Classification of Rice using Genetic Fuzzy cascading system

Dipin Nair and Kelly Cohen

Abstract Classification can be done using various AI methods currently available in the market. But most of the AI techniques are black box. We do not know what is going on inside it and hence explainability of the model is very limited. Fuzzy system can increase the explainability to certain degree. In this paper we classify two different type of rice with use of genetic fuzzy cascading system and compare the accuracy and explainability with other methods. A total of rice grain images converted to grey scale images and with help of computer-vision, the attributes are obtained. These data are used for classification using a 7 input 2 output Fuzzy Inference System(FIS) with multiple levels of cascading. Each of the input, output membership functions and the rule base are tuned using genetic algorithm

Key words: Fuzzy, Genetic algorithm, classification of rice, soft computing

1 Introduction

Rice being is one of the popular cereal corps in the world, it needs to be exported all around globe. For this purpose, it has to be classified, cleaned and sorted efficiently. The necessity of classifying rice varieties using a less time consuming and more efficient led the authors Cinar & Kokul [1] to come up with a computer-vision and machine learning combined method. Image take from fixed distance of 2 different rice grains (Cameo and Osmancik) as shown in figure. 1

This image then converted to binary images and 7 attributes has been iterated from each image. which are given in table 1.

Dipin Nair
Nair, University of Cincinnati e-mail: nairdk@mail.uc.edu

Kelly Cohen
Cohen, University of Cincinnati e-mail: cohenky@ucmail.uc.edu

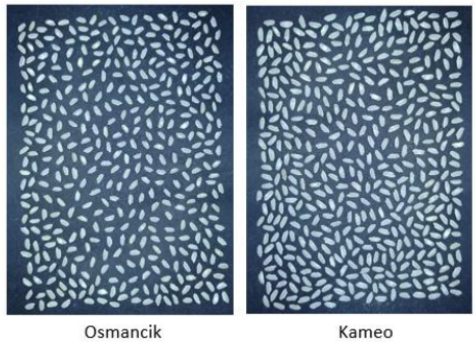


Fig. 1: Rice samples taken by camera

No	Name	Explanation
1	Area	Returns the number of pixels within the boundaries of the rice grain.
2	Perimeter	Calculates the circumference by calculating the distance between pixels around the boundaries of the rice grain.
3	MajorAxisLength	The longest line that can be drawn on the rice grain, i.e. the main axis distance, gives.
4	MinorAxisLength	The shortest line that can be drawn on the rice grain, i.e. the small axis distance, gives.
5	Eccentricity	It measures how round the ellipse, which has the same moments as the rice grain, is.
6	ConvexArea	Returns the pixel count of the smallest convex shell of the region formed by the rice grain.
7	Extent	Returns the ratio of the region formed by the rice grain to the bounding box pixels

Table 1: Attributes and the explanation

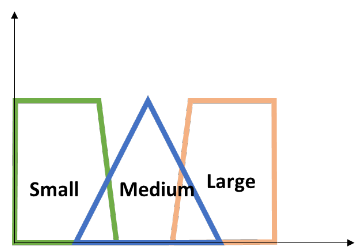
The seven morphological features extracted fed into FIS and trained on 3810 data. The classification accuracy then compared to the bench marked AI classification methods.

2 Methodology

In our study, the multiple FIS has been created and cascaded in 3 level to get the final classification. Each of the FIS consists of 2 input function which is further divided into 3 triangular membership functions and 2 triangular output functions. To fine tune the vertices of the triangular function , genetic algorithm has been used. The GA consists of vertices as well as the rules which defines the ouput of each FIS.

