Journal of Engineering Science and Technology
Special Issue on ICCSIT 2018, July (2018) 242 - 258
© School of Engineering, Taylor's University

GENETIC ALGORITHM FUZZY LOGIC FOR MEDICAL KNOWLEDGE-BASED PATTERN CLASSIFICATION

CHIN HOOI TAN[1,*], MEI SZE TAN[2], SIOW-WEE CHANG[2], KEEM SIAH YAP[3], HWA JEN YAP[4], SHEN YUONG WONG[5]

[1] Tenaga Nasional Berhad, No 129, Jalan Bangsar, 59200 Kuala Lumpur, Malaysia
[2] Bioinformatics Programme, Institute of Biological Sciences, Faculty of Science, University of Malaya, 50603 Kuala Lumpur, Malaysia
[3] Department of Electronics & Communication Engineering, College of Engineering, Universiti Tenaga Nasional, 43000 Kajang, Selangor, Malaysia
[4] Department of Mechanical Engineering, Faculty of Engineering, University of Malaya, 50603 Kuala Lumpur, Malaysia
[5] Faculty of Engineering, Xiamen University, 43900 Sepang, Selangor, Malaysia
*Corresponding Author: chinhooi.tan@tnb.com.my; vincentchtan@gmail.com

Abstract

Hybrid of genetic algorithm and fuzzy logic in genetic fuzzy system exemplifies the advantage of best heuristic search with ease of understanding and interpretability. This research proposed an algorithm named Genetic Algorithm Fuzzy Logic (GAFL) with Pittsburg approach for rules learning and induction in genetic fuzzy system knowledge discovery. The proposed algorithm was applied and tested in four critical illness datasets in medical knowledge pattern classification. GAFL, with simplistic binary coding scheme using Pittsburg approach managed to exploit the potential of genetic fuzzy inference system with ease of comprehension in fuzzy rules induction in knowledge pattern recognition. The proposed algorithm was tested with three public available medical datasets, which are Wisconsin Breast Cancer (WBC) dataset, Pima Indian Diabetes dataset (PID), Parkinson Disease dataset (PD) and one locally collected oral cancer dataset. The results obtained showed that GAFL outperformed most of the other models that acknowledged from the previous studies. GAFL possessed the advantage of fuzzy rules extraction feature apart from conventional classification technique compared to other models which are lack of fuzzy interpretation. It is easier to interpret and understand fuzzy value in contrast to continuous or range value. GAFL outperformed the other algorithms in terms of accuracy without compromising on interpretability. It is vital to obtain high accuracy in medical pattern recognition especially when dealing with critical illness.

Keywords: Computational intelligence, Fuzzy logic, Genetic algorithm, Genetic fuzzy system, Knowledge discovery, Pattern classification, Rules induction, Rules learning.

1. Introduction

Knowledge pattern recognition is vital in the discipline of knowledge discovery to identify and interpret important or optimum patterns. Knowledge pattern recognition is significant in medical application especially when dealing with critical illness such as cancer, diabetes, Parkinson Disease (PD), hepatitis, AIDS, strokes and others. Early detection and accurate diagnosis of critical illness are key components which greatly increase the chances of successful treatment.

Solving complex problems with hybrid computational intelligence approach is becoming an interesting topic in research community especially for fuzzy rules learning and extraction in Fuzzy Rule-based Systems (FRBS) [1]. It is utmost vital to obtain higher accuracy in medical pattern recognition especially when dealing with critical illness. In this study, an improved algorithm named Genetic Algorithm Fuzzy Logic (GAFL) was developed to perform the medical pattern recognition. Genetic fuzzy rules learning using Pittsburg approach was chosen as the methodology for rules induction in knowledge pattern recognition.

The proposed algorithm was tested on three public available datasets of Wisconsin Breast Cancer (WBC), Pima Indian Diabetes (PID), and PD and one locally collected Oral Cancer dataset in Malaysia. Only the rules for WBC and PID benchmarked models were in fuzzy format, whereas PD and Oral Cancer dataset were quite new and there were no rules available in fuzzy format yet. The objective of this research is to extract rules from critical illness medical datasets to facilitate early detection of critical illness. The study intends to improve generation of knowledge base for rules learning and induction in genetic fuzzy inference system.

