Medical Image Classification using a Many to Many Relation, Multilayered Fuzzy Systems and AI

KISHORE KUMAR AKULA[1], MAURA MARCUCCI[2,8], ROMAIN JOUFFROY[3], FARZAD ARABIKHAN[4], RAHELEH JAFARI[5], MONICA AKULA[6, *], ALEXANDER GEGOV[4,7]
[1]Statistics eTeachers Group, Royal Statistical Society,
100 Leeward Glenway, Toronto, Ontario, M3C 2Z1,
CANADA

[2]Clinical Epidemiology and Research Centre, Department of Biomedical Sciences,
Humanitas University and IRCCS Humanitas Research Hospital,
Milan,
ITALY

[3]Intensive Care Unit,
Ambroise Paré Hospital– Assistance Publique Hôpitaux,
Paris, 9 avenue Charles De Gaulle, 92100, Boulogne-Billancourt,
Paris,
FRANCE

[4]School of Computing,
University of Portsmouth,
Winston Churchill Ave, South Sea, Portsmouth PO1 2UP, Portsmouth,
UNITED KINGDOM

[5]School of Design,
University of Leeds,
Leeds LS2 9JT,
UNITED KINGDOM

[6]Department of Neuroscience,
McMaster University,
Hamilton, L8S 4L8,
CANADA

[7]English Faculty of Engineering,
Technical University of Sofia, Sofia 1756,
BULGARIA

[8]Department of Health Research Methods, Evidence, and Impact,
McMaster University,
Health Sciences Centre, 2C, 1280 Main Street West Hamilton, L8S 4L8,
CANADA

Abstract: - One of the research gaps in the medical sciences is the study of orphan diseases or rare diseases, due to limited data availability of rare diseases. Our previous study addressed this successfully by developing an Artificial Intelligence (AI)-based medical image classification method using a multilayer fuzzy approach (MFA), for detecting and classifying image abnormalities for large and very small datasets. A fuzzy system is an AI system used to handle imprecise data. There are more than three types of fuzziness in any image data set: 1) due to a projection of a 3D object on a 2D surface, 2) due to the digitalization of the scan, and 3) conversion of the digital image to grayscale, and more. Thus, this was referred to in the previous study as a multilayer

fuzzy system, since fuzziness arises from multiple sources. The method used in MFA involves comparing normal images containing abnormalities with the same kind of image without abnormalities, yielding a similarity measure percentage that, when subtracted from a hundred, reveals the abnormality. However, relying on a single standard image in the MFA reduces efficiency, since images vary in contrast, lighting, and patient demographics, impacting similarity percentages. To mitigate this, the current study focused on developing a more robust medical image classification method than MFA, using a many-to-many relation and a multilayer fuzzy approach (MCM) that employs multiple diverse standard images to compare with the abnormal image. For each abnormal image, the average similarity was calculated across multiple normal images, addressing issues encountered with MFA, and enhancing versatility. In this study, an AI-based method of image analysis automation that utilizes fuzzy systems was applied to a cancer data set for the first time. MCM proved to be highly efficient in detecting the abnormality in all types of images and sample sizes and surpassed the gold standard, the convolutional neural network (CNN), in detecting the abnormality in images from a very small data set. Moreover, MCM detects and classifies abnormality without any training, validation, or testing steps for large and small data sets. Hence, MCM may be used to address one of the research gaps in medicine, which detects, quantifies, and classifies images related to rare diseases with small data sets. This has the potential to assist a physician with early detection, diagnosis, monitoring, and treatment planning of several diseases, especially rare diseases.

Key-Words: - Fuzzy systems, CNN, AI; Image analysis, CT scans, medicine, Confusion matrix.

Received: November 9, 2023. Revised: April 17, 2024. Accepted: June 2, 2024. Published: July 3, 2024.

1 Introduction

Medical images like computed tomography (CT) scans are used by physicians and researchers to understand and diagnose disease, and guide treatments. There are many effective currently available automated image analysis tools to determine the abnormalities in the objects within an image, such as a tumor. However, the minimum number of images to run these tools is hundreds or thousands of images, and some methods require already classified data to train the model. Moreover, when performing image analysis of rare disease data sets, a large number of images may be unavailable.

In order to find the abnormalities in the objects present in images, sophisticated methods are available using the AI concept of deep learning, such as convolutional neural networks (CNNs). However, most of these methods need a bulk number of images that are already classified. Even the methods that require fewer images still need thousands of images. Consequently, these methods are not as helpful for analyzing medical images from rare diseases and diseases with limited available data. Hence, to address this gap, we previously developed an Artificial Intelligence (AI) based medical image classification method using a multilayer fuzzy approach (MFA), [1].

Fuzzy logic is a mathematical framework which deals with unlikelihood and imprecision. The concept of multilayered fuzziness used in MFA, as well as in the current study arises from three sources. The first source of fuzziness is present in the image due to the projection of a three-dimensional object, which is lungs, on a two-dimensional surface, which is a CT scan. The second source of fuzziness occurs in the image due to the digitalization of the image in the form of scans. The third source of fuzziness in the image occurs due to the conversion of the image to grayscale to implement the software used in the MFA study.

A fuzzy set is a set, into which fuzzy logic is incorporated. A fuzzy set has an identification (ID) and its membership, which is the extent to which an element of a set belongs to the set. A fuzzy set takes the following form: {ID, membership}. The fuzzy set in both the MFA method and the current study is {patient's ID, SSI}. The multilayered fuzziness involved in images and in the process of obtaining a similarity index will be propagated to the data set used in this study, which is {ID/serial number, similarity between the normal and abnormal image}.

The overall approach used in our previous study MFA was derived from the cognitive science concept of comparing two images, [1]. In MFA, we compared an image containing objects with abnormalities to a reference image with the same objects but without abnormalities [1], calculating the structural similarity index (SSI), [2]. That is, a part in the first image was compared with the corresponding part in the second image. In this way, the entire first image was compared with the second image to get the SSI. This approach not only involved the identification of abnormalities but also

quantification and classification based on the intensity of the abnormality in the images. This was done by subtracting the quantified similarity score between a normal and abnormal image from 100. This method required minimal training data and time. Achieving accurate results with just 22 images, MFA proves beneficial in medical image analysis, particularly with CT scans or images of rare diseases, saving physicians' time.