2.1 Fuzzy Inference System (FIS)

Each FIS has been created as two input one output system. Input is designed as combination of 3 membership function- left shoulder, triangle and right shoulder. Output has been just reduced to 2 triangular membership functions to match the classification of rice (in our study there are only 2 varieties). Rule matrix consists of 3 x 3 and min operator is used. Mamdani inference coupled with area center defuzzification method are deployed. A random example can be seen from figure .6



	Mu2 small	Mu2 medium	Mu3 large
Mu1 small	1	2	2
Mu1 medium	1	2	3
Mu1 large	2	3	3

Fig. 2: Membership function and rulebase representation

2.2 Fuzzy cascading

The number of rules required increases exponentially with increase in the inputs. In previous section.2.2 we saw that for 2 input 1 output, we need 3^2 rules (assuming 3 mem function for each input). But if we are taking 7 inputs which will require $3^7 = 2187$ rules. To alleviate the computational energy, we introduce fuzzy cascading or fuzzy trees as showing figure .3. Fuzzy cascading creates 3 levels between output and inputs and reduces the rules to a total of 54. The crisp output obtained from two inputs FIS goes to other FIS as input.

2.3 Genetic algorithm

A genetic algorithm (GA) is a class of evolutionary algorithms that offer a lot of robustness in searching for optimal solutions. GA is inspired from the biological

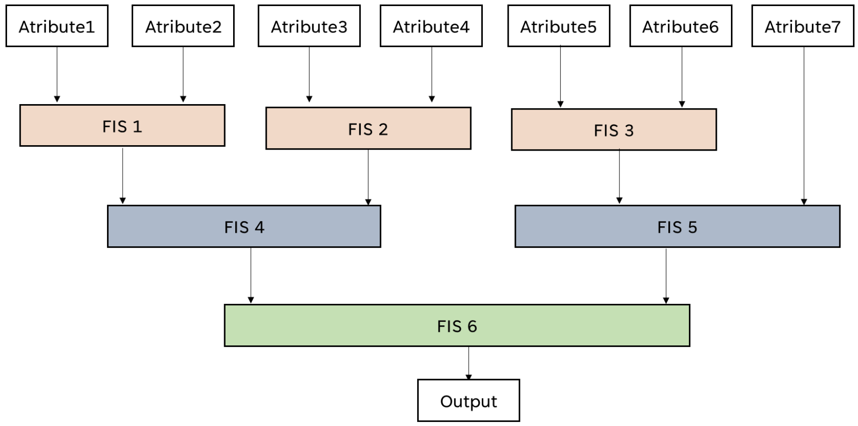


Fig. 3: Fuzzy cascading

evolution which involves reproduction, mutation and survival of the fittest. GA takes random solutions (not the best) and evaluates its fitness (depending on the objective function), then selects the individuals which have higher fitness to reproduce (cross over). The process gets terminated when the maximum generation reaches.
Here in this study we use GA to tune the vertices of input and output membership functions and rules.
The initial population are generated randomly but within the bounds of each attributes. Total of 126 genes are created in each chromosome as shown in figure **??**

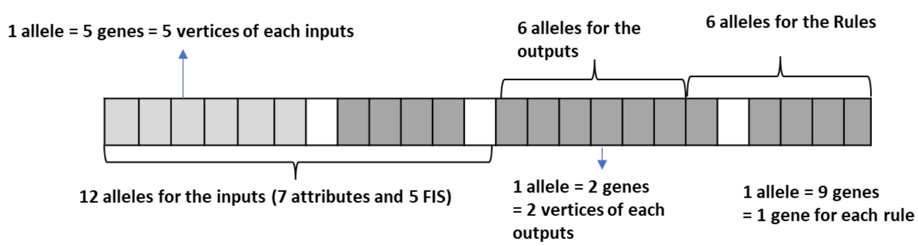


Fig. 4: Chromosome encoding

The 100 chromosomes in single population is then subjected to fitness calculation using the 3810 data we have on two different rice varieties. 80 % of the data set are then passed through the FIS cascading syste generated using each chromosome. The fuzzy ouput of each data then compared to the label (came0 labelled as 1, Osmancik labelled as zero) and reciprocal of the sum of Mean squared error is calculated as fitness of the respective chromosome.

$$f = 1/MSE \tag{1}$$

where MSE is the Mean Squared Error between fuzzy output and actual label

$$MSE = \frac{\sum_{i=1}^{N=3048} (Fuzzy_O utput - Actuallabel)^2}{N} \tag{2}$$

After 100 generation , best gene is selected for the final Fuzzy cascading. Rest of the 20 % data then passed through the current selected fuzzy model to validate the model.