In Multi-Objective Evolutionary Fuzzy Systems (MOEFSs), improvement of an objective always leads to deterioration of another parameter, since there are always different trade-offs between accuracy and interpretability [2]. The knowledge discovery from GAFL not only increases the accuracy of prediction in pattern recognition for medical applications but also extracts fuzzy rules from existing benchmarked medical datasets without compromising on interpretability. It manages to strike a balance between accuracy and interpretability. The extracted fuzzy rules are easily understood by a human as there is no lateral displacement for membership functions. The induced fuzzy rules could be applied in new datasets to predict the diagnosis probability of critical illness. A fuzzy value is easier to be understood by a human in contrast to continuous value.

This paper is formatted with an introduction of the motivation of the research work, followed by genetic fuzzy inference system. Proposed method and algorithm are presented in a later section, details of GAFL is elaborated and pseudo-code is included in the subsequent section. Results of the four medical datasets are then discussed, analyzed and benchmarked against other models. The paper concludes with the importance of rules induction in medical engineering.

2. Literature Review

2.1. Genetic fuzzy inference system

There are different categories of genetic Knowledge Base (KB) learning as illustrated in Fig. 1 [1]. A typical genetic KB learning involves genetic rules learning, selection, Database (DB) learning and simultaneous learning of KB components. Generally, genetic fuzzy rules learning involves the learning of existing Rule Base (RB) in FRBS [1, 3-5], which might change existing RB and/or

DB during the adaptive learning. It is reported in several researches [6-12] on fuzzy RB learning that eventually will develop a set of fuzzy rules from predefined DB.

A typical hybrid genetic-fuzzy rules inference system normally deals with RB learning and tuning process FRBS [1, 3, 4]. It is hard to differentiate between the learning and tuning FRBS [1, 3, 4] since they overlap with each other most of the time during the evolutionary optimization process. This study focuses on genetic rules learning that capitalizes on genetic RB learning for prediction in knowledge pattern recognition. The knowledge discovery process involves evolutionary learning and tuning of the RB. The process does not change the DB. It only induces the rules from existing KB after the optimization process.

Genetic algorithm has been widely used for extracting symbolic rules. There are two coding approach for rules learning, which are Pittsburgh approach and Michigan approach. The genetic learning and tuning of RB in this context followed the Pittsburg approach, a set of fuzzy if-then rules are coded as a string and handled as an individual, i.e., "Chromosome = Set of Rules [1, 13, 14] whereby each chromosome represents a set of rules. Each chromosome evolves as a complete RB and it competes with the rest in the evolutionary process. Pittsburg approach has been spearheaded by Thrift for RB learning [1]. Each fuzzy set is codified with an integer including null value with (0) in Pittsburg coding method, it is then referred to a decision table for both antecedents and consequents [1, 14].

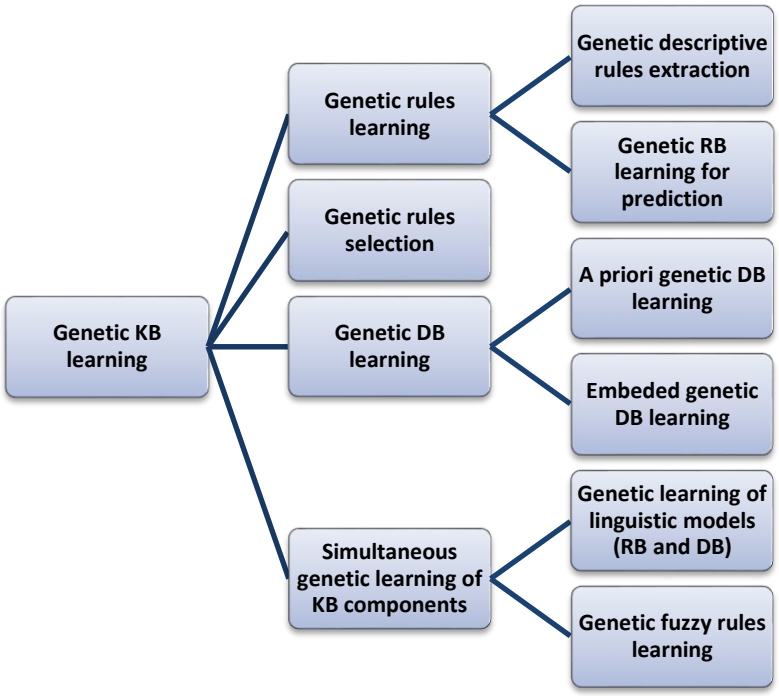

Fig. 1. Taxonomy of genetic knowledge base learning [1].

