doi.org/10.3390/e24081119
40. Renjini A, Swapna MNS, Raj V, Sreejyothi S, Sankararaman SI. Fractal and Time-series A nalyses based Rhonchi and Bronchial Auscultation: A Machine Learning Approach. Indian J Sci Technol. 2022 Jun 5; 15(21): 1041–51. https://doi.org/10.17485/IJST/v15i21.627
41. Sharma U, Sood M, Puthooran E, Kumar Y. A block-based arithmetic entropy encoding scheme for medical images Int. J Healthc Inf Syst Inform. 2020 Jul 1; 15(3): 65–81. https://doi.org/10.4018/IJHISI.2020070104
42. Hellman Z, Peretz R. A survey on entropy and economic behaviour. Entropy. 2020 Jan 29; 22(2): 157, https://doi.org/10.3390/e22020157
43. Sparavigna AC. Entropy in image analysis, Entropy. MDPI AG. 2019 May 17; 21(5): 502. https://doi.org/10.3390/e21050502
44. Taha TB, Nurtayeva T, Arif SA, Jamal AS. Partial Differential Equations and Digital Image Processing: A Review. In: 8th IEC 2022 - International Engineering Conference: Towards Engineering Innovations and Sustainability. Institute of Electrical and Electronics Engineers Inc. 2022; 235–40. https://doi.org/10.1109/IEC54822.2022.9807553
45. Jeon G, Chehri A. Entropy-based algorithms for signal processing, Entropy. MDPI AG. 2020 Jun 4; 22(6): 621. https://doi.org/10.3390/E22060621
46. Hao J, Ho TK. Machine Learning Made Easy: A Review of Scikit-learn Package in Python Programming Language. J Educ Behav Stat. SAGE Publications Inc. 2019 Jun; 44(3): 348-61. https://doi.org/10.3102/1076998619832248
47. Ferreira Júnior PE, Mello VM, Giraldi GA. Image thresholding through nonextensive entropies and long-range correlation. Multimed Tools Appl. 2023 Nov 1; 82(28): 43029–73. https://doi.org/10.1007/s11042-023-14978-x
48. Aljanabi MA, Hussain ZM, Lu SF. An Entropy-Histogram Approach for Image Similarity and Face Recognition. Math Probl Eng. 2018; 2018(1): 9801308. https://doi.org/10.1155/2018/9801308
49. Wu J, Yang S, Gou F, Zhou Z, Xie P, Xu N, et al. Intelligent Segmentation Medical Assistance System for MRI Images of Osteosarcoma in Developing Countries. Comput Math Methods Med. 2022; 2022(1): 7703583. https://doi.org/10.1155/2022/7703583
50. Delgado-Bonal A, Marshak A. Approximate entropy and sample entropy: A comprehensive tutorial. Entropy. MDPI AG. 2019 May 28; 21(6): pages= {541}, https://doi.org/10.3390/e21060541
51. Cannon-AmericanMathematicalMonthly-1984; 91(9), p. ebi. https://doi.org/10.1080/00029890.1984.11971488
52. Al-Kadi OS, Watson D. Texture analysis of aggressive and nonaggressive lung tumor CE CT images. IEEE Trans Biomed Eng. 2008 Jul; 55(7): 1822–30. https://doi.org/10.1109/TBME.2008.919735
53. Khalili M, Karamouzian M, Nasiri N, Javadi S, Mirzazadeh A, Sharifi H. Epidemiological Characteristics of COVID-19: A Systematic Review and Meta-Analysis. Epidemiol Infect. 2020 Jan; 148: e130; https://doi.org/10.1017/S0950268820001430
54. Waliszewski P. The quantitative criteria based on the fractal dimensions, entropy, and lacunarity for the spatial distribution of cancer cell nuclei enable identification of low or high aggressive prostate carcinomas. Front Physiol. 2016 Feb 11; 7:34, https://doi.org/10.3389/fphys.2016.00034
55. Wang H, Lei Z, Zhang X, Zhou B, Peng J. A review of deep learning for renewable energy forecasting, Energy Convers Manag. Elsevier Ltd; 2019 Oct 15; 198: 111799. https://doi.org/10.1016/j.enconman.2019.111799