1.1 Rational for the MCM Study

In the MFA study, only one standard image was used and compared with the abnormal image to find the abnormality. The main issue with this was biased results in detecting the abnormality in the images. For instance, in classifying CT scans for lung cancer, a high-contrast initial image can lead to misjudgments if later replaced with a lower-contrast image. The SSI depends on the standard image, and using the MFA method, SSI tended to change depending on the contrast, or age of the patient. Thus, there was some bias in the abnormality scores and hence, there was also bias in finding the thresholds. To address this issue, a Robust AI-based Medical Image Classification method using the Many to Many Relation and a Multilayer Fuzzy Approach (MCM) was conceived. A many-to-many relation [3] is a relation between two sets in which every element of the first set is related to every element of the second set. In this method, multiple standard, normal images are compared with each abnormal image in order to determine the abnormality.

In the current study, the algorithm of MFA was changed using a fuzzy many-to-many relation. In MCM, the average score obtained by comparing multiple normal images separately with one abnormal image was used to compute the SSI. The images were then classified based on abnormality. These classification thresholds can be used for any different image data set of the same kind of objects, and the normal image will not cause any bias in the abnormality scores, since multiple normal images were used as a standard. Thus, the MCM robustly improved the detection of abnormality and the efficiency of classification of the abnormality in images.

In the current study, a robust, AI-based image classification method using fuzzy systems was applied for the first time to a lung cancer data set. CT scans acquired to diagnose lung cancer were used to test the MCM method in the application part of the current study, with results showing that the MCM succeeded in detecting and classifying the abnormality in the CT scans more accurately than the MFA. The current study MCM makes our previous MFA study algorithm more robust and it was used to detect, classify, and predict a disease using a relatively smaller data set. In addition, although the MFA worked better than the current gold standard methods, it was slightly subjective to the standard image, and in the current study, the MCM method removed this bias in the MFA. This will be useful for physicians and scientists in finding the abnormalities in images.

Moreover, one of the most important research gaps in medical sciences is rare or orphan diseases, due to limited data availability, which the MCM method addresses, since the minimum size of the data set is more than one normal image and one abnormal image for each stage. Additionally, in this study, it was successfully demonstrated that MCM works with a small data set of 19 images for four classes, whereas the gold standard, CNN, is not as effective with such a small data set. Furthermore, MCM works more efficiently than MFA and works with data sets that are even smaller than the one used in this study. There is evidence, [4], that CNN works well for smaller datasets but not for datasets as small as 19.

2 Problem Formulation

2.1 The Primary Aim

The primary aim is to modify the previous MFA method [1] using the concept of fuzzy many-to-many relations, find abnormalities in images, and classify the images based on the abnormality, so as to improve the accuracy of classification and prediction of abnormalities in similar images.

2.2 The Secondary Aim

The secondary aim is to apply the method in the primary aim to detect lung cancer and classify the medical CT scans based on the severity of the lung cancer. In addition, the performance of MCM will be checked against the performance of MFA and CNN for a small data set of images.

3 Problem Solution

3.1 Materials

3.1.1 Image Data Set used to Develop MCM

As mentioned earlier in the above section, the images picked are a cancer type that may or may not be rare lung cancer types. The purpose is to apply

the MCM method to a data set of few images. Next, the current study method was applied to the spread of lung cancer in the CT scans. The open-source image dataset used in the current study for developing the MCM method consisted of CT scans taken to detect lung cancer, [5]. The file format of the CT scans used in this study is the '.dcm' format, which is the Digital imaging and communications in medicine (DICOM) format (Figure 1). The data sets considered had two random samples of different sizes, [5]. In the current study, the selection of data size was contingent on the nature of the analysis. The first analysis performed was to compare MFA and MCM, and for this, a data set of 367 images with confirmed lung cancer was analyzed. The severity ranged from stage 1 to stage 4. In addition, the 42 images for prediction were also randomly considered from the same domain but from a different image set. Furthermore, all manipulations related to images in the entire study were done in terms of the .dcm format of DICOM images.

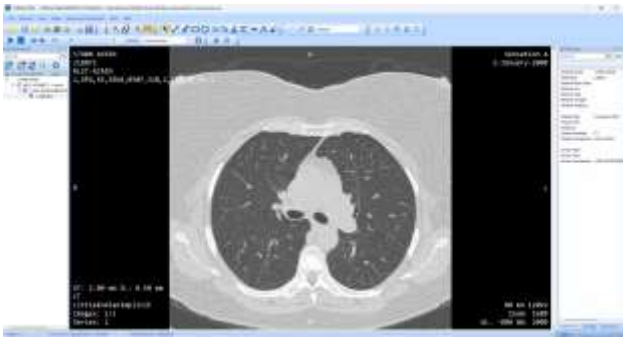

Fig. 1: Sample image of DICOM image of lungs saved in .dcm format (The script on either side of the image was not used in the study)

For the second analysis, MCM was compared with CNN to assess the MCM method for its effectiveness in evaluating a small data set. One of the important characteristics of the current study method is that it works with the smallest possible data. The data was taken has 19 images. To run CNN for training and validation, 13 and 2 images, respectively, were devoted, and for testing, 4 images were devoted. As MCM does not require training, validation, and testing steps, the entire data of 19 images were devoted to the detection of abnormality and classification of the images based on the abnormality. Additionally, for prediction, 35 images were devoted for both of the methods. The .dcm images were used for MCM, and as .dcm images were not detected by CNN, the .png format was used.

In the current study, the spread of lung cancer in the right lung was studied (Figure 2). The right lung was studied separately to avoid noise in the images caused by parts of the image other than the lungs. This is because the noise will influence the structural similarity index (SSI) score, and cause biased detection and classification of abnormality in the images. However, the spread can also be studied simultaneously in both of the lungs. The right lung was extracted by cropping the right lung present in the CT scan to study the spread of the cancer specifically in the right lung.

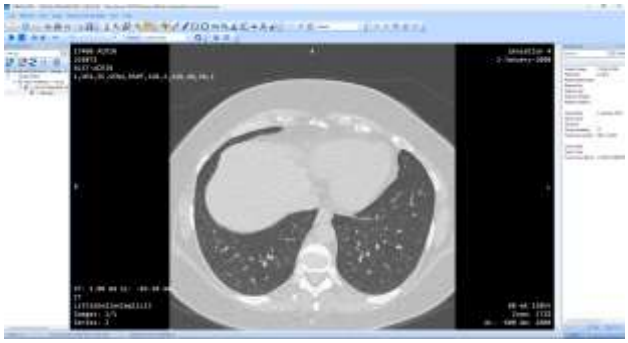

Fig. 2: A sample image of lungs having cancer

3.1.2 Normal Images or Standard Images

The images in which objects have very little or no abnormality are standard images (Figure 3). These were the images with which the abnormal images were compared to find the similarity percentage in the MCM method. An equal number of high-contrast and low-contrast standard images were used, and a total of 60 normal images were considered.