2.2. Related work in the medical applications

There are some related studies that utilized hybrid fuzzy rules inference system in the medical datasets. In [15], hybrid genetic fuzzy model was used to study the prediction related to heart disease. The rules set generated from the dataset was

selected using Genetic Algorithm (GA) to obtain the most important and relevant rules subset that in used in the fuzzy classifier. The dataset used in this study was obtained from UCI machine learning repository, which consisted of (297) samples and (13) attributes of heart disease cases. The classification accuracy archived was (86%) (specificity= 0.9, sensitivity=0.8) with stratified 10-fold technique.

Dennis and Muthukrishnan [16] proposed an augmented fuzzy system based on Adaptive Genetic Fuzzy Systems (AGFS). The process of AGFS was divided into three steps which firstly optimized the generation of the rule using GA, followed by automatic designing of fuzzy membership function and lastly fuzzy system classification. The proposed algorithm was tested with datasets available from UCI machine learning repository, namely, Cleveland Heart Disease database (76 attributes), Indian Liver data (11 attributes), Mammogram data (6 attributes), Glass data (10 attributes), Wine data (13 attributes), PID data (8 attributes) and Iris data (4 attributes). The proposed algorithm could test quantitatively, qualitatively and comparatively, with the highest accuracy of (89.80%) using PID dataset compared with another existing system.

3. Methodology

3.1. GAFL algorithm

This study presented an innovative GAFL approach of learning and optimizing on the rule base using Pittsburg approach. However, the difference between the GAFL and the Pittsburg approach is that there was no lookup table used in GAFL. Binary coding either (1) or (0) was applied to each antecedent of the fuzzy rules. The consequents of the fuzzy rules were fixed. During the learning and optimization process, binary value 1/0 was applied to each linguistic term of antecedents in RB for every chromosome for each run. The individuals evolved as a complete RB to compete with each other in every generation for the least mean absolute error. Before the evolutionary process began, parameters such as a number of populations, generations, mutation rate, crossover fractions and a number of rules were defined. Crossover fraction and mutation probability were predefined at (0.8) and (0.01) respectively.

During the evolutionary process, the binary string was applied to each of the antecedents of the fuzzy rules. Training data was then fetched and fuzzy predicted outputs of each data were derived. The predicted output was derived using a weighted average defuzzification method. The derivation of the fuzzy predicted output of each data was based on the binary coding of linguistic terms determined during every evolutionary cycle. If the bit is (0), then the respective linguistic term will be discarded and not used in the defuzzification process. Errors of the prediction were counted and mean absolute error was calculated. Moreover, the mean absolute error was selected as the fitness function in this study since all the benchmark datasets dealt with (2) classes, i.e.: positive class (+1, diagnose with the disease) and negative class (-1, normal). When generations were completed, the best offspring derived from the evolutionary process was applied to the predefined fuzzy rules. If all linguistic terms for a particular antecedent binary value are (0), then the antecedent will be ignored in the defuzzification process.

Validation data then are fetched and evaluated. Original targeted class and predicted class value are compared and calculated. After the validation process, testing data are fetched and evaluated. Original targeted class and predicted class value were compared and calculated. For instance, if the targeted class is different

than the predicted class, then one error is counted. Accuracy was calculated and stored for benchmarking. Details of the algorithm is elucidated in Fig. 2.

```
Initialize population, generation, crossover, mutation rate and number of fuzzy rules

While generation not exceeded

        While population not exceeded
                Fetch training data
                        While number of rules not exceeded
                                Apply binary value to each of the antecedents of the fuzzy rules
                        End while

                While number of rules not exceeded
                        Derive the output, ypredict (predicted fuzzy output of fuzzy rules) based on the linguistic
                terms for each antecedent of fuzzy rules determined by genetic algorithm (if the bit is 0, then
                the respective linguistic term will not be used in defuzzification process)

                        Evaluate the fitness function (mean absolute error)
                End while
        End while

        Perform crossover and mutation
End while

Apply the best offspring returned from genetic algorithm process to each linguistic term of antecedents of the fuzzy rules

Fetch training data
While number of rules not exceeded
        Derive ypredict based on the generated fuzzy rules returned from genetic algorithm process (if all linguistic
        terms for an antecedent binary value are 0, then the antecedents will be discarded in defuzzification process)
        Evaluate mean absolute error for the training data
        Evaluate accuracy for the training data
End while

Fetch validation data
While number of rules not exceeded
        Derive ypredict based on the generated fuzzy rules returned from genetic algorithm process
        Evaluate mean absolute error for the validation data
        Evaluate accuracy for the validation data
End while

Fetch testing data
While number of rules not exceeded
        Derive ypredict based on the generated rules returned from genetic algorithm process
        Evaluate the mean absolute error for the testing data (y-ypredict, i.e.: y is the original targeted class)
        Evaluate and store accuracy for the testing data
        End while
```