3 Results

In this section, we provide the accuracy and the GA performance for the tuning. Also the final fuzzy model input and output membership functions are plotted. The fitness and the error for each generation of GA is given in figure .**??** and **??** respectively.

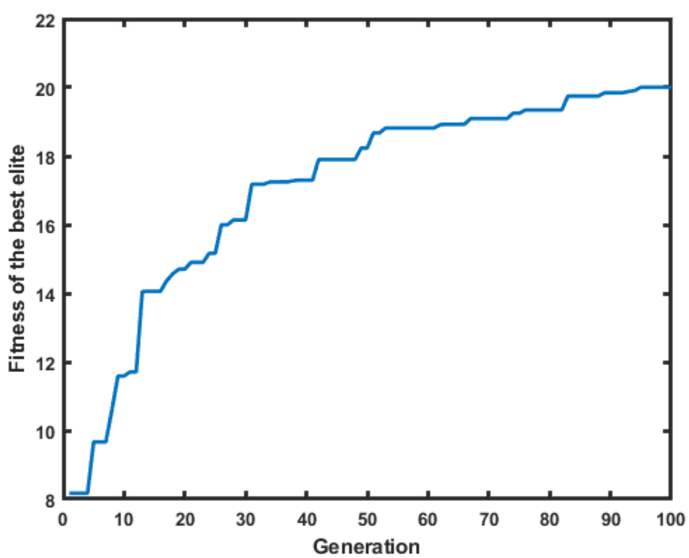


Fig. 5: Fitness vs generation

The final FIS model and its membership functions can be show as in figure 7. The out functions are not shown because of limitation of the space. To check the accuracy, the confusion matrix is generated for both training set and validation set.

The accuracy on training set was found to be 92.81 % and for validation set is 94.36 %.

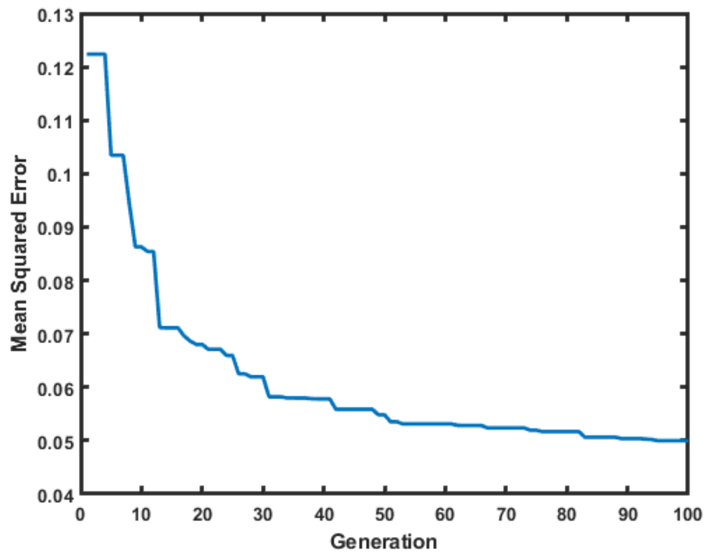


Fig. 6: MSE vs Generation

	Predicted as Cameo	**Predicted as Osmancik**
Labelled as Cameo	2055	125
Labelled as Osmancik	149	1481

Table 2: Confusion matrix for training data (80 % of total data)

	Predicted as Cameo	**Predicted as Osmancik**
Labelled as Cameo	408	18
Labelled as Osmancik	25	312

Table 3: Confusion matrix for validationdata (20 % of total data)

4 Conclusion

The current method of genetic fuzzy system seems to be very efficient and able to produce 94.36 % in validation set in a single epoch. It is important to note that this can be improved by training on more epoch which was done in previous research work.
Compared to other AI technique, it can be concluded that explainability also increase well because user can get some understanding when checking the membership functions and see how big is the dependency of each parameter.

