Classification of diabetes maculopathy images using data-adaptive neuro-fuzzy inference classifier

Article *in* Medical & Biological Engineering & Computing · June 2015

DOI: 10.1007/s11517-015-1329-0

7 authors, including:

Pradeep Chowriappa
Louisiana Tech University
20 PUBLICATIONS **237** CITATIONS
SEE PROFILE

U Rajendra Acharya
Ngee Ann, Singapore University of Social Science, Singapore; Asia University Taiw…
634 PUBLICATIONS **20,112** CITATIONS
SEE PROFILE

Kevin P. Noronha
St.Francis Institute Of Technology, Borivili(w),Mumbai
20 PUBLICATIONS **375** CITATIONS
SEE PROFILE

Sulatha V. Bhandary
Kasturba medical college, Manipal. Manipal Academy of Higher education
41 PUBLICATIONS **688** CITATIONS
SEE PROFILE

Med Biol Eng Comput
DOI 10.1007/s11517-015-1329-0

ORIGINAL ARTICLE

Classification of diabetes maculopathy images using data-adaptive neuro-fuzzy inference classifier

Sulaimon Ibrahim[1] · Pradeep Chowriappa[1] · Sumeet Dua[1] · U. Rajendra Acharya[2] · Kevin Noronha[3] · Sulatha Bhandary[4] · Hatwib Mugasa[1]

Received: 7 April 2014 / Accepted: 5 June 2015

Abstract Prolonged diabetes retinopathy leads to diabetes maculopathy, which causes gradual and irreversible loss of vision. It is important for physicians to have a decision system that detects the early symptoms of the disease. This can be achieved by building a classification model using machine learning algorithms. Fuzzy logic classifiers group data elements with a degree of membership in multiple classes by defining membership functions for each attribute. Various methods have been proposed to determine the partitioning of membership functions in a fuzzy logic inference system. A *clustering method* partitions the membership functions by grouping data that have high similarity into clusters, while an *equalized universe method* partitions data into predefined equal clusters. The distribution of each attribute determines its partitioning as fine or coarse. A simple grid partitioning partitions each attribute equally and is therefore not effective in handling varying distribution amongst the attributes. A data-adaptive method uses a data frequency-driven approach to partition each attribute based on the distribution of data in that attribute. A data-adaptive neuro-fuzzy inference system creates corresponding rules for both finely distributed and coarsely distributed attributes. This method produced more useful rules and a more effective classification system. We obtained an overall accuracy of 98.55 %.

Keywords Diabetic retinopathy · Diagnosis · Fuzzy logic · Classification · Image analysis

1 Introduction

Diabetic retinopathy (DR) is a manifestation of diabetes that causes irreversible damage to the retina [15]. Prolonged DR leads to diabetic maculopathy (DM), which can cause permanent loss of the central vision in diabetic patients. Ophthalmologists characterize DR by using features such as hemorrhages, micro-aneurysms, and blood vessels. The objective of this research is to develop a decision system for early detection of DM using ANFIS based on a data-adaptive partitioning approach. The specific aim of the proposed decision system is twofold: firstly, the extraction of textural features from a set of DM fundus images and secondly, to build a model that classifies fundus into three classes, namely normal, clinically significant macular edema (CMSE), and non-clinically significant macular edema (non-CMSE). These classes signify the severity of DM [10].

Decision systems [10, 25] that classifies fundus by the degree of severity are sensitive to two factors, namely the type of features extracted from the fundus images and by the nature of classifiers that are used to build the decision models [24, 31]. They both assumed that the classes (i.e., disease severity) are mutually exclusive. This assumption of class exclusivity may not always hold true, as some image fundus may have a degree of membership in more than one class. We believe that unlike automatic classification, fuzzy classifiers can be designed to be transparent,

✉ Sumeet Dua
sdua@latech.edu

[1] Computer Science, Louisiana Tech University, Nethken Hall 121, 600 Dan Reneau Dr., #10348, Ruston, LA 71272, USA

[2] Department of Electronics & Computer Engineering, Ngee Ann Polytechnic, 535 Clementi Road, Singapore 599489, Singapore

[3] Department of Electronics & Communications, MIT Manipal, Manipal 576104, India

[4] Department of Ophthalmology, KMC Manipal, Manipal 576104, India

Published online: 25 June 2015

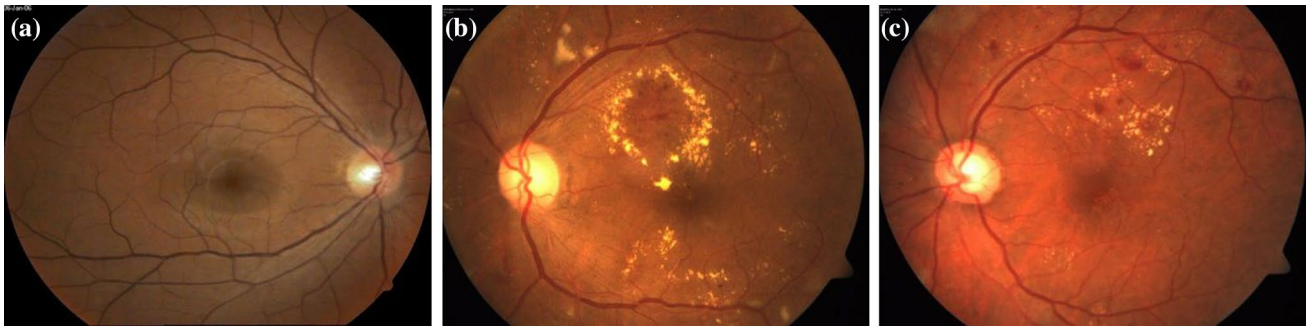

Fig. 1 Typical fundus images **a** normal; **b** CSME; **c** non-CSME

with classification steps and logic statements that are traceable and comprehensible. We therefore propose a novel data-adaptive neuro-fuzzy rule-based classifier to test this belief.

Related works [2, 14] in fundus classification rely on novel features extraction techniques to classify images to their corresponding classes. These features describe morphological characteristics of the fundus image such as cup-to-disk (c/d) ratios [14], diameter of the optic disk, the distance between the optic disk center and optic nerve head (ONH), and ratio of area of blood vessels in inferior–superior side to the area of blood vessels in the nasal–temporal side [25].

In recent years, textural features have gained prominence in fundus image classification. Enhancement to the extraction of textural features for classification is described in the work by Acharya et al. [2]. They proposed the use of texture and higher-order spectra features for the classification of glaucoma using fundus images. They classified fundus images into normal and glaucoma classes and achieved a classification accuracy of 91 %. Similarly, the work by Dua et al. [14] also proposed the use of wavelet-based energy features on fundus images to classify glaucomatous images. They reported an overall accuracy of 93.33 %.

In the literature, we find many computer-aided diagnosis techniques (CAD), which are based on feature extraction and classification [24]. As observed, Nayak et al. [25] used textural features in conjunction with a feed-forward artificial neural network (ANN) for the classification of DM. They reported a sensitivity of 95.4 % and a specificity of 100 %. Similarly, the work presented by Chowriappa et al. [10] proposed an ensemble selection technique to build an ensemble classifier from a library of models. They reported an average accuracy of 96.7 %, with highest recorded accuracy of 97.8 %.

Fuzzy rule-based classifiers have gained considerable interest in decision support systems because of their ease of interpretability [19]. Typically, frameworks for fuzzy classifiers consist of fuzzification techniques, inference systems, and defuzzification techniques [33]. The work by Jang et al. [21] suggested the use of ANFIS. It uses the fuzzy modeling procedure, whereby it learns information about a dataset from the dataset itself. Here, the membership functions are determined from the given set of features, and rules are generated by adjusting the weights using both the forward pass and back-propagation algorithms.

In this work, we intend to create a data-ANFIS that works around known limitations of existing neuro-fuzzy inference systems. We believe that the way fuzzy membership is computed by traditional approaches overlooks the inherent distribution of data by assuming that the data are normally distributed. We therefore hypothesize that with the inclusion of a data-adaptive partitioning during the fuzzification process we can greatly boost the performance of fuzzy classifiers. We adopt a data-adaptive partitioning method [20] to partition each extracted feature based on its underlying distribution. The approach uses a data frequency-based partitioning to generate peaks and valleys.