Fig. 2. Pseudo-code for the proposed algorithm.

3.2. Fitness function

The process of defining fitness measurement function in genetic fuzzy inference system is problem dependent. Mean square error or mean absolute error is commonly used in prediction and estimation of the related problems [17]. In this study, Mean Absolute Error (MAE) was selected as the fitness function during the evolutionary process. Furthermore, during the rules induction process, the objective is to minimize the error for each consequent. MAE is an estimation error function used to quantify the difference between the original targeted fuzzy value and predicted fuzzy value [15]. It is illustrated in Eq. (1) that S is the fuzzy system to be evaluated, *N* represents sample data size and yi: ypredicti on which the i represents a targeted-estimated pair of the particular rule.

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left(S(|y^i - ypredict^i|)\right) \quad (1)$$

For instance, if a targeted output is (7) but the estimated output is (6), and if a targeted output is (2) but the estimated output is (1), it would generate the same absolute error of (1) and each would have the same contribution to the error function. However, (6) is a better estimation of (7) than (1) is of (2). Thus, using MAE to calculate the fitness would produce better accuracy for large target outputs than for small ones [18].

3.3. Testing of GAFL and datasets

In this study, three publicly available datasets and one locally collected dataset were used for GAFL performance. WBC [19], PID [20] and PD [21] are downloaded from the UCI Machine Learning Repository, Center for Machine Learning and Intelligent Systems. The datasets have been widely used by researchers as a primary source of machine learning dataset to study the performance on small-sized dataset, GAFL was tested with the locally collected oral cancer dataset that was collected from the Malaysia Oral Cancer Database and Tissue Bank System (MOCDTBS) coordinated by the Oral Cancer Research and Coordinating Centre (OCRCC), Faculty of Dentistry, University of Malaya [22, 23]. There are three performance indicators reported in [24] which are commonly used in the medical diagnosis/prognosis classification. They are illustrated in Eq. 2, 3 and 4. Another non-parametric statistical parameter used for estimation was a bootstrap method whereby the underlying sampling distribution is unknown or difficult to estimate [25]. Furthermore, Bootstrap mean was applied for accuracy comparison.

$$Accuracy = 100\% \times \text{Number of correctly classified cases/Total number of cases} \quad (2)$$

$$Sensitivity = 100\% \times \text{Number of correctly classified positives cases/Total number of positive cases} \quad (3)$$

$$Specificity = 100\% \times \text{Number of correctly classified negative cases/Total number of negative cases} \quad (4)$$

The results of GAFL with WBC and PID were presented under two experiments: i) rules induction models (EGART-FIS and MFM-FIS) accuracy comparison ii) other models (MFMM, C4.5, C4.5 rules ITI, LMDT, CN2, LVQ, OCI, and Nevprop) results benchmarking. To the best of our knowledge, there is no related study in rule induction that was conducted using PD dataset. Therefore, this dataset is tested and benchmarked with other non-rules induction classification methods to verify the feasibility of GAFL. The models included Neural Network, DMNeural, Regression and Decision trees. Similarly, there is no related rule induction applied to oral cancer dataset as well. The result of other models Artificial Neural Network (ANN), Support Vector Machine (SVM), Logistic Regression (LR), Adaptive Neuro-Fuzzy Inference System (ANFIS), Genetic Programming (GP) using the same oral cancer dataset is then compared with the results of GAFL.