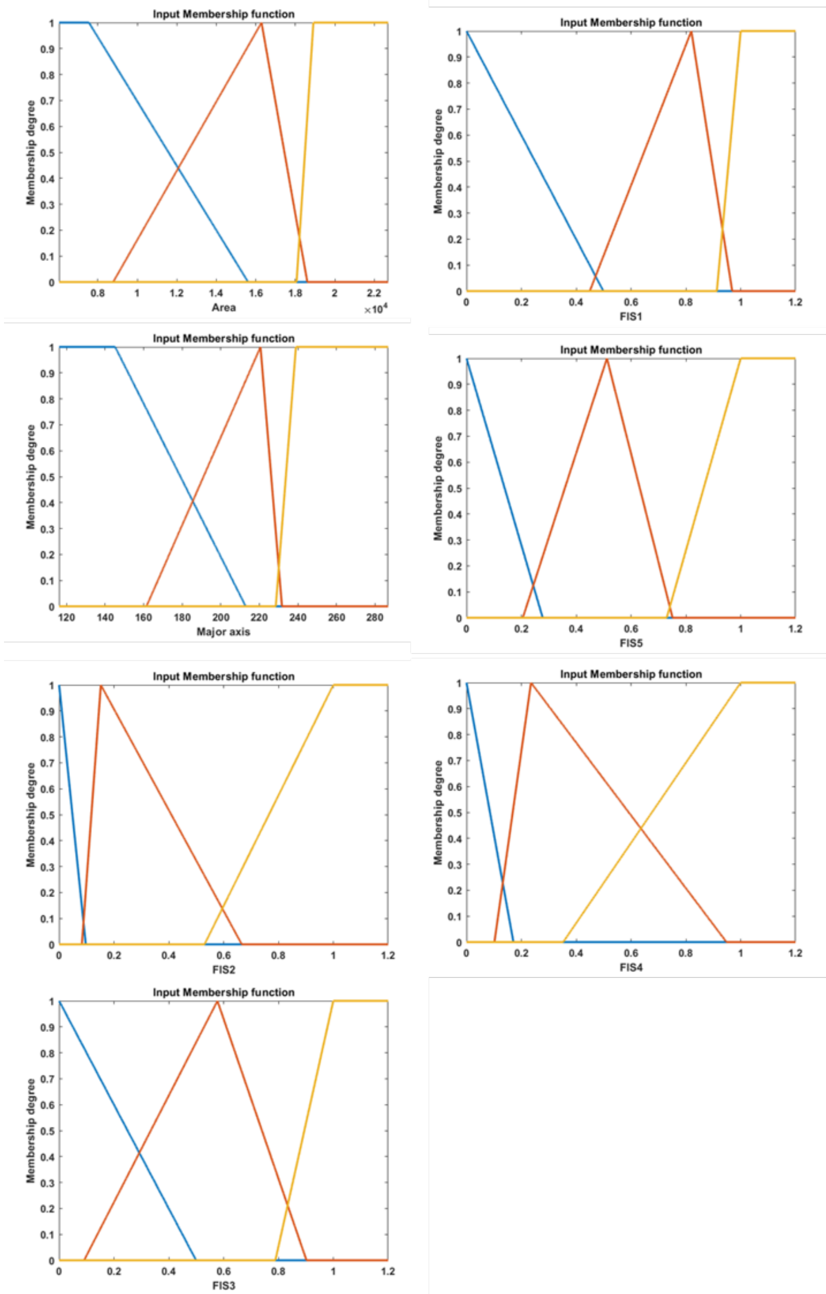


Fig. 7: Membership functions for Best Chromosome

4.1 Challenges

Challenges faced throughout this research works are as follows:

- Data set was huge which increased the overall computational time
- setting up fuzzy cascading system can have multiple combination which we are not considering
- Creating chromosome of length 126 and applying genetic operators.

4.2 Future work

Eventhoug, current work seems to produce astonishing results with the limited range of input membership functions and data, there are lot of room for improvement.

- Run the simulation for multiple epochs with randomizing the training and validation data in each epoch
- Tune genetic parameters such as cross over , mutation probabilities, and elitism factor.
- Increase the number of input functions for each FIS system so that we can upscale the dimension.

References

1. Cinar, Ilkay, and Murat Koklu. *"Classification of Rice Varieties Using Artificial Intelligence Methods."* International Journal of Intelligent Systems and Applications in Engineering 7, no. 3 (2019): 188-194.