Fig. 3: Normal image of the right lung

3.1.3 Software Used

The software programs used were Python 3.7, and Anaconda 3, with an editor Spyder 5 to run a CNN [6], [7], detect abnormality, compare the images, and classify images as per the abnormality. The visualization of prediction of the CRAN-R software was used.

3.2 Methods

In the current study, when the efficiency of MCM and CNN were compared, the data set considered was very small. Hence, the time taken to run these methods and the memory used to run these methods were negligible and were not analyzed further.

3.2.1 The SSI Metric Used in the Current Study to Compare the Images

The mathematical formula [2] used in the current study MCM to find the SSI is as follows:

$$SSI(N, A) = [(2\mu_N \mu_A + c_1)\cdot(2\sigma_{NA} + c_2)] / [(\mu^2_N + \mu^2_A + c_1)\cdot(\sigma^2_N + \sigma^2_A + c_2)],$$

where N is the normal image, A is the abnormal image being compared, SSI(N, A) stands for the SSI between images, N and A, μ_N is the mean of x; μ_A is the mean of A, σ_N represents the variances of N, σ_A represents the variances of A, and c_1 and c_2 are the weak denominator steadying constants.

The reason why the SSI is specifically used among many such metrics in the MFA and the current study, among many available similarity measure metrics between two images, is that the SSI is a metric for measuring the similarity between two images, with a focus on quantifying the similarity in structure, luminance, and contrast between the reference image and the abnormal image.

3.3 Multilayer Fuzzy Dataset and the Fuzzy Operations on This Set used in the Current MCM Study

3.3.1 Fuzziness

In the current study, fuzziness can be defined as the unclear nature of an image. The fuzzy set is a set of fuzzy objects together with the object's degree of membership. For this study, the fuzzy set is as follows: {ID of the object or patient, abnormality in percentages}.

3.3.2 Multilayer Fuzzy Notion

As introduced in our previous study MFA [1], as well as in the introductory part of the current study, fuzziness arises in an image in numerous ways. One mode is when a three-dimensional object is projected on a two-dimensional exterior. The second mode is through conversion of an image into pixels when it is uploaded to a computer, and the third possible way is by conversion of the digital image to a grayscale image. Another possibility is the change of natural colors of the object in the image to digital shades.

3.3.3 Many to Many Relation

Many-to-many relation [2] is a set theory mathematical concept in which the objects of one set are related or compared with all of the objects of another set with a certain relation among them (Figure 4). A similar operation exists even if the sets are fuzzy sets, which was used in the current study to compare multiple normal images with each abnormal image.

3.4 The Stages of Cancer for Classification to Develop the MCM Method

Clinical staging of lung cancer as performed by healthcare professionals is slightly different from the classification done for the purposes of developing the MCM method in the current study. In the clinical staging of lung cancer, CT images of the liver, bones, and other surrounding organs are also taken into consideration to classify the severity of the spread of cancer, whereas, in the current study, the CT scans consist only of lung images. Thus, the classification of disease here is based on the spread of cancer as seen on the CT scan with the focus on only the lungs. The classification of cancer in this study provides a rough estimation of the spread or stages of lung cancer to the physician for a greater number of images within a short span of time.

3.5 The Threshold of Classification of Abnormality in Terms of the SSI

For developing the MCM method, each of the stages is taken as the following:

Stage 1 is a mild abnormality as seen in the image,
Stage 2 is a moderate abnormality,
Stage 3 is a severe abnormality, and
Stage 4 is a very severe abnormality.

3.6 The CNN

A kind of deep learning model known as a CNN, [6], [7], is utilized for the processing and analysis of visual data. It can be described by multiple layers, including convolutional layers, pooling layers, and fully connected layers. Filters are applied to input data in the convolutional layers, enabling the network to autonomously learn hierarchical features. Spatial dimensions are reduced through pooling layers, and final predictions are made by fully connected layers. CNNs have demonstrated success in diverse applications, such as computer vision, medical image analysis, and natural language processing. A comparative analysis was conducted between MCM and the gold standard, CNN. It is

known that CNN can operate effectively with a minimal amount of data, typically ranging from a few hundred to a few thousand data points [4].

3.7 Method used in the Current Study MCM to Find the Abnormality in Images

The method used in MCM is similar to the method used in the previous MFA study, except that in MFA, only one standard image was used to compare with the abnormal image, whereas in the current MCM method, multiple standard images were used to compare them with the image with abnormalities. Additionally, all the SSI scores, obtained when multiple standard images were compared with one abnormal image, were averaged. When subtracted from 100, this averaged SSI score in percentage form provides quantitative information on the abnormality in the image. This process was continued for all the remaining images with abnormalities allocated for this analysis. According to the fuzzy set theory, the method used in MCM described is a fuzzy many-to-many relation (Figure 4). With the introduction of the fuzzy many-to-many relation, the MCM proved to be a more robust method compared with MFA in reducing the bias and improving the reliability of abnormality detection.

3.7.1 Methods to Accomplish the Primary Aim

The primary aim of the current MCM study has two parts. The first part is developing a method to find abnormalities in images using the concept of fuzzy many to many relations, and the second part is to classify images based on abnormalities.

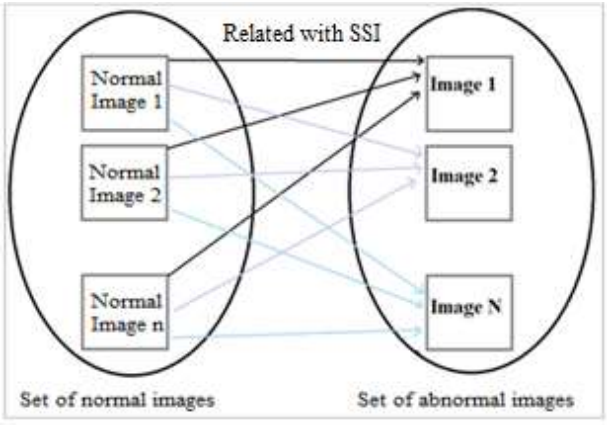

Fig. 4. The method of comparing images in the MCM method using the concept of the fuzzy many-to-many relation.

Firstly, to quantify the abnormalities in images using the fuzzy many-to-many relation, N (say 20) normal images and one abnormal image were considered, then these N normal images were compared with the abnormal image to get N number of structural similarity indices (SSI).

The average of all these SSIs was computed to obtain the SSI of the abnormal image. This process was continued for all the abnormal images, as shown in Figure 4.