4. Results

4.1. Wisconsin breast cancer and pima Indian diabetes

The WBC dataset consists of (699) samples which include (458 samples of benign and (241) samples of breast cancer [19]. The experimental set up followed the procedure reported in [26]. The dataset was divided into (50% (350)), (30% (209)), and (20% (140)) for training, validation, and test, respectively. The PID dataset

consists of (768) samples in which (268) samples are classified as diabetic and the remainder as non-diabetic [20]. There were two experiments conducted for PID dataset. The first experimental set followed the procedure reported in [27, 28]. The dataset was divided into three subsets, ((348) (50%)) samples were allocated for training, (192 (25%)) samples for validation, and the remainder (192 (25%)) samples for testing.

In experiment (i), GAFL result was compared with EGART-FIS [29] and MFMM-FIS [28, 30]. Out of (50) runs, the one with the best classification rate was selected for fuzzy rule extraction with seven quantization levels, listed as.: extremely low (1) to extremely high (7). The fuzzy rules extracted from WBC and PID are presented in Tables 1 and 2 respectively. These rules can be interpreted in terms of the IF-THEN format as well as the equivalent fuzzy equation. An example with explanation is shown in Supplementary Table S1 for WBC and Supplementary Table S2 for PID.

Table 1. The rules extracted from the algorithm with quantization level *Q*= 7 for WBC data set.

	IF attributes is									**THEN** class is
	i	ii	iii	iv	v	vi	vii	viii	ix	
RULE 1	1-2,4-6	2,4,6	1-2,4-5	1-3,6	2,5-6	1	2,5,7	1,3-4	1,3	-1
RULE 2	1-4,6-7	3-4,6	2-4,7	1-2,5-6	1-3,6	1-2	7	1,3-4,6-7	1,4	-1
RULE 3	2-3,5-7	4-7	2,4-7	1-2,4-5,7	1,4,6	1,3-7	1,4,6	1,3-6	1,3,6-7	+1

Quantization Level: 1(extremely low); 2(very low); 3(low); 4(nominal); 5(high); 6(very high); 7(extremely high). Classes: −1 (benign); +1 (malignant).

Table 2. Rules extracted from the algorithm with quantization level *Q*= 7 for PID data set.

	IF attributes is								**THEN** class is
	i	ii	iii	iv	v	vi	vii	viii	
RULE 1	2-5,7	4	3-5	2-4	1-2,5-7	1-3	1-2,7	2,6-7	-1
RULE 2	1,3,5,7	2	6	1,5,7	2,4-6	1,5,7	3,6	2,4-6	-1
RULE 3	2-5,7	6-7	1-4,6-7	1-2,7	-	4-7	3-6	3	+1

Quantization Level: 1(extremely low); 2(very low); 3(low); 4(nominal); 5(high); 6(very high); 7(extremely high). Classes: −1 (non-diabetic); +1 (diabetic).

The performance indicators result in accuracy, sensitivity, and specificity for WBC are (96.35%), (97.3%) and (83.5%) respectively as depicted in Table 3. Whereas, accuracy, sensitivity and specificity parameter for PID dataset is (78.12%), (77.87%), and (75.32%) respectively. Moreover, as compared with other rules induction models, GAFL outperformed other two algorithms for rules induction with bootstrap mean average accuracy of (96.35%) for WBC and (78.03%) for PID with 3 rules respectively. The results are illustrated in Table 4.

In experiment (ii), GAFL was compared with other non-rules induction models. The classification results showed GAFL achieved a comparatively better accuracy if compared to results generated from other models [31] as illustrated in Table 5.

Table 3. Accuracy, sensitivity and specificity parameters for WBC and PID dataset.

Parameter	Result (%)	
	WBC	**PID**
Accuracy	96.34	78.03
Sensitivity	97.30	77.87
Specificity	83.50	75.32

Table 4. Comparison of rules induction accuracy comparison.

Model	WBC		PID	
	No. of Rules	Average Accuracy (%)	No. of Rules	Average Accuracy (%)
GAFL (bootstrap mean)	3	96.35	3	78.03
EGART-FIS	3	93.56	6	73.05
MFMM-FIS	4	92.56	5	72.92

Table 5. Benchmarking with other models for WBC and PID dataset.