2 Methods

2.1 Image preprocessing

The fundus images of patients with DM can be categorized into four classes based on the severity of the disease: normal, mild, moderate, and severe. Cases of mild and moderate maculopathy are classified further as non-CSME, and cases of severe maculopathy are categorized as CSME (refer to Fig. 1). Here, the image dataset consists of three hundred retinal images of normal eyes (100 images), retinal images of eyes affected by non-CSME (100 images), and retinal images of eyes affected by CSME (100 images), respectively. The images from each class were obtained from the Kasturba Medical Hospital, Manipal, India. The images were stored in a 24-bit JPEG lossless format with an image size of 720 × 576 pixels. Each image in the set was subject to the following image preprocessing steps:

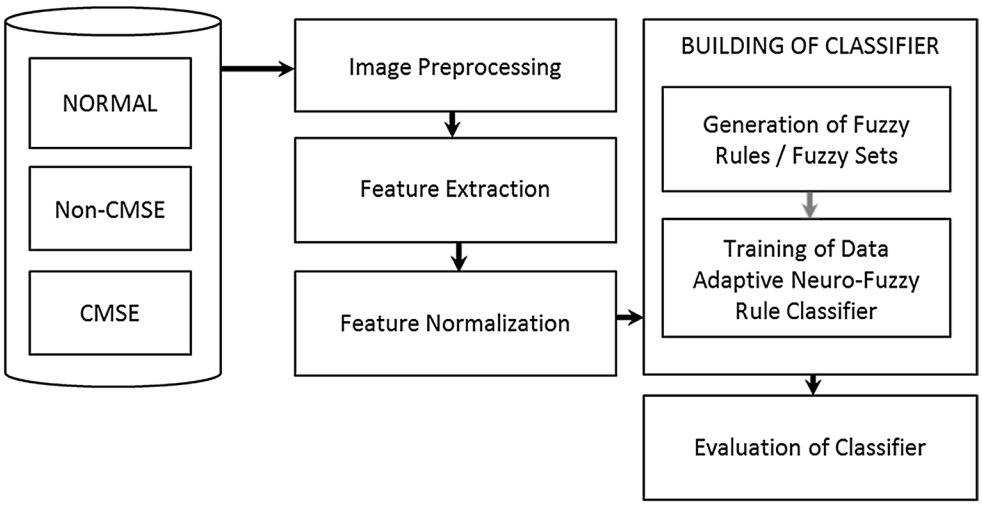

Fig. 2 Schematic overview of the proposed early detection DM as a decision system

(1) color normalization where the RGB image is converted into the habitat suitability index (HSI) model for contrast enhancement. (2) We use the median filter to eliminate noise. (3) We use histogram specification to remove biases caused by the skin tone of the subject [25]. (4) We then convert the image into grayscale.

2.2 The proposed early detection system of DM

The proposed data-ANFIS-based classification of DM image fundus aims to improve the accuracy and interpretability of an early detection of DM. In this section, we describe the methodology adopted that is captured in the schematic overview shown in Fig. 2. The following sections describe the feature extraction strategies and details of building the classifier. The proposed system was executed in two steps, firstly by dividing the dataset into training and testing data and secondly by using a *k*-fold cross-validation method. In the first experiment, the system was trained using 70 % of the data provided. Here, samples for the training purposes were chosen at random and the remaining 30 % used for testing the system. The training data have known class labels. To test the system, the class labels of the test samples were predicted by the model. The second experiment uses a tenfold cross-validation where 90 % of the dataset was randomly divided from the original dataset and used as training data and 10 % as testing data. The results obtained here will be compared to non-fuzzy classification approaches and recent works in the area. The schematic of the proposed early detection DM decision system is seen in Fig. 2 below.

2.2.1 Textural feature extraction

The textural features of the DM fundus image were extracted using the following approaches from related literatures. Textural features are useful in capturing morphological features that are reflected as nonlinear changes in texture in fundus images. These nonlinear changes are captured as suitable texture features that can quantify the changes in the intensity, regularity, coarseness, contrast, homogeneity, etc. of pixels of the image. Traditional texture descriptors are local binary pattern (LBP), laws texture energy (LTE), entropies, and Hu's invariant moments [3]. In this work, we employed two popular techniques, namely discrete wavelet transform (DWT) [23] and the Fourier spectrum [9]. We then extracted the entropy coefficients from these descriptors and used these features to classify images.

2.2.1.1 Discrete wavelet transform (DWT) DWT basically acts as low-pass filter and high-pass filter. When an image of size (M × N) passes through DWT, it yields four different coefficients. Approximate coefficients of level 1 (A_1) are obtained by applying a low-pass filter to both the horizontal row and vertical columns of pixels. Detailed horizontal coefficients of level 1 (Dh_1) are obtained by applying low-pass filter to the horizontal rows of pixels and high-pass filter to the vertical columns of pixels. Detailed vertical coefficients of level 1 (Dv_1) are obtained due to high-pass filter to the horizontal rows of pixels and low-pass filter to the vertical columns of pixels. Finally, detailed diagonal coefficients of level 1 (Dd_1) are due to high-pass filter to both horizontal rows and vertical columns of pixels. Similarly, Dh_2, Dv_2, Dd_2, and A_2 are the resultant matrices of second level of 2D DWT.

The wavelet basis bi-orthogonal wavelet 3.7 is selected to decompose the fundus images. The energy values in the various sub-bands are used as features as given in Eq. 1.

$$\text{Energy}_{\text{sub-band}} = \frac{1}{M \times N} \sum_{x=\{M\}} \sum_{y=\{N\}} \left(D_{x,y}^{\text{sub-band}} \right)^2 \quad (1)$$

where $D_{x,y}^{\text{sub-band}}$ are various sub-band matrices obtained after first and second level of wavelet decomposition, $M \times N$ is the size of matrix.

2.2.1.2 Texture from Fourier spectrum In any image (x_i, y_i) where $i = 1, 2, \ldots, K$ represents the edge points of an object, Fourier descriptors of that edge can be represented by the following approach [9]. Each point can be treated as a complex number such that

$$s(k) = x_i + jy_i \tag{2}$$

Now, the discrete Fourier transform (DFT) of $s(k)$ is

$$a(u) = \sum_{k=0}^{k-1} s(k) e^{-j2uk/K} \tag{3}$$

If the length of DFT of any sequence is the same as an original sequence, the number of the descriptors varies as the length of the edge changes. Here, the power coefficient of Fourier descriptors is computed as follows

$$\text{PAC} = \sum_{\substack{u \neq 0 \\ v \neq 0}} \left(F_R^2(u,v) + F_1^2(u,v) \right) \tag{4}$$

where $F_R(u, v)$ and $F_1(u, v)$ are real and imaginary segments of the Fourier transforms of the image, and u and v are the frequencies along the x and y axes of the image, respectively. Fourier descriptors are not invariants to scaling and translation.

2.2.1.3 Entropy The measure of entropy is a measure of uncertainties in an image. Let $f(x, y)$ be the image with $P_i (i = 0, 1, 2, 3, \ldots, R - 1)$ various gray levels. The normalized histogram C_i for a fundus image of size (A × B) is given by

$$C_i = \frac{P_i}{A \times B} \tag{5}$$

Shannon entropy:

$$S_e = -\sum_{i=0}^{R-1} C_i \log_2 (C_i) \tag{6}$$

where R represents number of gray levels present in the given image. Renyi entropy [26]:

$$R_e = \frac{1}{1-\alpha} \log_2 \sum_{i=0}^{R-1} C_i^{\alpha} \quad \textit{where } \alpha \neq 1, \alpha > 0 \tag{7}$$

where α is diversity index. Here, $\alpha = 3$. Kapur entropy [4, 26]:

$$K_e = \frac{1}{\beta - \alpha} \log_2 \frac{\sum_{i=0}^{R-1} C_i^{\alpha}}{\sum_{i=0}^{R-1} C_i^{\beta}} \quad \textit{where } \alpha \neq \beta, \alpha > 0, \beta > 0 \tag{8}$$

For $\alpha \neq 1$; $\beta > 0$; $\alpha + \beta - 1 > 0$, where α and β are diversity indices. Here, $\alpha = 0{:}5$ and $\beta = 0{:}7$ Yager's measure [4, 26]:

$$Y_e = 1 - \frac{\sum_{i=0}^{R-1} |2C_i - 1|}{|A \times B|} \tag{9}$$

where A and B are rows and columns of the given image, respectively.

2.2.2 Data normalization and transformation

Normalization is a process, whereby the features are scaled so as to fall within an equal range. Due to the varying range of the features, in this work we employed the min–max normalization with predetermined range of minimum of zero and maximum of one. Min–max normalization performs a linear transformation on the features. The normalization data value v_i' of value v_i of an attribute A from range [min$A$, max$A$] to a new range [$new_{\text{min}A}$, $new_{\text{max}A}$] is computed as:

$$v_i' = \frac{v_i - \text{min}A}{\text{max}A - \text{min}A} \left(new_{\text{max}A} - new_{\text{min}A} \right) + new_{\text{min}A} \tag{10}$$

where v_i' is the new value in the required range.

We then subject the normalized features to the log transformation. The log transformation is used to make highly skewed data distributions less skewed. It can be computed as:

$$V_i = \log \left(v_i' \right) \tag{11}$$

2.2.3 Building the classifier

The building of the classifier was carried out in two phases, namely training and testing. We used 70 % of the dataset for construction of fuzzy rule learner as part of the training phase. The membership function for the data representing the three classes normal, non-CSME, and CSME was generated as part of the training phase. The remaining 30 % data were then used to test the model. We then used a tenfold cross-validation method with 90 % of the dataset used for the training phase and 10 % for the training phase. A bootstrapping process was also used to select a random subset from the original data with replacement. This subset was further subdivided into training and testing data, and the process was repeated 10 times. In this section, we describe the key concepts employed for the creation of the classification model.