The SSI of N normal images, when compared with the first abnormal image, the second, third, and so on for N fuzzy sets are as follows:

SSI_1 = average of $\{S_{11}, S_{12} \ldots S_{1N}\}$.
SSI_2 = average of $\{S_{21}, S_{21} \ldots S_{2N}\}$.
SSI_3 = average of $\{S_{31}, S_{32} \ldots S_{3N}\}$.
…
SSI_N = average of $\{S_{N1}, S_{N2} \ldots S_{NN}\}$,

where S_{ij} is the SSI between the i^{th} abnormal image and with j^{th} normal image. That is, the SSI_N is the average fuzzy similarity of the N^{th} abnormal image in N normal images, and S_{ij} is the fuzzy similarity between the i^{th} abnormal image and the j^{th} normal image. The SSI between the given normal images and the abnormal images is SSI = $\{SSI_1, SSI_2, SSI_3 \ldots SSI_N\}$, where SSI_k is the mean of SSI_k. Hence, the fuzzy set is {Serial number, SSI}, where SSIs are the memberships of the fuzzy set.

3.7.2 Thresholds of Classification

The second part of the primary aim was to find the classification thresholds, which were the stages of the abnormality and obtained by using the schema in Figure 5 and manual software testing strategies, [1], until the classification was done correctly. The general classification stages looked like the following:

If SSI_N <= a%, then the abnormality is at stage 1;
If SSI_N >= a% and SSI_N <= b%, then the abnormality is at stage 2;
If SSI_N >= b% and SSI_N <= c%, then the abnormality is at stage 3;
If SSI_N >= c% and, then the abnormality is at stage 4;

where a, b, and c are SSI values and are fuzzy thresholds of the classification representing a specific SSI to be determined using the schema in Figure 5. To determine the values of these thresholds, the following method was followed:

Step 1: A folder was created with the data images.
Step 2: Folders were formed for each stage (These folders were initially empty).

Step 3: Initially, the a% was assumed to be 25% abnormality as seen in the image, the b% was assumed to be 50% abnormality, and the c% was taken to be 75% abnormality.
Step 4: The code was run.
Step 5: The folders were checked for each stage, and the spread of the abnormality in the images was observed.
Step 6: If the folder contained images with varying degrees of spread of abnormality, the thresholds were adjusted accordingly. For instance, if Stage 2 images were classified within the Stage 1 folder, the classification threshold for Stage 1 was modified. This process was completed for each folder, adjusting the thresholds within the software.
Step 7: After the thresholds were adjusted, the code was run and steps 5 and 6 were repeated until the images were classified properly.

The above steps involve a simple manual software testing strategy and a logic on the fuzzy set. That is, simple fuzzy logic was used to find the classification thresholds of the fuzzy set.

3.7.3 Schema used in MCM

The schema in Figure 5 shows the process for the MCM method, wherein normal images that were of a multilayer fuzzy input were compared with abnormal images to find the SSI. Subsequently, all SSIs were stored, and the identity number of each image was added to the SSI to get the fuzzy set. After acquiring the fuzzy set, logic, and intelligence rules were used to classify the SSI score as per the abnormality, and finally the images were classified. If the thresholds of classification were not classifying some images or if they were classifying many images incorrectly, the thresholds were modified as mentioned above, and manual software testing was used.

3.8 Confusion Matrix or Contingency Matrix to Check the Best among MCM and CNN

The format for the confusion matrix, [8], used to analyze the accuracy of MCM versus CNN was done by the confusion matrix, which is shown in Table 1.

An effective method for summarizing how well a classification rule performs is through the use of a confusion matrix. A confusion matrix was deemed the most appropriate statistical tool over other methods since one of the aims of the study was to evaluate if MCM could be better than the CNN for this very small data set or not. Moreover, the MCM and CNN methods were used on a very small dataset, which is 19 images only. This matrix essentially provides a breakdown of how the predicted classes align with the true classes for a group of objects that the rule has categorized. This mathematical tool will be used in the current study to compare the efficacy of MCM and CNN using a small data set.

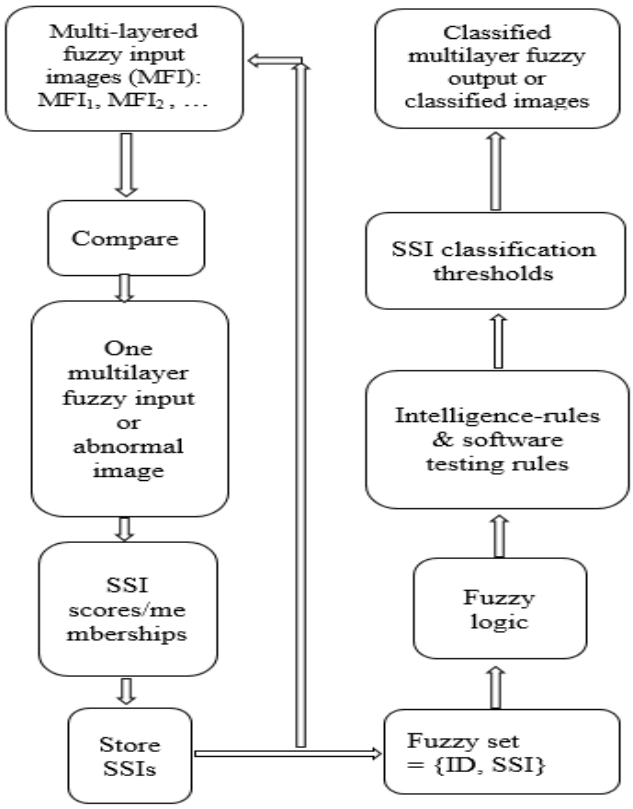

Fig. 5: The schema used in the current study

Table 1. Format for the confusion matrix used for the prediction of stages of lung cancer using MCM and CNN

	PC[a] 1	PC 2	PC 3	PC 4
AC[b] 1	TP[c] 1	FN[d] 2	FN 3	FN 4
AC 2	FP[e] 1	TP 2	FN 4	FN 5
AC 3	FP 2	FP 3	TP 3	FN 6
AC 4	FP 4	FP 5	FP 6	TP 4

[a] predicted class, [b]actual class, [c]true positive, [d]false negative, and [e]false positive

3.8.1 Metrics to Analyze the Confusion Matrix for MCM and CNN for a Small Data Set

The following simple mathematics and statistics metrics, [9], [10] were used to analyze the confusion matrix to decide which of the methods, either MCM from the current study or the gold standard CNN method, is effective for small data.

1. Accuracy: accuracy was measured as the overall correctness of the model's predictions, calculated as

the ratio of correctly predicted instances to the total number of instances.

Accuracy = (Number of correct predictions) / (Total number of predictions)

2. Precision: precision, alternatively recognized as Positive Predictive Value, was measured for the accuracy of positive predictions, and was defined as the ratio of true positives to the total number of positive predictions.