Model	Average accuracy (%)	
	WBC	PID
GAFL (bootstrap mean)	96.35	78.03
MFMM (single run)	96.42	71.02
C4.5	94.25	71.55
C4.5 rules	94.68	73.16
ITI	91.14	73.51
LMDT	95.75	72.19
CN2	94.39	71.28
LVQ	94.82	50.00
OCI	93.24	68.52
Nevprop	95.05	78.12

4.2. Parkinson disease

Parkinson disease dataset consists of (195) samples from two classes, with., a total of (48) samples from patients without PD (Class -1) and (147) samples from patients with PD (Class +1) respectively [21]. The experimental set up in this research followed the exact procedure in [18], the dataset was randomly partitioned into training and testing dataset. (65%) of the input, dataset was used for training and the rest of the dataset was used for testing. Out of (50) runs, the accuracy rate was fed into a bootstrap method with (1000) resampling's to obtain the bootstrap mean of the accuracy rates. The run with the best classification rate was selected for fuzzy rule extraction with seven quantization levels. The extracted fuzzy rules are illustrated in Table 6. Supplementary Table S3 showed an example in the interpretation of the Rule and the respective fuzzy equation that extracted from GAFL. The rules can be interpreted in terms of the IF-THEN format as well as the equivalent fuzzy equation.

The average accuracy, sensitivity, and specificity are (96.23%), (95.96%) and (94.28%) respectively, as shown in Table 7. There is no rules induction related study that used PD as the dataset, therefore, for benchmark purposes, GAFL result for PD was compared with other non-rules induction classification models (Neural Network, DMNeural, Regression and Decision trees). The comparison results were shown in Table 8. The results showed that GAFL was comparable with the neural network while outperformed the rests.

Table 6. Rules extracted from the algorithm with quantization level *Q*= 7 for PD dataset.

	IF attributes is											THEN class is
	i	ii	iii	iv	v	vi	vii	viii	ix	x	xi	
RULE 1	1,7	2,4,6	3,5-6	1,3,6	1-2,5-7	1,3,6-7	2-3,6	2,4-6	1	2-3,7	1-2,4,6	-1
RULE 2	2-3,7	1,6	-	2	1,3,5-7	1-2,4-6	1,3-6	1,3-4,6	1,4-6	1,5,7	1-2,5-6	+1
RULE 3	2-5,7	1,3-7	2,5,7	1,4,6-7	1-2,4,7	3,7	2,4	2-3,5,7	1-4,6	4-7	1,3	+1
RULE 4	1,3,5	1,3,5-7	1,3-4,6	1,4,6	5,7	1-6	4-5	3-6	2,5-7	2,5,7	1-4	+1
	xii	xiii	xiv	xv	xvi	xvii	xviii	xix	xx	xxi	xxii	
RULE 1	1,5	1-3,5	1-2,4	1-2,4,7	1-3,5	2,4-7	1,3-4,7	1-2,5	2,4-7	1,4-5	1,3,5,7	-1
RULE 2	1,3,6	1-2,5,7	1-2,5,7	1-2,4,6	1-3,5-7	1-3,5,7	1,3-7	4-7	3,5-6	2-4	2,4	+1
RULE 3	1,3-7	2,5	2-3,6	3-4,6	2-6	3,5,7	2,4-5,7	2-3	1,6-7	1-2,4-7	4,6-7	+1
RULE 4	2-3	1,6-7	2-5	3-4,6-7	3,6	1,4	2-5,7	1,3,5-6	1-2,6	2-4	3-7	+1

Quantization Level: 1(extremely low); 2(very low); 3(low); 4(nominal); 5(high); 6(very high); 7(extremely high). Classes: −1 (absence); +1 (presence).

Table 7. Accuracy, sensitivity and specificity parameters of PD.

Parameter	Result (%)
Accuracy	96.23
Sensitivity	95.96
Specificity	94.28

Table 8. Benchmarking with other models for PD.

Method	Average accuracy (%)
GAFL (bootstrap mean)	96.23
Neural network	96.45
DMNeural	86.95
Regression	88.80
Decision tree	88.95

4.3. Oral cancer dataset

The oral cancer dataset consists of (31) oral cancer cases which were collected from the Malaysia Oral Cancer Database and Tissue Bank System (MOCDTBS). The dataset was acquired through the procedure as described in [22]. Table 9 lists the (17) features in the dataset. Each case of oral cancer was followed up for 3-year from the date of diagnosed. At the end of 3 years, the outcome of each case is either dead or alive.