First introduced by Zadeh [33], fuzzy set is an extension of classical notion of sets. In this work, fuzzy sets refer to sets whose elements have degrees of membership. These sets are used to represent and manipulate data and information in which there are alternative uncertainties. In classical

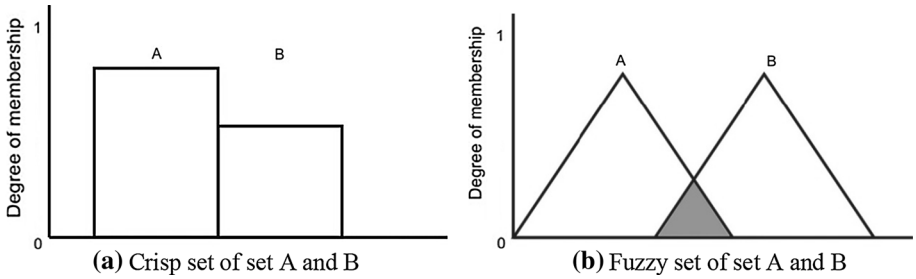

Fig. 3 **a** Crisp set of set *A* and *B*; **b** Fuzzy set of set *A* and *B*

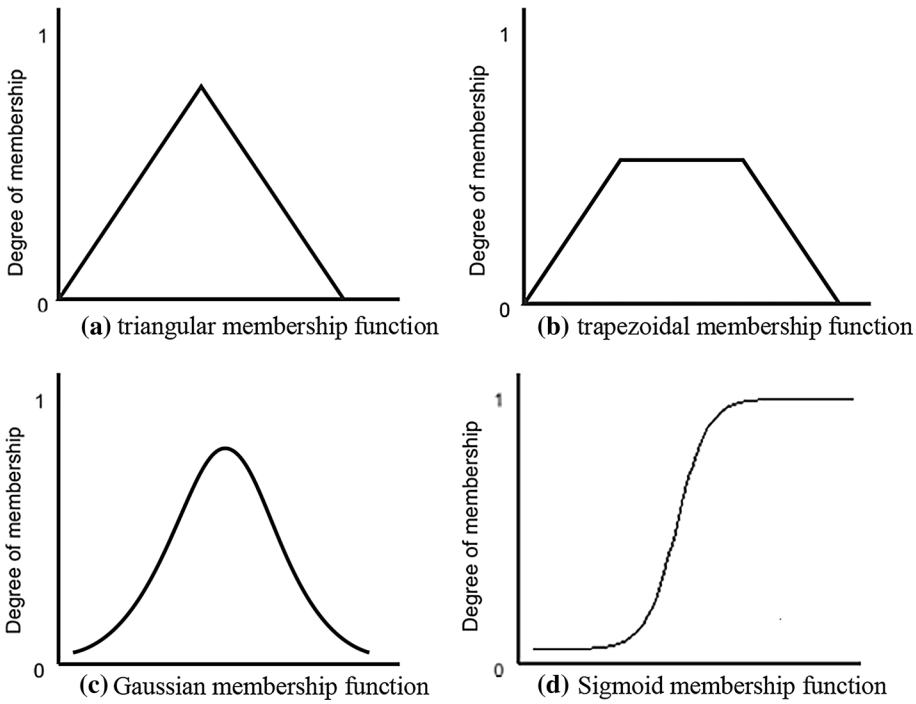

Fig. 4 Known membership functions include **a** the triangular membership function, **b** the trapezoidal membership functions, **c** the Gaussian membership functions, and **d** the sigmoid membership function

set theory, the membership of elements either belongs or does not belong to the set. By contrast, fuzzy set theory uses gradual degree of the membership of elements in a set. This is defined by a membership function, which is set at an interval [0, 1].

For example, an element x is either a member or nonmember of the set A; thus, the membership μ of $A(x)$ of x into A is defined as follows:

$$\begin{aligned} \mu A(x) &= 1 \quad \textit{if } x \textit{ is totally in } A \\ \mu A(x) &= 0 \quad \textit{if } x \textit{ is not in } A \\ 0 < \mu A(x) &< 1 \quad \textit{if } x \textit{ partly in } A \end{aligned} \tag{12}$$

The distinction between a classical crisp set and a fuzzy set is best captured schematically as shown in Fig. 3. The shaded area in Fig. 3b shows elements that belong to both A and B with varying degree of membership in both sets.

Using the process of fuzzification, crisp variables are converted to their corresponding fuzzy set. Fuzzification determines the degree to which a crisp variable belongs to the appropriate fuzzy sets using a membership function. There are various types of membership functions for obtaining the degree of membership of fuzzy sets. As shown in Fig. 4, traditional membership functions include the triangular membership function, the trapezoidal membership functions, the Gaussian membership functions, and the sigmoid membership. The choice of membership function type for fuzzy sets is determined based on its suitability such as convenience, simplicity, and effectiveness.

Fuzzy if–then rules have two distinct parts; the antecedent, which is fuzzifying the input and the consequent in which the fuzzy input is applied to obtain the output. This process is known as fuzzy reasoning. These rules typically take the form as shown below.

IF antecedent(s)THEN consequent(s)

Fuzzy sets can have multiple antecedents, which are connected with fuzzy operators (OR, AND, NOT). As shown in Fig. 5, if two fuzzy sets A and B have membership functions μA(x) and μB(x), respectively, then the

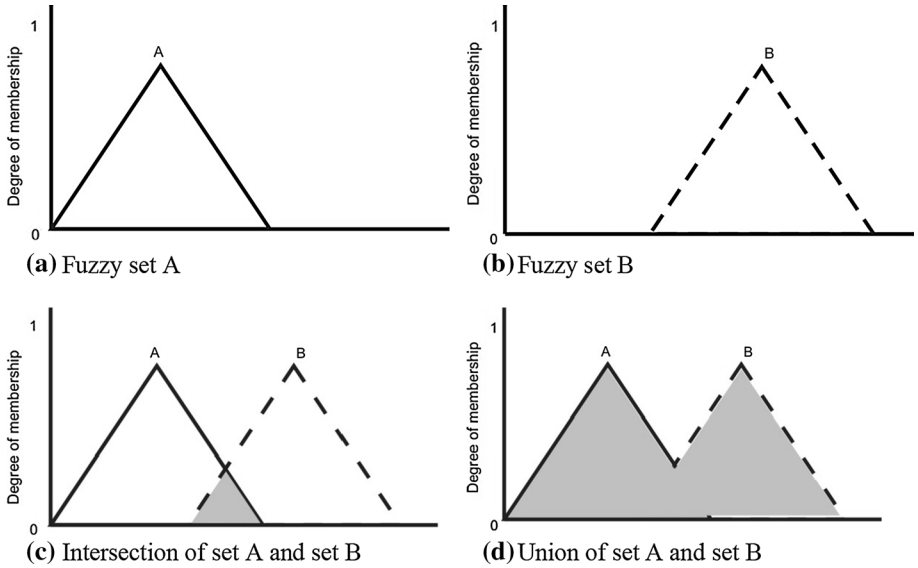

Fig. 5 Intersection, union, and complement of two fuzzy sets *A* and *B* and their corresponding membership functions $\mu A(x)$ and $\mu B(x)$. **a** Fuzzy set *A*; **b** Fuzzy set *B*; **c** Intersection of set *A* and set *B*; **d** Union of set *A* and set *B*

intersection, union, and complement of $\mu A(x)$ and $\mu B(x)$ are described using the following equations (Eqs. 13–15).

The intersection (AB) of two fuzzy sets *A* and *B* can be defined as:

$$\mu AB(x) = \min(\mu A(x), \mu B(x)) \tag{13}$$

The union (AB) of two fuzzy sets *A* and *B* can be defined as:

$$\mu AB(x) = max(\mu A(x), \mu B(x)) \tag{14}$$

The complement $\left(\bar{A}\right)$ of fuzzy sets *A* can be defined as:

$$\mu \bar{A}(x) = 1 - \mu A(x) \tag{15}$$

2.2.3.1 Fuzzy inference The process of fuzzy inference generation involves the use of three components. These include (a) the membership functions, (b) a logical operation, and (c) the if–then rules to formulate the mapping function from a given input to an output using fuzzy logic. The results of the fuzzy inference can be aggregated and then defuzzified so that the results can be a basis from which patterns can be discerned to enable decision making. In this work, we employed the following process.

Step 1—*The creation of the antecedent for fuzzy sets:* Here, we mapped the feature set to class label by using appropriate fuzzy operator (AND or OR) on the fuzzy if–then rules. We assigned weights to each membership function of the if–then rule. The weights ranged between the value 0 and 1, respectively.

Step 2—*Obtaining the consequent of the fuzzy inference:* The consequent of a fuzzy set is obtained by combining the rule strength and the output membership function. Here, we ensured that the rules were combined by aggregation in order to make decisions from the rules.

Defuzzification is the process of producing crisp result values for a given fuzzy set using the derived membership functions. In this work, the process of defuzzification was envisaged as a process that takes the aggregated output fuzzy set as input and generates a single crisp value as output. Traditional methods of defuzzification include the centroid, the bisector, the middle of maximum (the average of the maximum value of the output set), the largest of maximum, and the smallest of maximum [22]. We employed the centroid method to defuzzify the outputs to generate the measured area under the curve (AUC).