Precision = (True positives) / (True positives + False positives)

3. Recall: recall assesses how proficiently the model is able to recognize the instances in which the attribute being evaluated is present. It can also be defined using the following formula:

Recall = number of true positive values / sum of the number of true positive values and the number of false negative values

4. F1-Score takes both precision and recall into account in a balanced way, and is the harmonic mean of these two variables.

F1-Score = (2x Precision x Recall) / (Precision + Recall)

5. Specificity evaluates how well the model recognizes cases in which the characteristic being assessed does not occur. It can also be defined using the formula given below:

Specificity = number of true negative values / sum of the number of true negative values and number of false positive values

6. False positive rate evaluates the fraction of cases without the characteristic being assessed that are inaccurately categorized as having the characteristic.

False positive rate = false positive values / sum of the number of false positive values and number of true negative values

7. False negative rate:
False negative rate = number of false negative values / sum of number of false negatives and number of true positive values

4 Results

The application of the main objective discussed in section 1 and the methodology presented in section 3 was applied to the CT scan data set taken to detect lung cancer. Specifically, in this study, the right lung was arbitrarily chosen to detect lung cancer. The right lung was cropped from the CT scan to find the abnormality and the spread of the cancer, and to classify the cancer in the right lung.

4.1 Results for MCM

4.1.1 Detection and Classification of the Cancer and Thresholds of Classification by the MCM

The methodology explained in the previous section, which is based on the MFA method [1], was used to find the SSI among normal images and abnormal images using the fuzzy many-to-many relation (Figure 4). This process was continued until all the abnormal images were exhausted. Upon obtaining all the SSI as described above, the schema of the study was used (Figure 5) to detect and classify the CT scans. After using the method described in section 3, the classification thresholds of fuzzy abnormality on the basis of the SSI using the CT scans of the right lung were set as follows:

If SSI >= 0.884, this indicates that the spread of the cancer is Stage 1,
If SSI >= 0.80 and SSI <= 0.884, this indicates that the spread of the cancer is Stage 2,
If SSI >= 0.57 and SSI <= 0.80, this indicates that the spread of the cancer is Stage 3, and
If SSI <= 0.57, this indicates that the spread of cancer is Stage 4.

Using these fuzzy thresholds and a different data set not used for these thresholds, the MCM model was next tested by making predictions.

4.1.2 Prediction by MCM

To make predictions using the MCM method, the standard images for prediction were the same as the images used when the thresholds of classification were set. This is because standard images include a variety of different types, and so images with varying contrasts, as well as images from different age groups were covered. Moreover, the prediction data was used with the same classification thresholds obtained in section 4.1.1, which were obtained by using the data from the study. The results for the prediction by MCM are presented in Table 2.

Table 2. Prediction of identification and classification of the abnormality by MCM

Stages	Stage1	Stage2	Stage3	Stage4
MCMc	0(0)	2(100)	15(94)	22 (100)
MCMw	0(0)	0(0)	1(6)	0
Total	0	2	16	22

c correctly classified, w wrongly classified,
**Numbers in parentheses are percentages.*

Next, as part of the secondary aim, MCM was compared with the MFA method. In the following section, the required computations related to the MFA were performed.

4.2 Results for MFA

In this section, the calculations needed to compare the MCM method with the MFA were performed in order to check whether the MCM is more efficient than the MFA. The major calculations were as follows: detection and classification of cancer using a high contrast or dark standard image, but prediction with a light or low contrast image (MFA_c), and detection and classification of cancer using a low contrast or light standard image but prediction with a high contrast image (MFA_l). The aforementioned two types of detection and classification were compared with MCM, to help determine how using only a single standard image influences the SSI score.

4.2.1 The Thresholds of Detection and Classification of the Cancer by MFA_c

The fuzzy classification and the fuzzy thresholds of classification for MFA_c are as follows:

If SSI >= 0.89, then the spread of the cancer is at Stage 1,
If SSI >= 0.83 and SSI <= 0.89, the spread of cancer is at Stage 2,
If SSI >= 0.605 and SSI <= 0.83, the spread of cancer is at Stage 3,
If SSI <= 0.605, the spread of cancer is at Stage 4.

4.2.2 Prediction by MFA_c

In section 3.3.1, the thresholds of classification of abnormalities were obtained by using dark or high contrast standard images. Next, predictions were made using light or low contrast standard images. The predictions were done by using the prediction data mentioned in the previous sections which can be seen in Table 3.

Table 3. Prediction with the MFA_c method.

Stages	Stage1	Stage2	Stage3	Stage4
MFAc	1(100)	2(67)	9(64)	17(77)
MFAw	0(0)	1(33)	6(36)	5(23)
Total	1	3	15	22

[c] Correctly classified, [w]Wrongly classified

4.2.3 The Thresholds of Detection and Classification of the Cancer by MFA_l

The classification and the fuzzy threshold of classification using the MFA method using a low-contrast image of the right lung as the standard image are as follows:

If SSI >= 0.898, the spread of the cancer is at Stage 1,
If SSI >= 0.85 and SSI <= 0.898, the spread of the cancer is at Stage 2,
If SSI >= 0.55 and SSI <= 0.85, the spread of the cancer is at Stage 3,
If SSI <= 0.55, the spread of cancer is at Stage 4.

4.2.4 Prediction of MFA_l

In section 4.2.3, the thresholds of classification of abnormalities were obtained by using a low-contrast standard image. The prediction was then done using a low-contrast standard image, which is given in Table 4.

Table 4. Prediction by MFA_l

Stages	Stage1	Stage2	Stage3	Stage4
MFA_lc	3(100)	3(75)	26(79)	7(100)
MFA_lw	0(0)	1(25)	7(21)	0(0)
Total	3	4	33	7

[c] correctly classified, wrongly classified

4.3 Results for Comparing the Efficacy of Prediction by MCM versus MFA (MFA_c, and MFA_l)

4.3.1 Comparing All Stages of MCM, MFA_c, and MFA_l

The stages of MCM, MFA_c, and MFA_l are shown in Figure 6 for the correct predictions. The classification done by MCM, MFA_c, and MFA_l shows a clear difference (Figure 6), and just changing the normal image affected the prediction. Although the predictions are borderline correct, the prediction is influenced by the type of standard image used, like whether it is a high-contrast or low-contrast image. Hence, the aim of this study was to

eliminate the subjective nature of the analysis due to the use of a single standard image in order to make it more robust.