This dataset was tested in the validation step to confirm the practicality of GAFL by using the locally collected samples. Similar to the previous experiment, among the (10) runs, the one that shows the best accuracy was selected for fuzzy rule extraction with four quantization levels. Table 10 records the fuzzy rules extracted and an example in the interpretation of the Rule and the respective fuzzy equation that extracted from GAFL was demonstrated in Supplementary Table S4. The rules can be interpreted in terms of the IF-THEN format as well as the equivalent fuzzy equation. With the ratio of 80:20 for training and testing set of the

dataset, GAFL successfully resulted in the consistent accuracy of (87.1%) in 10 runs. The results on average accuracy, specificity and sensitivity were (87.1%), (100%) and (85.00%) respectively as recorded in Table 11.

Besides that, the average accuracy result of GAFL in oral cancer was compared with the results from other common and non-rules induction classification models for benchmarking purposes. The benchmarking comparison is shown in Table 12.

Table 9. Features available in the oral cancer prognosis dataset.

Feature			
Age	*(Age)*	**Pattern of Invasion**	*(Inv)*
Ethnicity	*(Eth)*	**Nodes**	*(Nodes)*
Gender	*(Gen)*	**PT**	*(PT)*
Smoke	*(Smo)*	**PN**	*(PN)*
Drink	*(Dri)*	**Stage**	*(Sta)*
Chew	*(Chew)*	**Size**	*(Size)*
Site	*(Site)*	**Treatment**	*(Tre)*
Histological differentiation of SCC	*(Diff)*	**p53**	*(p53)*
		p63	*(p63)*

Table 10. Rules extracted from the algorithm with quantization level *Q*= 4 for oral cancer data set.

	IF attributes is								**THEN** class is
	i	ii	iii	iv	v	vi	vii	viii	
RULE 1	1-3	2,3	2-4	1,3	1,3	2,3	3	1	0
RULE 2	4	1,3	1-3	3	3,4	4	1-4	2,4	0

Classes: 0 (Survival); 1 (Dead)

Table 11. Accuracy, sensitivity and specificity parameters of oral cancer.

Parameter	**Result (%)**
Accuracy	87.10
Sensitivity	100.00
Specificity	85.00

Table 12. Benchmarking with other models for oral cancer.

Method	**Average accuracy (%)**
GAFL (bootstrap mean)	87.10
SVM-GA [28]	74.76
ANN [28]	84.62
LR [28]	74.76
ANFIS [28]	93.81
GP [29]	83.87

5. Discussion

From Tables 3, 4, and 7, it can be seen that GAFL yielded higher average accuracy results compared to other models in both of the experiments (rules induction and non-rules induction models' comparison). For every table mentioned above, the

average accuracy obtained using each of the respective datasets is consistent (>75%). Thus, it is fair to infer that GAFL is reliable and consistent with rules learning and extraction. In Table 12, the average accuracy obtained by GAFL is 87.1% which positioned the purposed model as the second-best result compared to ANFIS model. As the size of the oral cancer dataset is small and due to nature of the data, not all of the features in the original oral cancer dataset (number of features= 17) is suitable to be scaled quantitatively, the size of the dataset becomes even smaller when compared to the previous study. Hence, only 8 features (Age, Ethnicity, Site, PT, PN, Stage, Size and Treatment) were tested in GAFL.

By referring to this reason, the result obtained by GAFL is comparable to the other common classification methods and performing reliably with such a small-sized local dataset. The motivation of GAFL is somehow similar to FURIA [32] that was developed for fuzzy rules induction. Nevertheless, FURIA cannot develop rules from raw data. There is a novel rule stretching technique applicable to FURIA and the idea is to generalize the existing rules until they cover all examples [30]. On the other hand, GAFL can develop rules from sample data and can be used to predict the output based on new inputs.

Besides, the implementation of GAFL is somehow different compared to fuzzy association rule extraction for classification model (FARC-HD) [33] which involves a search tree and subgroup discovery with candidate rule prescreening. This involves rule selection and lateral tuning for FARC-HD [33] where there is lateral displacement as the linguistic label is moving between its two lateral labels. Hence, the parameter of membership function requires tuning. As a result, the positioning of membership function does not guarantee human interpretation, which is leftmost is very small, the center is medium, the rightmost is very big. Nevertheless, there is no membership function tuning required for GAFL and it directly constructs fuzzy rules from the optimization process. It is also worthwhile to mention that the rules induction algorithm, implementation procedure and experimental set up reported in [31] are different than what is implemented in GAFL.