Published Online First: October, 2024
https://doi.org/10.21123/bsj.2024.10835
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

56. Pinheiro W, Santos D, Santana M, Gomes J, Nematollahi MA, Marefat A, et al. CCTCOVID: COVID-19 detection from chest X-ray images using Compact Convolutional Transformers. Front Public Health. 2023 Feb 27; 11: 1025746, https://doi.org/10.3389/fpubh.2023.1025746
57. Wu Y, Zhou Y, Saveriades G, Agaian S, Noonan JP, Natarajan P. Local Shannon entropy measure with statistical tests for image randomness. Inf Sci (N Y). 2013 Feb 10; 222: 323–42. https://doi.org/10.1016/j.ins.2012.07.049
58. Yang D, Martinez C, Visuña L, Khandhar H, Bhatt C, Carretero J. Detection and analysis of COVID-19 in medical images using deep learning techniques. Sci Rep. 2021 Oct 4; 11(1): 19638. https://doi.org/10.1038/s41598-021-99015-3
59. El Hajjar S, Dornaika F, Abdallah F. Recognizing and detecting COVID-19 in chest X-ray images using constrained multi-view spectral clustering. Prog Artif Intell. 2024 Feb 9; 1-14; https://doi.org/10.1007/s13748-023-00312-x
60. Christie DC. Efficient estimation of directional wave buoy spectra using a reformulated Maximum Shannon Entropy Method: Analysis and comparisons for coastal wave datasets. Appl Ocean Res 2024 Jan 1; 142. https://doi.org/10.1016/j.apor.2023.103830
61. Mohammadi H, Gupta S, Sharma S. A large-scale performance study of entropy-based image thresholding techniques using new SAD metric. Pattern Anal and Appl. 2023 May 1; 26(2): 473–86. https://doi.org/10.1007/s10044-022-01121-z

البعد الكسري وتحليل الإنتروبيا للصور الطبية لتصنيف الأمراض على أساس KNN

كريشنا ماهاباترا، سيلفاكومار آر

مدرسة العلوم المتقدمة، معهد فيلور للتكنولوجيا، فيلور، 632014، تاميل نادو، الهند.

الخلاصة

شهدت دراسة الكشف عن الأمراض المستندة إلى الصور الطبية زيادة ملحوظة في الاهتمام والنجاح باستخدام الأساليب الحسابية. يقترح هذا العمل إطارًا جديدًا للكشف عن الأمراض في مراحل مبكرة من الصور الطبية، باستخدام النمذجة الرياضية والتعلم الآلي. نقدم مقياسين كميين جديدين للكشف عن مرض كوفيد-19 وأمراض الأورام: عدم اليقين في الصورة بناءً على إنتروبيا شانون وتعقيد الصورة بناءً على البعد الكسري. أظهرت نتائجنا أنه في الصور الإيجابية لـCOVID-19 أظهرت توزيعًا منحرفًا للبكسل بسبب المناطق الضبابية مما أدى إلى انخفاض قيم الإنتروبيا للحالات المريضة مقارنة بالحالات السليمة. أما الكمية الثانية، وهي البعد الكسوري، فقد تم قياسها بطريقة العد المربع التي تحدد مدى تعقيد الصورة. تم تطبيق نتائج كلا التقنيتين على تصنيف الصور باستخدام نموذج التعلم الآلي (k-NN (ML)أقرب جار). يوفر هذا الإطار الكامل نهجًا جديدًا وفريدًا لتحديد وتصنيف أنواع مختلفة من الصور بدقة تصنيف تبلغ ≈ 90% لفيروس Covid و≈ 70% للورم. يُظهر عملنا أن الأنتروبيا والأبعاد الكسرية يمكن أن تميز بين كوفيد-19 والمرضى الأصحاء، مما يجعلها واعدة بالتشخيص المبكر. تقدم هذه المخطوطة منهجية جديدة وفعالة حسابياً وقابلة للتفسير لتصنيف الأمراض والتي توفر الكشف عن المرض في مرحلة مبكرة.

الكلمات المفتاحية: البعد الكسري، إنتروبيا شانون، تحليل الصور الطبية، التصوير بالرنين المغناطيسي، الأشعة المقطعية، التحليل الإحصائي، التعلم الآلي، أقرب الجيران.