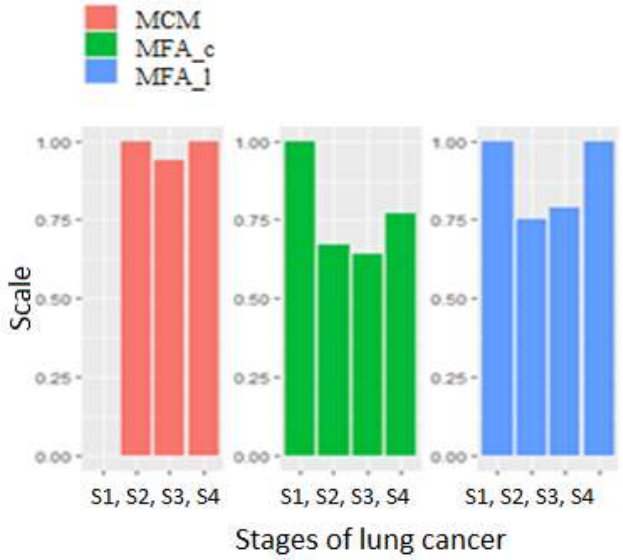

Fig. 6: Comparison of the percentages of correctly predicted values at each stage of the MCM, MFA_c, and MFA_l methods – where S1, S2, S3, and S4 are stages of lung cancer

4.3.2 Stage-wise Comparison of Prediction by MCM versus MFA_c, and MFA_l

Figure 7 shows the stage-wise comparison at stage 1 for MCM, MFA_c, and MFA_l. As shown in Table 2, the correctly classified percentages acquired using each method in each stage are shown. There were no images in stage 1 when classified by the MCM method. The images were at the border, so they may fall under stage 1 or stage 2, as there was only a slight difference in the SSI. The MCM was more sensitive was able to grasp the minute change, and correctly classified these images into stage 2. However, the MFA_c and MFA_l methods classified them into stage 1. Both are still correct, as there is only a slight difference in the SSI. However, the maximum amount of correct classification was done by MCA.

Stage 2 images were classified into Stage 1 with a slight difference in the SSI when classified by the MFA_c and MFA_l, whereas the MCM identified the very minor difference that led to the image being classified as Stage 2 (Figure 7). As shown in Figure 7, a greater number of images were correctly classified as stage 2 by the MCM method compared with the MFA_c and MFA_l. The MCM classifies 100% of the images correctly (Table 1), whereas the images correctly classified by MFA and MFA_l were only 67% and 75%, respectively (Table 2 & Table 3).

At stage 3, 94% of images were classified correctly by the MCM method. Table 1, Table 2 and Table 3 and Figure 7 show that 30% more images were incorrectly classified by MFA than by the MCM method. Additionally, 16% more images were incorrectly classified by the MFA_l than the MCM. Importantly, the MCM did not classify any images incorrectly.

At stage 4, the MFA method misclassified 5 more images than the MCM, whereas the MFA_l misclassified 7 more images in this category than the MCM (Table 2, Table 3 and Table 4). These results also demonstrate that the misclassification rate using the MCM method is lower than that of the MFA_c and MFA_l methods (Figure 7). In addition, Figure 6 shows how a change in the standard image affects the misclassification between the MFA and MFA_l methods, and demonstrates that the MCM method is more robust than the MFA method.

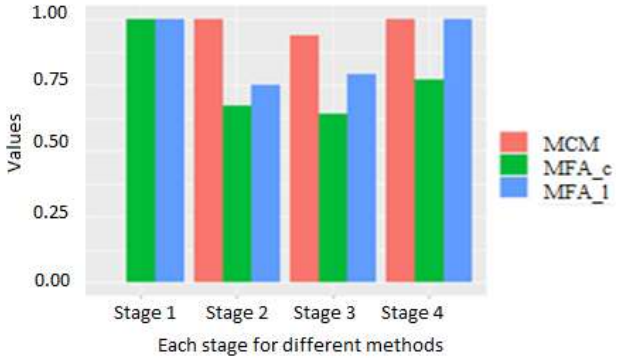

Fig. 7: Stage wise comparison of MCM, MFA_c, and MFA_l

4.4 Predictions using MCM when Classification Thresholds are Obtained with a Small Data Set

4.4.1 Classification of Images using MCM for a Small Data Set

For this section, a small data set was used for the classification. The thresholds obtained after finding SSI scores were as follows:

If SSI > 0.78, the spread of the lung cancer is at Stage 1;
If SSI > 0.70 and SSI ≤ 0.78, the spread of the lung cancer is at Stage 2;
If SSI > 0.49 and SSI ≤ 0.70, the spread of the lung cancer is at Stage 3; and
If SSI < 0.49, the spread of the lung cancer is at Stage 4.

4.4.2 Predictions for MCM using Small Data

For MCM, these thresholds from the previous section were used to predict the images and classify these images based on the abnormality. In addition, the confusion matrix with a graph was constructed (Table 5, and Figure 9). The pictorial representation of the above confusion matrix on a continuous scale is shown in the below graph (Figure 9).

Table 5. Confusion matrix for predictions by MCM for a small data set

Predicted class	True stage			
	Stage1	Stage2	Stage3	Stage4
Stage1	1	0	0	0
Stage2	0	2	0	0
Stage3	0	0	3	0
Stage4	0	0	1	21

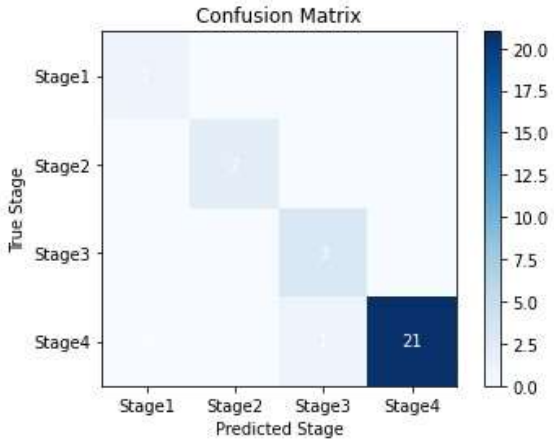

Fig. 8: Pictorial representation of a confusion matrix for MCM for a small data set

4.5 The Classification of Images using CNN for a Small Data Set

4.5.1 The Epochs while CNN was Running for the Small Dataset

The CNN was run for the above-mentioned small data set, followed by manual classification of the images, and the images were then fed into the CNN. A few results were obtained to show how the CNN overfitted the model for the small dataset in Table 5. Accuracies for training validation showed insufficiency of data.

Table 6. Sample epochs for a CNN

Epoch number	Training accuracy	Validation accuracy	Validation loss
1	33.3%	0.0%	1.450
2	100.0%	33.3%	1.401
3	33.3%	0.0%	1.399
4	33.3%	0.0%	1.396

4.5.2 The Predictions for the CNN

The CNN constructed for the above data was used to predict the accuracy and the confusion matrix for the prediction, as shown in Table 7. To analyze the predictions using the CNN, a confusion matrix and its graph were constructed (Table 6 and Figure 8).