2.2.3.2 Proposed fuzzy inference system There are two major types of fuzzy inference systems, namely the Mamdani-type fuzzy inference and the Sugeno-type fuzzy inference. Proposed by Maamdani in 1975 [22], the Mamdani-type inference expects the output membership functions to be fuzzy sets. When this system is being used after the aggregation process, the process of defuzzification needs to be performed on the fuzzy set for each output variable.

The Takagi–Sugeno–Kang method of fuzzy inference also known as the Sugeno-type fuzzy inference was first introduced by Takagi–Sugeno–Kang in 1985 [29]. The method is similar to the Mamdani in that the first two parts of the fuzzy inference process follow the Mamdani process, i.e., fuzzifying the inputs and applying the fuzzy operator to generate if–then rules. But Sugeno method outputs membership functions that are either linear or constant.

The Mamdani fuzzy inference system uses only fixed membership functions that are chosen arbitrarily. This method applies fuzzy inference models whose rule structure is essentially predetermined by the user's interpretation of the characteristics of variables in the model. This becomes a challenge for our dataset as we cannot determine membership

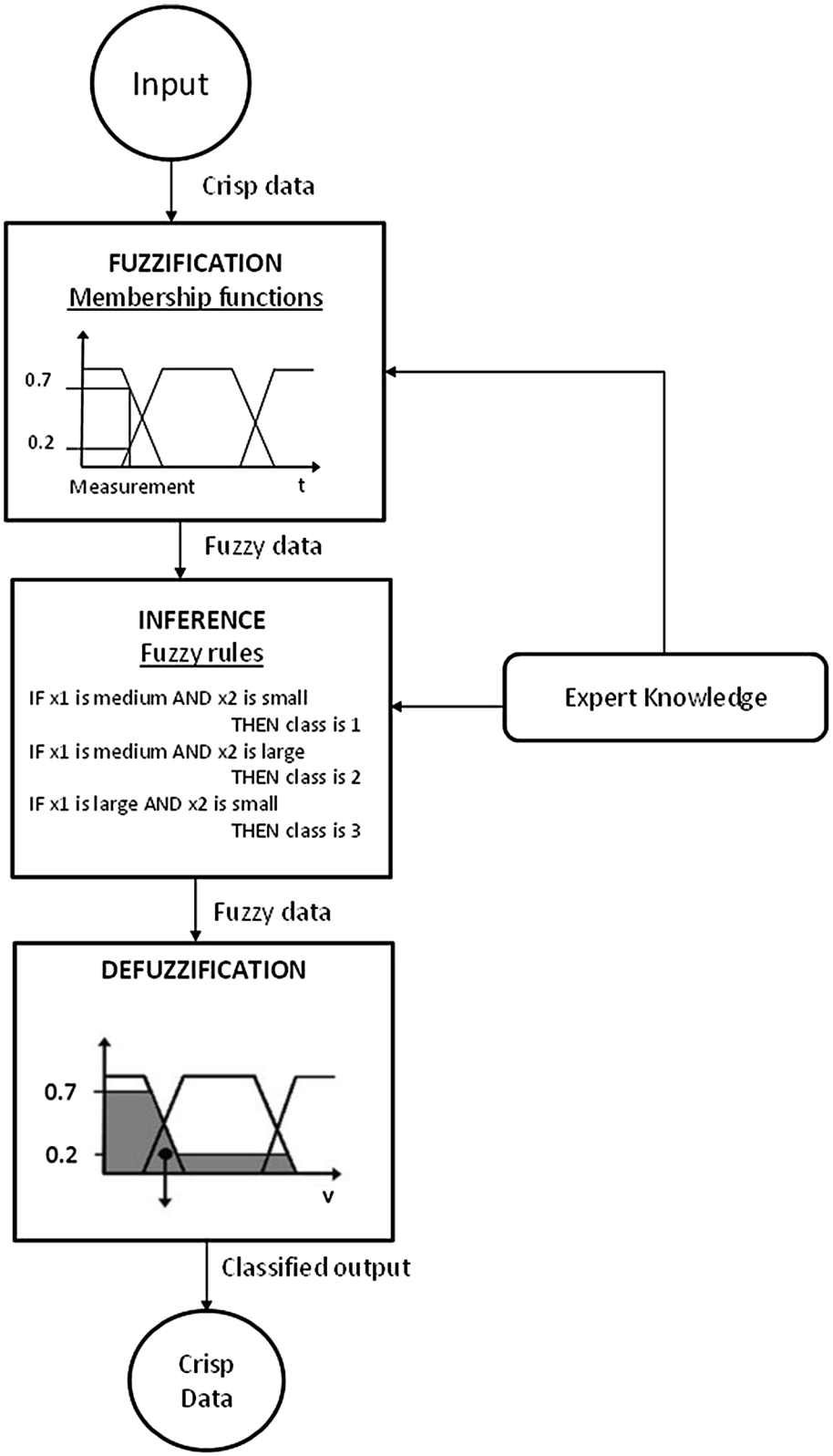

Fig. 6 Architecture of a fuzzy inference system

functions simply by looking at the data. Furthermore, we do not have a predetermined model structure derived from the features of the fundus images. The proposed system of fuzzy inference used in this work is best captured in Fig. 6.

2.2.4 Automatic determination of fuzzy sets and rules

Fuzzy sets are determined by membership functions. The selection of membership functions method and the

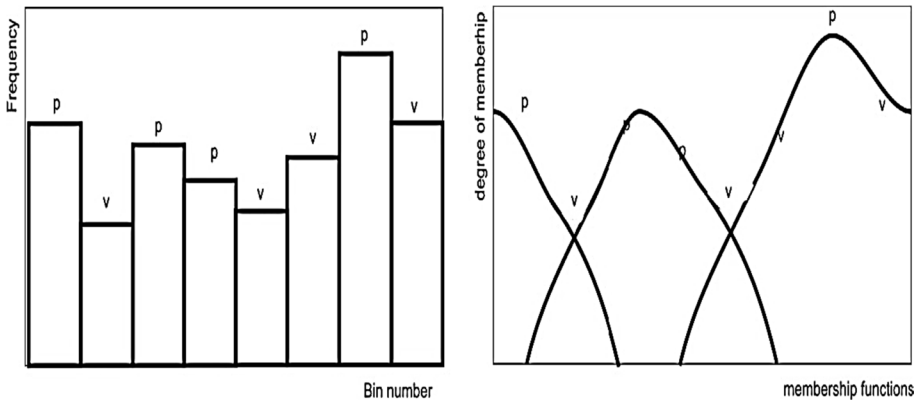

Fig. 7 Mapping of a peak–valley–peak histogram to a membership function

partitioning of the input into fuzzy sets are usually determined by expert knowledge through interpretation of the data. Obtaining fuzzy sets this way can only be time efficient if the number of inputs and range of values are less. There are various techniques that can automate this process and significantly improve performance while reducing the development time and interpretability of the fuzzy model. The methods, (a) equalized universe and (b) clustering, are examples of the methods used for automatic determination of fuzzy sets [21].

In equalized universe method, the inputs are equally partitioned with the chosen number of clusters and the membership functions are set at an equal distance along the range of each data input [21]. The grid partition method is used to generate a set of rules for equalized universe method. The grid partition method works by enumerating all possible combinations of membership functions of all inputs. For n inputs with m membership functions, n^m fuzzy if–then rules will be generated.

Similarly, clustering is the process of grouping a set of data that have high similarity and are very dissimilar to other objects in other groups. Clustering can be used to determine the degree of membership functions of the input from a given dataset. Fuzzy c-means (FCM) clustering is often used to determine initial membership functions [12]. FCM clusters or groups one piece of data to belong to multiple clusters simultaneously, with different degrees of membership.

However, the choice of the appropriate automatic determination method depends on its suitability to the application domain. In this work, we propose to adopt a novel data-adaptive partitioning method by Jain, 2012 [20]. This method partitions inputs into fuzzy sets based on the data distribution rather than conforming to fixed data models as other techniques assume.

2.2.5 Data-adaptive partitioning method

Data-adaptive partitioning method partitions each attribute of the dataset based on the distribution of the data by understanding the underlying nature of the data [20]. It adopts a data frequency-based partitioning to generate peaks and valleys from the distribution of the input data. The process of generating data-adaptive partitions involves histogram analysis, and identifying adaptive partitions using peak–valley–peak triple test.

Histogram analysis is used to identify peaks and valleys of each input. A vertical bar in histogram is called a peak (p) if its frequency count of data points is greater than the frequency count of data points of the next vertical bar. A vertical bar in histogram is called a valley (v) if its frequency count of data points is less than the frequency count of data points of the next vertical bar. A vertical bar in histogram is called an unassigned histogram (p) if the histogram is neither a peak nor a valley. Peak, valley, and unassigned histogram are defined in Eq. 16 below.

$$\begin{aligned} Peak(p_i) &= f_i > f_j, \\ Valley(v_i) &= f_i < f_j, \text{ and} \\ Unassigned\ Histogram(x_i) &= f_i = f_j \end{aligned} \tag{16}$$

where i is the i^{th} vertical bar of the histogram. j is the next vertical bar of i. f_i is the frequency count of data points of vertical bar i, and f_j is the frequency count of data points of vertical bar j, respectively.