Table 7. Confusion matrix for the predictions by CNN for a small data set

Predicted Class	True stage			
	Stage1	Stage2	Stage3	Stage4
Stage1	4	1	0	6
Stage2	8	1	0	1
Stage3	0	3	1	3
Stage4	1	1	4	5

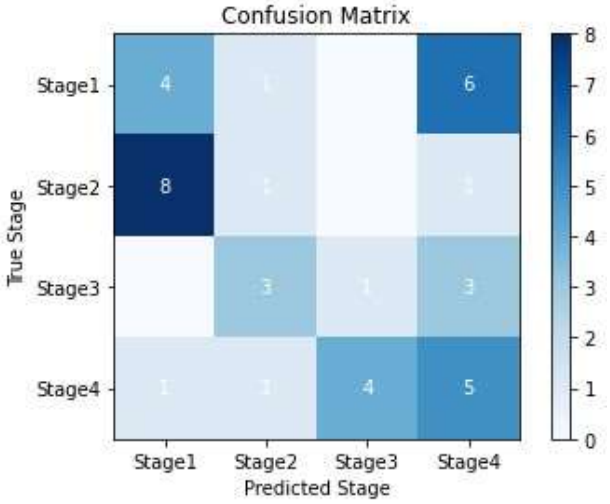

Fig. 9: Pictorial representation of the confusion matrix for CNN for a small data set

4.6 Results for the Detailed Examination of MCM versus CNN for a Small Data Set using Metrics Calculated on the Confusion Matrix

An advantage of MCM is working efficiently with both large and small data sets. It was already established that MCM was robust when compared with MFA in terms of accurately classifying images. In this section, the efficiency of MCM versus the gold standard, CNN, was determined using the following tools discussed in the previous section, as shown in Table 8 and Table 9.

The accuracy of the CNN model in classifying different stages was approximately 48.78%, signifying that around 48.78% of the predictions were correct. Nevertheless, the average accuracy for all stages using MCM was 98.9%. That is, for small data sets MCM is 50.1% more accurate than CNN (Figure 7, Figure 8). Moreover, the highest precision

that the CNN resulted in was for stage 4, which was 33.33%, whereas for MCM the precision was 95.45. For recall, MCM also had better scores. The recall values were high for all stages, indicating that the MCM model effectively identifies most of the positive cases.

The F1 scores for all stages were generally high, suggesting that the MCM model provides a balanced performance between precision and recall (Table 8 and Table 9). The MCM model had a specificity of 100% for all stages of MCM, meaning it correctly identifies negatives all the time, whereas the CNN had the highest value only for stage 1, and it was only 60%. This rate represents the frequency at which the model incorrectly predicts a positive case when it is, in fact, negative. For MCM the false positive rate was 0% for all stages but for CNN, only stage 3 had a minimum false positive rate, which was 7%. The false negative rate represents the rate at which the model incorrectly predicts a negative case when it is actually positive. Even in this regard, the MCM was superior.

Table 8. Various metrics were calculated using the confusion matrix for MCM

Metric	Stage1	Stage2	Stage3	Stage4
Accuracy	1	1	1	0.9545
Precision	1	1	1	0.9545
Recall	1	1	1	0.9545
F1-Score	1	1	1	0.9545
Specificity	1	1	1	1.0000
False positive rate	1	1	1	0.0000
False negative rate	1	1	1	0.0455

Table 9. Various metrics are calculated using the confusion matrix for CNN stage-wise

Metric	Stage1	Stage2	Stage3	Stage4
Accuracy	0.6250	0.5000	0.5714	0.4762
Precision	0.2857	0.1667	0.2000	0.3333
Recall	0.6667	0.0833	0.2000	0.3333
F1-Score	0.4000	0.1111	0.2000	0.3333
Specificity	0.6000	0.1250	0.2308	0.3333
False positive rate	0.1000	0.3750	0.0769	0.4000
False negative rate	0.1667	0.6667	0.0000	0.4615

Table 8 and Table 9 show that MCM performs better than CNN for small data sets. For example, accuracy for precision and recall, up to false positive rates were greater for MCM, demonstrating that the predictions are accurate and robust for MCM compared with CNN for a small data set. In addition to this, the last metric, which is the false negative rate, was very low.

5 Discussion

In the current study, the MCM method was developed as a generalization of the MFA method using fuzzy many-to-many relations to increase the robustness in the classification and prediction of the images with abnormalities. The MCM method was then applied to a medical image data set of CT scans from lung cancer patients. MCM was used for the detection and classification of lung cancer as seen on the CT scans. The purpose of both the MCM and MFA methods is to quantify abnormality in visual form, that is, quantifying an abnormality in a cancer tumor as seen on the CT scan into the form of an SSI score in the case of the currently used lung cancer data set.

MCM did not result in biased classification, whereas MFA sometimes did, possibly because MFA is not as sensitive as MCM in predicting minute abnormalities. Classification thresholds once formed are independent of normal or standard images, whereas thresholds are dependent on the standard image for MFA. In MCM, the SSI was also stable compared to MFA, as many normal images were considered in MCM.

One of the unique features of the MCM is that it also works for a small data set of images. Hence the MCM was compared with the gold standard CNN for a small data set. It is known that CNN works for small data sets; however, the data set must at least contain around a few hundred or thousands of data points [1], [3]. To compare MCM and CNN, only 19 images were taken, of which 13 images were used for training, including two images for validation, and the rest were used for testing. A CNN was run by taking all the precautions to run it successfully for this smallest data set, such as using very clear images.

MCM performed better than CNN with smaller data sets. MCM does not need classified data, but a CNN needs this step most of the time. Writing the code, debugging, and running the program takes very little time for MCM, but takes more time for CNN. Furthermore, MCM uses few software functions, but CNN requires many more than MCM. Moreover, MCM can analyze and quantify the abnormality, whereas CNN cannot quantify the abnormality in the images. For MCM, the minimum data required is more than one normal image and one abnormal image to compare with the normal images. On the other hand, for CNN, it might need a few hundred if the images are of good quality and are clear. For this study, as stated above, the minimum data to run MCM is 4 plus 1, where 4 is the number of normal or standard images and 1 is the number of abnormal images needed. With this

data, the images can be classified and the disease stage predicted. With this data, it is not possible to run a CNN, as the CNN is overfitted for the data.

There are many methods like CNN that cannot perform well with very small data sets and when limited computational power issues exist. However, it was demonstrated in this study using a confusion matrix that MCM works even with a very small data set. MCM can be customized or adapted more easily to the specific characteristics of small datasets. MCM also fits smaller and bigger data sets to find patterns in the data easily and can understand the patterns of the data with a much smaller data sample size like 10, whereas CNN needs bigger data sets to find a pattern. Although the CNN works with smaller datasets, it does not work as effectively for data as small as the one used in this study.

Another important property of MCM is that even if a small data set is used and the classification thresholds are calculated, it can be generalized to a big data set, which is an important data augmentation property. That is, the thresholds of classification obtained by small data could be applied to a large data set to classify images. While CNN can also do this, it cannot do so with smaller data of as few images as 10 or 20. Moreover, MCM can successfully find patterns in very small data sets and work effectively with larger data sets, that is when the same thresholds of classification obtained using a small data set are applied to larger data sets

Domain expertise is needed to work with CNN. For example, this data set is related to the classification of the stages of lung cancer. If CNN is used, then in most cases, the user must know what is stage 1, stage 2, stage 3, and stage 4, because CNN needs manually classified data, so the user has to first classify the data manually, whereas MCM uses a normal image to classify images with abnormalities.

A quantitative comparison of MCM and CNN showed significant differences in performance between MCM and CNN, such that MCM showed better performance than CNN.

The results of this study demonstrate that the MCM was more effective than the MFA for all kinds of data, and for small data sets, the MCM worked better than the CNN. The main limitation of MCM is that many kinds of standard or normal images have to be used. In addition, noise in the images should be removed before using MCM, which was done in this study by cropping the lung images.

6 Conclusion

To conclude, MCM is a generalization of the MFA method, showing that MCM more accurately classifies images. Specifically, MCM is 21% more accurate than MFA_c and 9.5% more accurate than MCA_l. Both the MCM and MFA methods are successful in quantifying the abnormality in an image, such as a cancer tumor. However, the MCM is very sensitive and can catch small changes in the abnormality when compared to the MFA. On the other hand, the MFA method is subjective to the standard image. Both work with a very small data set, so they are useful for studying rare diseases or abnormalities in the form of images. The main problem with the MFA method is that it is based on comparing a single normal image with an abnormal image. That is, if a single normal image is replaced with another, then the classification thresholds will be altered. However, this problem was rectified in this study using multiple normal images for comparison with one abnormal image. Thus, the SSI score and the classification determined using the MCM method were made more robust than with the MFA. Thus, physicians and scientists could use the MCM with confidence to obtain an accurate initial overview of an abnormality or disease in a patient. Furthermore, the MCM method can be used to make accurate predictions for rare diseases or problems with very little data.

When comparing MCM with the gold standard, CNN, for a small data set, the results of all the statistical tools used show that the MCM performed better than CNN. Moreover, MCM accepted DICOM images and conversion to PNG. Hence, some of the ‘fuzziness’ was avoided. The MCM can also be used to detect the abnormality in a small data set with two images. Rare diseases typically do not have a lot of data to train, test, or validate the CNN process. Thus, MCM can be used to detect rare diseases using a limited number of diagnostic images, as the minimum data needed to run the MCM is more than one normal image and one image for each group or stage of abnormality. Additionally, although the MCM method was applied to cancer images in the current study, it could be applied to any image type, like other medical images, or images from any other field of science, such as astronomy and geography.

One of the research gaps that needs more detailed study is rare events, such as rare diseases. These rare diseases have limited data and using traditional tools, it is not possible to study these diseases. However, MCM is designed for both smaller and larger data sets, and the comparison performed between CNN and MCM in the current

study demonstrates that MCM is efficient for very small data sets. Future studies based on the MCM method can be in epidemiology, clinical or medical science, and rare fields of many sciences that have small image data sets. Future studies can also focus on applying the MCM to make a connection between the abnormality in medical images and the risk associated with that abnormality. In future projects, the mortality risk present in the patient will be estimated using the quantified abnormality computed with the MCM method.

References:

[1] Akula K.K; Gegov, A; Arabikhan, F. Artificial Intelligence-Based Medical Image Classification Using a Multilayer Fuzzy Approach, *WSEAS Transactions on Computers*, 2023 vol. 22, pp. 206-217. https://doi.org/10.37394/23205.2023.22.24.

[2] Wang, Z. Multi-scale structural similarity for image quality assessment. *Proceedings of the 37th IEEE Asilomar Conference on Signals, Systems and Computers, Pacific Grove*, CA, 2003. DOI: 10.1109/ACSSC.2003.1292216.

[3] Rodgers, N. Learning to Reason, A. Wiley-Interscience Publication John Wiley & Sons 2000, INC, [Online]. https://www.wiley.com/en-us/Learning+to+Reason%3A+An+Introduction+to+Logic%2C+Sets%2C+and+Relations-p-9781118165706 (Accessed Date: February 11, 2024).

[4] Rhett N. D'souza; Po-Yao Huang; Fang-Cheng Yeh. Structural Analysis and Optimization of Convolutional Neural Networks with a Small Sample Size, *Scientific Reports*, 2020, 10:834. DOI: 10.1038/s41598-020-57866-2.

[5] National Cancer Institute, 2021, [Online], https://www.cancerimagingarchive.net/ (Accessed Date: December 20, 2023).

[6] Sewak, M.; Poojari, P. *Practical Convolutional Neural Networks*, Packt Publishing, Birmingham, England 2018; pp. 46-75, [Online], https://www.packtpub.com/product/practical-convolutional-neural-networks/9781788392303 (Accessed Date: January 10, 2024).

[7] Loy, J. *Neural Network Projects with Python Packt Publishing*, Birmingham, England 2019, [Online]. https://www.packtpub.com/product/neural-network-projects-with-python/9781789138900 (Accessed Date: January 10, 2024).

[8] Simpson, E.H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society*, 1951 Series B 13, 238–241, [Online]. https://www.jstor.org/stable/2984065 (Accessed Date: January 6, 2024).

[9] Armitage, P.; Colton, T. *Encyclopedia of Biostatistics*, John Wiley, 2005. DOI: 10.1002/0470011815.

[10] Rao, C.R. *Handbook of Statistics: Epidemiology and Medical Statistics*, Vol. 27, North-Holland, 2007, [Online]. https://shop.elsevier.com/books/epidemiology-and-medical-statistics/rao/978-0-444-52801-8 (Accessed Date: January 2, 2024).

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

KKA conceived the study, wrote the software code, and prepared and reviewed the manuscript. AG supervised the idea, and the research and reviewed the manuscript. MA reviewed the research related to medical concepts, cognitive processes, language and logic, and the manuscript. RJ and FA reviewed the manuscript. MM and RJ reviewed the medical concepts.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflict of Interest

The authors do not have any conflicts of interest to disclose.