2.2.6 Finding peak–valley–peak triplet sets in a histogram

The peak–valley–peak triplet set is used to generate the adaptive partitions of the inputs. Each peak–valley–peak triplet gives one partition, and each partition represents a membership function of the input. To obtain peak–valley–peak combination, some assumptions have been made (refer to Fig. 7):

If there is any valley (v) or unassigned data point (x) or both valley (v) and unassigned point (x) in between two peaks $\left(p_i \text{ and } p_j\right)$, then peaks $\left(p_i \text{ and } p_j\right)$, valley ($v$), and unassigned point (x) will be combined into one adaptive partition.

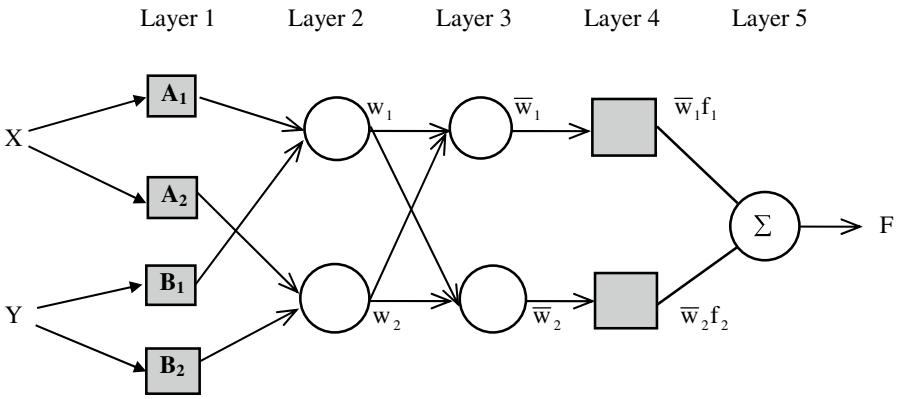

Fig. 8 ANFIS architecture for a two rule system

Table 1 Data description of normalized data (Min–max normalization of range 0–1)

Role	Feature	Type	Statistics	Range
Regular	cA2_Kapoor	Real	Avg = 0.715 ± 0.179	[0.000; 1.000]
Regular	cA2_Renyi	Real	Avg = 0.108 ± 0.133	[0.000; 1.000]
Regular	cA2_Shanon	Real	Avg = 0.244 ± 0.203	[0.000; 1.000]
Regular	cA2_Yager	Real	Avg = 0.331 ± 0.195	[0.000; 1.000]
Regular	cA2_Fourier Energy	Real	Avg = 0.459 ± 0.171	[0.000; 1.000]
Regular	cH2_Kapoor	Real	Avg = 0.666 ± 0.175	[0.000; 1.000]
Regular	cH2_Renyi	Real	Avg = 0.558 ± 0.161	[0.000; 1.000]
Regular	cH2_Shanon	Real	Avg = 0.609 ± 0.195	[0.000; 1.000]
Regular	cH2_Yager	Real	Avg = 0.389 ± 0.209	[0.000; 1.000]
Regular	cH2_Fourier Energy	Real	Avg = 0.306 ± 0.163	[0.000; 1.000]
Regular	cH1_Kapoor	Real	Avg = 0.617 ± 0.183	[0.000; 1.000]
Regular	cH1_Renyi	Real	Avg = 0.620 ± 0.218	[0.000; 1.000]
Regular	cH1_Shanon	Real	Avg = 0.616 ± 0.191	[0.000; 1.000]
Regular	cH1_Yager	Real	Avg = 0.490 ± 0.204	[0.000; 1.000]
Regular	cH1_Fourier Energy	Real	Avg = 0.336 ± 0.178	[0.000; 1.000]
Regular	cV2_Kapoor	Real	Avg = 0.633 ± 0.181	[0.000; 1.000]
Regular	cV2_Renyi	Real	Avg = 0.579 ± 0.132	[0.000; 1.000]
Regular	cV2_Shanon	Real	Avg = 0.695 ± 0.147	[0.000; 1.000]
Regular	cV2_Yager	Real	Avg = 0.380 ± 0.165	[0.000; 1.000]
Regular	cV2_Fourier Energy	Real	Avg = 0.355 ± 0.175	[0.000; 1.000]
Regular	cV1_Kapoor	Real	Avg = 0.573 ± 0.186	[0.000; 1.000]
Regular	cV1_Renyi	Real	Avg = 0.701 ± 0.155	[0.000; 1.000]
Regular	cV1_Shanon	Real	Avg = 0.653 ± 0.170	[0.000; 1.000]
Regular	cV1_Yager	Real	Avg = 0.539 ± 0.185	[0.000; 1.000]
Regular	cV1_Fourier Energy	Real	Avg = 0.340 ± 0.188	[0.000; 1.000]

If there is one peak in the histogram, then the partition range will be from 0 to p_1 and p_1 to 1, i.e., two adaptive partitions.

If there is a combination of peak–valley–peak $(p_i{-}v{-}p_j)$ triplets in the histogram, then the partition range will be from half of the partition from the left of p_i and half of the partition from the left of p_j because both the peaks $(p_i \, and \, p_j)$ are being shared with neighboring partitions (intercepts).

2.2.7 *Validating the partitions*

We used the $\beta = 10$ threshold, $\beta = 20$ threshold and Xie and Beni's index to validate the partitions and mine for outliers after applying the peak–valley–peak procedure. The $\beta = 10$ and $\beta = 20$ threshold are a fixed partitioning methods that split each feature so that a minimum number of 10 and 20 %, respectively, of data samples are represented in each partition. The Xie and Beni's [32] index

Table 2 Analysis of variance (ANOVA) of the 25 features

Features	Normal		Non-CSME		CSME		Significance	
	Mean	SD	Mean	SD	Mean	SD	*p* value	*F* value
cA2_Shanon	3.6286	1.3144	2.0342	0.7084	2.5102	0.6936	0.0000	74.1398
cA2_Renyi	1.2252	0.8117	0.5096	0.3333	0.6612	0.2467	0.0000	51.3390
cA2_Kapoor	7.5973	0.1168	7.3691	0.2236	7.5087	0.1779	0.0000	41.6647
cA2_Fourier Energy	3.54E+07	1.05E+07	4.70E+07	1.11E+07	4.68E+07	1.05E+07	0.0000	38.4955
cV2_Shanon	3.6717	0.2193	3.9107	0.1882	3.8284	0.2311	0.0000	32.2917
cH2_Shanon	3.6223	0.1947	3.8297	0.1934	3.7582	0.2301	0.0000	25.9529
cV1_Shanon	3.0578	0.2002	3.2404	0.1841	3.1864	0.2086	0.0000	22.4754
cV2_Kapoor	6.8984	0.3126	7.1304	0.2573	7.1076	0.2664	0.0000	20.8876
cV2_Fourier Energy	2.96E+06	1.19E+06	4.17E+06	1.52E+06	3.89E+06	1.45E+06	0.0000	20.6473
cH1_Renyi	1.2192	0.0508	1.2654	0.0461	1.2341	0.0594	0.0000	20.2603
cV1_Renyi	1.2273	0.0623	1.2773	0.0481	1.2491	0.0612	0.0000	18.9508
cH1_Shanon	3.0604	0.2024	3.2254	0.1937	3.1598	0.2157	0.0000	16.5528
cV2_Yager	0.9843	0.0023	0.9829	0.0019	0.9832	0.0019	0.0000	12.6956
cH2_Renyi	1.2946	0.0674	1.3404	0.0593	1.3069	0.0773	0.0000	12.0199
cV1_Fourier Energy	8.77E+06	3.76E+06	1.14E+07	4.39E+06	1.09E+07	4.70E+06	0.0000	10.4918
cV2_Renyi	1.3464	0.0842	1.3913	0.0654	1.3517	0.0837	0.0001	9.8331
cH2_Kapoor	6.9714	0.3588	7.1350	0.2940	7.1267	0.2761	0.0002	8.7401
cH2_Fourier Energy	3.15E+06	1.63E+06	3.89E+06	1.67E+06	3.72E+06	1.41E+06	0.0026	6.0885
cV1_Kapoor	5.9187	0.3577	6.0654	0.3287	6.0601	0.3388	0.0030	5.9187
cA2_Yager	0.9810	0.0013	0.9816	0.0011	0.9814	0.0013	0.0043	5.5604
cH1_Fourier Energy	9.25E+06	4.20E+06	1.14E+07	4.97E+06	1.07E+07	4.69E+06	0.0045	5.4941
cH2_Yager	0.9838	0.0028	0.9830	0.0024	0.9831	0.0020	0.0333	3.4417
cH1_Kapoor	5.9551	0.4372	6.0987	0.3768	6.0396	0.3937	0.0422	3.1991
cH1_Yager	0.9971	0.0007	0.9970	0.0006	0.9971	0.0006	0.3295	1.1143
cV1_Yager	0.9972	0.0006	0.9971	0.0006	0.9971	0.0006	0.7688	0.2631

is proposed to find and validate the best partitions so as to determine an optimal number of clusters for the FCM algorithm.

2.2.8 *Data-adaptive neuro-fuzzy inference system*

Here, we extend the adaptive neuro-fuzzy inference system (ANFIS), which is based on Takagi–Sugeno fuzzy inference system. This learning technique uses an approach, whereby the fuzzy modeling procedure learns information about a dataset from the dataset. The membership functions are determined from the features. The rules from the ANFIS are generated by adjusting the assigned weights using the forward pass and back-propagation algorithms.

The parameters associated with the membership functions change throughout the learning process. The adjustment of these parameters is computed by a gradient vector. The gradient vector measures the performance of the fuzzy inference system in mapping the input or output data for a given set of parameters. The sum of the squared difference between actual and desired outputs is then used to calculate the error, and the weights are adjusted accordingly. To present the ANFIS architecture, two fuzzy if–then rules are considered:

Rule 1: $IF\ x$ is A_1 and y is $B_1\ THEN\ f_1 = p_1x + q_1y + r_1$

Rule 2: $IF\ x$ is A_2 and y is $B_2\ THEN\ f_2 = p_2x + q_2y + r_2$

where x and y are the inputs, A_i and B_i are the fuzzy sets, f_i are the outputs within the fuzzy region specified by the fuzzy rule, p_i, q_i, and r_i are the design parameters for training the model. In this work, we use the ANFIS architecture as shown in Fig. 8. The circular nodes represent nodes that are fixed, whereas the square nodes are nodes that have parameters to be learnt and are referred to as adaptive nodes.

2.2.9 *Evaluation of model performance*

In this study, we employ the tenfold cross-validation to evaluate model performance. Here, the entire dataset is divided into ten equal folds. During evaluation, ninefolds of the divided dataset are chosen at random to train the

Table 3 Measures of kurtosis and skewness for the twenty-five features in the dataset

Features	Kurtosis	Skewness	Variance
cA2_Renyi	11.48819	2.976931	0.370218
cA2_Kapoor	1.982845	−1.29726	0.04041
cA2_Shanon	1.771223	1.453815	1.345627
cH2_Kapoor	0.977189	−0.98388	0.102174
cA2_Fourier energy	0.494757	0.28613	1.44E+14
cA2_Yager	−0.05212	0.837173	1.60E−06
cH2_Fourier energy	1.551473	0.83616	2.56E+12
cH1_Kapoor	0.65345	−0.87976	0.165105
cH2_Yager	0.182438	0.769163	5.90E−06
cH2_Renyi	0.124982	−0.32501	0.005023
cH2_Shanon	0.049147	−0.52478	0.049894
cH1_Shanon	−0.05264	−0.56969	0.046003
cH1_Renyi	−0.45937	−0.54209	0.003097
cV2_Renyi	1.71844	−0.71698	0.006486
cV2_Shanon	1.52904	−0.87046	0.055185
cV2_Kapoor	1.225208	−0.98441	0.088723
cV2_Yager	0.85869	0.76724	4.51E−06
cH1_Fourier energy	0.281195	0.517865	2.21E+13
cH1_Yager	−0.20895	0.429413	3.76E−07
cV1_Renyi	1.248789	−0.96036	0.003708
cV1_Fourier energy	0.410619	0.666816	1.97E+13
cV1_Kapoor	0.296821	−0.55307	0.120757
cV1_Shanon	0.269442	−0.57017	0.04478
cV2_Fourier energy	0.148951	0.502189	2.20E+12
cV1_Yager	−0.093	−0.16179	3.39E−07

model, and the remaining fold is used for the purpose of testing. This process is repeated 10 times, ensuring in every iteration we choose a unique test set.

Apart from the accuracy of the model, we used the measurements of root mean square error (RMSE), percentage root mean square error (% RMSE) to estimate the models performance for every iteration of the tenfold cross-validation. Both RMSE and % RMSE are effective means to establish the ability of the model to generalize across classes. The RMSE and % RMSE are represented as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(t(i) - a(i))^2}$$

$$\%\ \mathrm{RMSE} = \frac{\mathrm{RMSE}}{\mathrm{Mean}(a)}$$

where t is the predicted target value, and a is the actual output value, and n is the number of data points.

3 Results

In this section, we describe the performance of the proposed decision system using textural features for the classification of DM images into three known classes.

3.1 Normalization

To curtail the variance in features, the dataset is normalized using min–max normalization of range 0–1 as shown in Table 1. Normalization is a process, whereby the data features are scaled so as to fall within an equal range.

3.2 Feature significance

In our work, a total of 25 features were extracted after the feature extraction method. Feature ranking and selection is a critical step to select a subset of significant features from the entire set of features. The feature with the top rank is more valued for the classification rather than a feature with a lower rank. Also, ignoring the features of lower rank increases the classification speed. The significance of extracted features was obtained using the *analysis of variance* (ANOVA) and then with the *kurtosis and skewness* test. After the test of significance, the features were ranked using *Fisher's discrimination index (F).*

The analysis of variance (ANOVA) The ANOVA test assumes that the data in each feature are normally distributed. Table 2 provides an overview of the results obtained. When subjected to a significance threshold of p value <0.05, it was observed that all except for two features were significant. Both features cH1_Yager and cV1_Yager exceeded the threshold and are believed to have excess variance.

Test for kurtosis and skewness Here, both measures of skewness and kurtosis are used to establish if the features exhibit any distortions in their distributions. According to popular belief, all learning algorithms assume that data fed as input are normally distributed. These algorithms would fail if the data are skewed away from a zero mean. As shown in Table 3, the features were sorted based on descending order of kurtosis. It was observed that features cA2_Renyi exhibited the highest kurtosis and a positive skewness (Table 3), despite being reported as significant as shown in Table 2. It was further observed that the remaining variables exhibited high kurtosis and were skewed even though they had p values <0.05.

A visual representation of the effect of kurtosis and skewness on a feature CA2_Yager that had a kurtosis of −0.05212 and a positive skew of 0.837 is represented in Fig. 9. This representation is subjected to the adaptive data partitioning to generate an ideal membership function that captures the inherent data distribution.

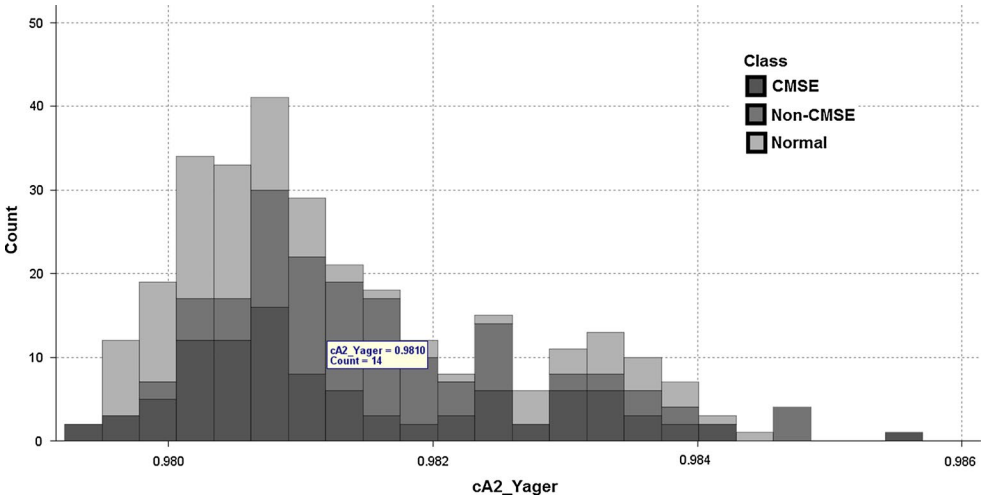

Fig. 9 Histogram representation of feature cA2_Yager. The feature is negatively skewed and exhibits kurtosis of 0.837

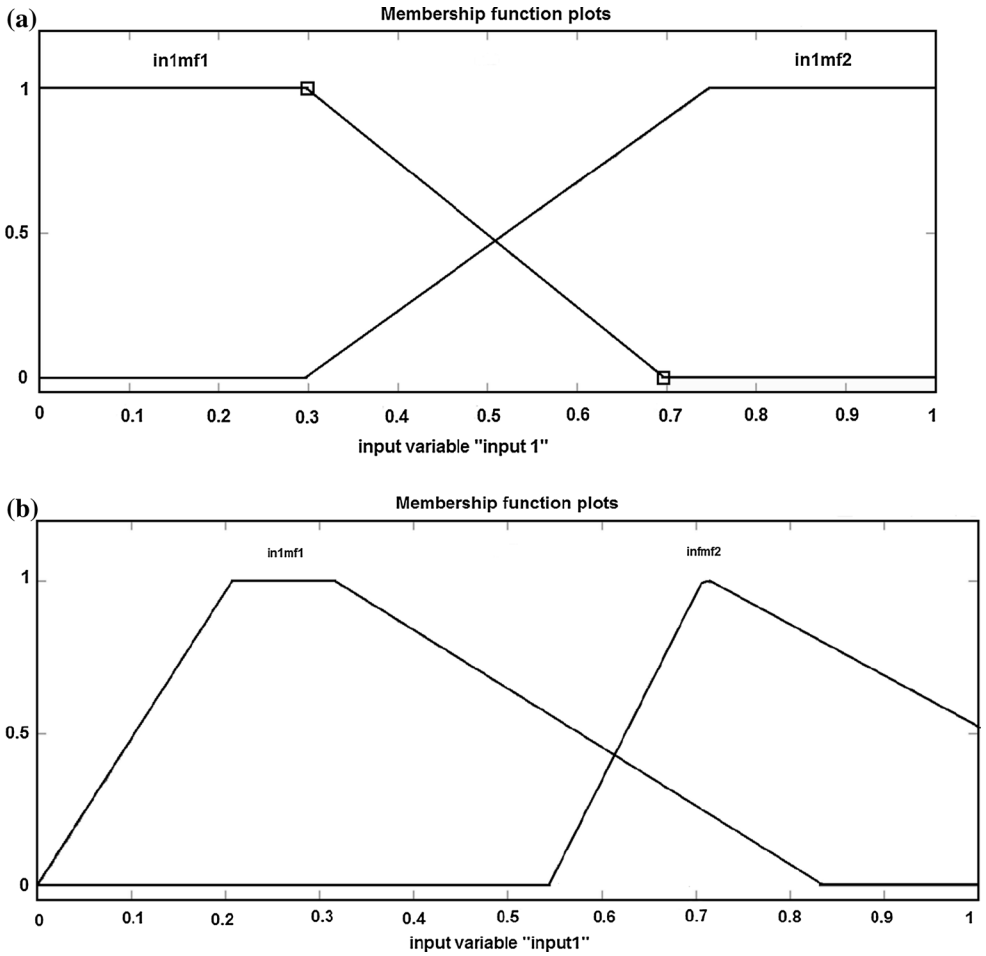

Fig. 10 Resulting membership function of feature cA2_Yager using **a** equalized universe method and **b** the proposed peak-valley-peak method

3.3 Adaptive data partitioning

Each feature in the dataset is subjected to adaptive partitioning using the peak–valley–peak method. Figure 10 provides a graphical representation of the resultant membership function using peak–valley–peak method of adaptive data partitioning as compared to the universal equalized method. This was repeated on each of the 25 features in the dataset. The resultant membership functions were then used to generate rules using the adaptive neuro-fuzzy classifier.

3.4 Classification

The normalized data were classified using both fuzzy rule-based and non-fuzzy rule-based algorithms. ANFIS, based on fuzzy inference systems generated from a fuzzy *c*-means (FCM) clustering method, was used to classify the dataset with different automatic fuzzy set and fuzzy rules detection methods. The prediction model's performance consists of RMSE, % RMSE, and the accuracy achieve (refer Table 4) for each fold of cross-validation.

As shown in Figure 11, the training RMSE continues to decrease until the 200th epoch, while the testing RMSE before it reduces when it reaches the 180th epoch and slightly increases and the training RMSE increases.

The performance was evaluated using overall accuracy, weighted mean precision, and weighted mean recall.

Table 4 Model performance

Cross-validation fold	RMSE	% RMSE	Accuracy (%)
1	0.2753	13.77	96.67
2	0.3292	16.46	96.67
3	0.3439	17.19	90.00
4	0.2771	13.86	100.00
5	0.2700	13.50	93.33
6	0.3222	16.11	100.00
7	0.2898	14.49	100.00
8	0.2746	13.73	98.33
9	0.2912	14.56	86.67
10	0.3072	15.36	96.67

Table 5 shows the performance of the system with the different algorithms used using tenfold cross-validation.

4 Discussion

It was observed that there was a boost in accuracies using the data-adaptive partitioning method to generate the membership function. To validate this finding, we compared our results with traditional non-fuzzy rule-based classification algorithms such as ANN, SVM, decision trees, Naïve Bayes, KNN, rule induction, and ID3 (As shown in Table 5). We know that these classification techniques work on the assumption that all the features are independent of each other and parametric classifiers largely rely on assumptions of data distribution. The independent component analysis (ICA) was used to obtain a class-conditional representation to support this assumption. In ICA, multidimensional data are decomposed into components that are maximally independent in the negentropy sense, uncovering disjoint underlying trends in the data. Amongst the non-fuzzy classifiers, it was observed that ANN performed the best with a reported overall accuracy of 83.33 %. With the inclusion of ANFIS, we observed a boost in accuracy from 83.33 to 92.10 % using the clustering method and furthermore an additional 5 % increase in accuracy using the equalized universe method. With the inclusion of the data-adaptive partitioning method, we observed the highest accuracy of 98.55 %, respectively, as shown in Table 6. We further subjected our results to a detailed comparison of recent related works in this area as shown in Table 7 and observed that the performance of the proposed data-ANFIS showed comparative improvement in overall accuracy.

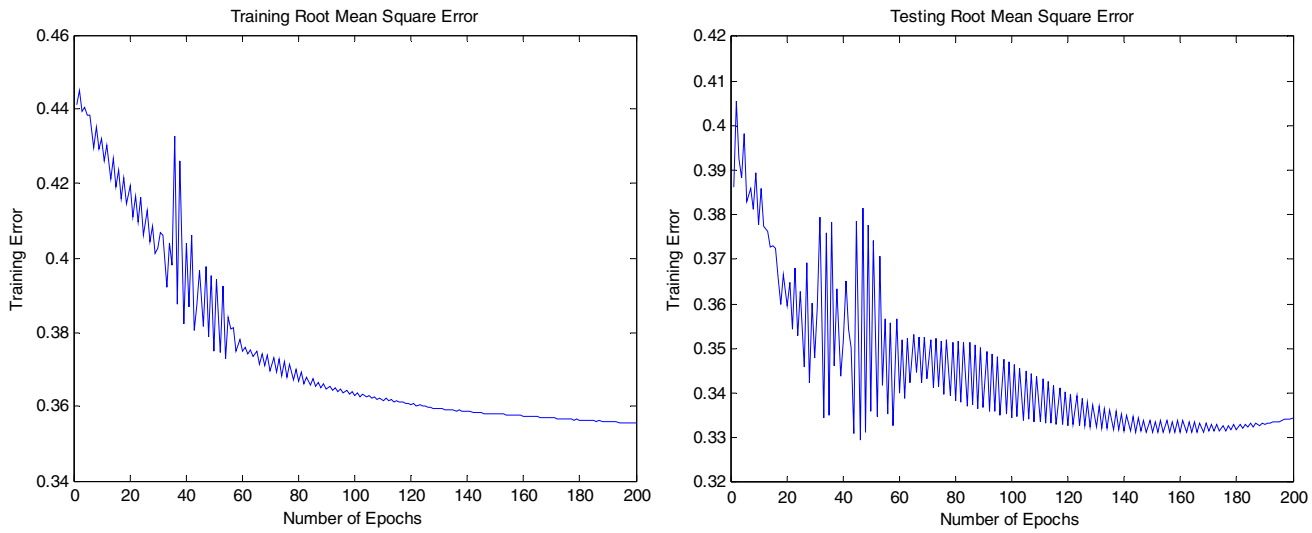

Fig. 11 Training and testing root mean square errors

Table 5 Performance of the different non-fuzzy rule-based classifiers on the diabetes maculopathy dataset using tenfold cross-validation

Classifier	Overall accuracy (%)	Weight mean precision (%)	Weighted mean recall (%)
ANN	83.33	83.53	83.33
SVM	71.11	70.51	71.11
Decision tree	70.00	73.96	70.00
Naïve Bayes	66.67	66.10	66.67
KNN	54.44	82.36	54.44
Rule induction	75.56	75.64	75.56
ID3	70.00	69.33	70.00

Table 6 Performance of the different fuzzy rule-based classifiers using tenfold cross-validation on the diabetes maculopathy dataset

Classifier	Overall accuracy (%)	Weight mean precision (%)	Weighted mean recall (%)
ANFIS + clustering method	92.10	92.47	92.10
ANFIS + equalized universe method	97.55	96.88	97.55
ANFIS + data-adaptive partitioning method (with cross-validation)	98.55	98.45	98.55
ANFIS + data-adaptive partitioning method (with bootstrapping)	98.33	98.33	98.41

Table 7 Comparison of recent works in the area

References	Methods	Classifiers	Performance
Nayak et al. [25]	Adaptive contrast enhancement and matched correlation	ANN	Sensitivity—95.40 % Specificity—100 %
Fleming et al. [16]	Nonspecific morphological filter	k-NN	Accuracy—99.2 % (NCSME) Accuracy—97.3 % (CSME)
Siddalingaswamy et al. [28]	Geometrical relation and clustering	Positional constrains	Sensitivity—95.60 % Specificity—96.15 %
Ang et al. [1]	Morphological image processing	ANN	Sensitivity—90 % Specificity—100 % Accuracy—96.67 %
Hunter et al. [18]	Peak point detection, morphology, color and texture feature	ANN	Sensitivity—97 % Specificity—80 %
Akram et al. [5]	Morphology, energy, filter bank response of exudates	SVM	Sensitivity—92.60 % Specificity—97.80 % Accuracy—97.30 %
Alipour et al. [6]	Curvelet transform	ANN	Sensitivity—93 % Specificity—86 %
Deepak et al., [13]	Motion pattern analysis	Gaussian data description (GDD) and Principal Component Analysis Data Description (PCADD)	AUC = 0.92–0.98
Giancardo et al. [17]	Kirsch edge operator, color and wavelet analysis	SVM	AUC—0.94
Chua et al. [11]	Adaptive histogram equalization and texture	Fuzzy	Sensitivity—100 % Specificity—100 % Accuracy—86.67 %
Chowriappa et al. [10]	Texture features	Ensemble learning	Accuracy—96.70 %
Punnolil et al. [27]	Morphology and texture features of exudates	SVM	Sensitivity—96.89 % Specificity—97.15 %
Tariq et al. [30]	Gabor filter bank And Otsu thresholding	SVM	Accuracy >97 %
This work	DWT entropy and energy	Data-adaptive neuro-fuzzy inference system (ANFIS)	Accuracy—98.5 %

5 Conclusions

Fuzzy rule-based classification offers improved performance and better interpretability of results. Adaptive neuro-fuzzy classification combines both the accuracy of neural networks and interpretability of fuzzy rule-based classification. Rule-based classification algorithms depreciate in performance when faced with a large number of rules. This is due to inefficient partitioning, which determines the number of rules and classification performance.

This paper adopts a data-adaptive partitioning which uses a data frequency-driven approach to partition each attribute based on the distribution of data in that attribute. Feature extraction, transformation, and data normalization were applied in the pre-processing stage while a data-adaptive partitioning method had a significant role in developing an efficient data-ANFIS classifier model. Weights are assigned to membership functions to determine optimum output for the classifier. It is observed from Tables 5, 6, and 7 that when we compared the performance of classification of maculopathy dataset using non-fuzzy rule-based classifier, fuzzy rule-based classifiers with clustering based, equalized universe and data-adaptive partitioning methods, we obtained the maximum overall accuracy of 98.55 % from data-adaptive methods. Results obtained from the classification indicate that adaptive partitions increase the separation between the classes and therefore produce more effective rules and make a more effective classification system.

Acknowledgement The project described was supported in part by Grant Number P20RR016456 from the National Center for Research Resources (NCRR). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NCRR or the National Institutes of Health (NIH).

References

1. Ang MH, Acharya, UR, Sree SV, Lim TC, Suri, JS (2011) Computer-based identification of diabetic maculopathy stages using fundus images. In: Multi modality state-of-the-art medical image segmentation and registration methodologies. Springer US. pp 377–399
2. Acharya UR, Dua S, Du X, Sree SV, Chua CK (2011) Automated diagnosis of glaucoma using texture and higher order spectra features. Inf Technol Biomed IEEE Trans 15(3):449–455
3. Acharya UR, Sree VS, Saba L, Molinari F, Guerriero S, Suri JS (2012) Ovarian tumor characterization and classification using ultrasound: a new on-line paradigm. J Digit Imaging 26(3):544–553
4. Acharya UR, Mookiah MRK, Sree SV, Yanti R, Martis R, Saba L, Suri JS (2013) Evolutionary algorithm-based classifier parameter tuning for automatic ovarian cancer tissue characterization and classification. In: Saba L, Acharya UR, Guerriero S, Suri JS (eds) Ovarian neoplasm imaging. Springer. pp 425–440
5. Akram MU, Akhtar M, Javed MY (2012) An automated system for the grading of diabetic maculopathy in fundus images. In: Huang T, Zeng Z, Li C, Leung C (eds) Neural information processing. Springer, Berlin. pp 36–43
6. Alipour SHM, Rabbani H, Akhlaghi M, Dehnavi AM, Javanmard SH (2012) Analysis of foveal avascular zone for grading of diabetic retinopathy severity based on curvelet transform. Graefe's Arch Clin Exp Ophthalmol 250(11):1607–1614
7. Azeez D, Ali M, Gan K, Saiboon I, (2013) Comparison of adaptive neuro-fuzzy inference system and artificial neutral networks model to categorize patients in the emergency department. Springer 2013 2:416, vol. 2, no. 1, pp 1–10
8. Bishop CM (1995) Neural networks for pattern recognition. Oxford University Press, Oxford
9. Bracewell RN, Bracewell RN (1986) The fourier transform and its applications, vol 31999. McGraw-Hill, New York
10. Chowriappa P, Dua S, Acharya UR, Mookiah MRK (2013) Ensemble selection for feature-based classification of diabetic maculopathy images. Comput Biol Med 43(12):2156–2162
11. Chua CK, Mookiah MRK, Koh JEW, Acharya UR, Lim CM, Laude A, Ng EYK (2013) Automated diagnosis of maculopathy stages using texture features. Int J Integr Care 13:1–8
12. Cintra ME, de Arruda Camargo H, Martin T (2009) Optimizing the fuzzy granulation of attribute domains. IFSA/EUSFLAT Conf. pp 742–747
13. Deepak KS, Sivaswamy J (2012) Automatic assessment of macular edema from color retinal images. Med Imaging, IEEE Trans 31(3):766–776
14. Dua S, Acharya UR, Chowriappa P, Sree SV (2012) Wavelet-based energy features for glaucomatous image classification. Inf Technol Biomed IEEE Trans 16(1):80–87
15. Fadzil MA, Izhar LI, Nugroho H, Nugroho HA (2011) Analysis of retinal fundus images for grading of diabetic retinopathy severity. Med Biol Eng Compu 49(6):693–700
16. Fleming AD, Goatman KA, Philip S, Prescott GJ, Sharp PF, Olson JA (2010) Automated grading for diabetic retinopathy: a large-scale audit using arbitration by clinical experts. Br J Ophthalmol 94(12):1606–1610
17. Giancardo L, Meriaudeau F, Karnowski TP, Li Y, Garg S, Tobin KW Jr, Chaum E (2012) Exudate-based diabetic macular edema detection in fundus images using publicly available datasets. Med Image Anal 16(1):216–226
18. Hunter A, Lowell JA, Ryder B, Basu A, Steel D (2011) Automated diagnosis of referable maculopathy in diabetic retinopathy screening. In: Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE. pp 3375–3378
19. Ishibuchi H, Kaisho Y, Nojima Y (2008) Designing fuzzy rule-based classifiers that can visually explain their classification results to human users. In: Genetic and evolving systems, 2008. GEFS 2008. 3rd International Workshop on IEEE. pp 5–10
20. Jain M (2012) Data adaptive rule-based classification system. M.S. thesis, Louisiana Tech University, USA
21. Jang JS (1993) ANFIS: adaptive-network-based fuzzy inference system. Syst Man Cybern IEEE Trans 23(3):665–685
22. Mamdani EH, Assilian S (1975) An experiment in linguistic synthesis with a fuzzy logic controller. Int J Man Mach Stud 7(1):1–13
23. Mookiah MRK, Acharya UR, Lim C, Petznick A, Suri JS (2012) Data mining technique for automated diagnosis of glaucoma using higher order spectra and wavelet energy features. Knowl Based Syst 33:73–82
24. Mookiah MRK, Acharya UR, Chua CK, Lim CM, Ng EYK, Laude A (2013) Computer-aided diagnosis of diabetic retinopathy: a review. Comput Biol Med 43(12):2136–2155

25. Nayak J, Bhat PS, Acharya UR (2009) Automatic identification of diabetic maculopathy stages using fundus images. J Med Eng Technol 33(2):119–129
26. Pharwaha APS, Singh B (2009) Shannon and non-shannon measures of entropy for statistical texture feature extraction in digitized mammograms. Proc World Congr Eng Comput Sci 2:20–22
27. Punnolil A (2013) A novel approach for diagnosis and severity grading of diabetic maculopathy. Advances in Computing, Communications and Informatics (ICACCI), 2013 International Conference IEEE pp 1230–1235
28. Siddalingaswamy PC, Prabhu KG (2010) Automatic grading of diabetic maculopathy severity levels. Systems in Medicine and Biology (ICSMB), 2010 International Conference IEEE. pp 331–334
29. Sugeno M (1985) Industrial applications of fuzzy control. Elsevier, Amsterdam
30. Tariq A, Akram MU, Shaukat A, Khan SA (2013) Automated detection and grading of diabetic maculopathy in digital retinal images. J Digit Imaging 26(4):803–812
31. Teng T, Lefley M, Claremont D (2002) Progress towards automated diabetic ocular screening: a review of image analysis and intelligent systems for diabetic retinopathy. Med Biol Eng Comput 40(1):2–13
32. Xie X, Beni G (1991) A validity measure for fuzzy clustering. IEEE Trans Pattern Anal Mach Intell (PAMI) 13(8):841–847
33. Zadeh LA (1965) Fuzzy sets. Inf Control 8(3):338–353