6. Conclusions

In this research, the average accuracy obtained by GAFL using WBC, PID, PD and Oral Cancer dataset is (96.35%), (78.03%), (96.23%) and (87.1%) respectively. As a comparison, GAFL outperformed most of the other models or algorithms in terms of accuracy without compromising on interpretability and it managed to strike a balance between accuracy and interpretability. It is vital to obtain higher accuracy in medical knowledge pattern recognition, especially when dealing with critical illness. GAFL possesses the advantage of fuzzy rules extraction feature apart from the conventional classification technique compared to other models that lack fuzzy interpretation. Moreover, the extracted fuzzy rules can be directly interpreted by a human without machine intervention since the membership functions for each linguistic variable are fixed.

Besides, it is easier to interpret and understand the fuzzy value in contrast to continuous or range value. Although this research is focused on the medical rules induction and pattern recognition, it is believed that GAFL can be applied to other areas of interest for pattern classification and rules extraction.

This research used binary classification for critical illness; nevertheless, GAFL is expected to be extended and improved to cater for multi-classification problems in medical pattern recognition or other domains in future work. It is

vital to extend GAFL to multi-class critical illness pattern recognition for early detection and treatment.

SUPPLEMENTARY MATERIALS

- Supplementary Table S1: Interpretation of Rule 1 and the respective fuzzy equation extracted from GAFL (WBC dataset).
- Supplementary Table S2: Interpretation of Rule 1 and the respective fuzzy equation extracted from GAFL (PID dataset).
- Supplementary Table S3: Interpretation of Rule 4 and the respective fuzzy equation extracted from GAFL (Parkinson dataset).
- Supplementary Table S4: Interpretation of Rule 1 and the respective fuzzy equation extracted from GAFL (Oral Cancer dataset).

Acknowledgment

This work was supported by the University of Malaya, UMRG Programme with grant number RP038A-15AET, RP038C-15AET, Universiti Tenaga Nasional with grant number J510050684, and funded by Tenaga Nasional Berhad.

Abbreviations

AGFS	Adaptive Genetic Fuzzy Systems
ANFIS	Adaptive Neuro-Fuzzy Inference System
ANN	Artificial Neural Network
DB	Database
EGART-FIS	Enhanced Generalized Adaptive Resonance Theory-Fuzzy Inference System
FARC-HD	fuzzy association rule extraction for classification model
FRBS	Fuzzy Rule-based Systems
GA	Genetic Algorithm
GAFL	Genetic Algorithm Fuzzy Logic
GP	Genetic Programming
ITI	Incremental Tree Inducer
KB	Knowledge Base
LMDT	Linear Machine Decision Trees
LR	Logistic Regression
LVQ	Learning Vector Quantization
MAE	Mean Absolute Error
MFM-FIS	Membership function modification-Fuzzy Inference System
MFMM	Modified **f**uzzy min–max
MOCDTBS	Malaysia Oral Cancer Database and Tissue Bank System
MOEFSs	Multi-Objective Evolutionary Fuzzy Systems
Nevprop	Nevada backpropagation
OC1	Oblique Classifier
OCRCC	Oral Cancer Research & Coordinating Centre
PD	Parkinson Disease dataset
PID	Pima Indian Diabetes dataset
RB	Rule Base
SVM	Support Vector Machine
WBC	Wisconsin Breast Cancer

References

1. Herrera F. (2008). Genetic fuzzy systems: taxonomy, current research trends and prospects. *Evolutionary Intelligence*, 1(1), 27-46.
2. Fazzolari, M.; Alcala, R.; Nojima, Y.; Ishibuchi, H.; and Herrera, F. (2013). A review of the application of multiobjective evolutionary fuzzy systems: Current status and further directions. *IEEE Transactions on Fuzzy systems*, 21(1), 45-65.
3. Cordón, O.; Gomide, F.; Herrera, F.; Hoffmann, F.; and Magdalena, L. (2004). Ten years of genetic fuzzy systems: current framework and new trends. *Fuzzy sets and systems*, 141(1), 5-31.
4. Alcalá, R.; Alcalá-Fdez, J.; Herrera, F.; and Otero, J. (2007). Genetic learning of accurate and compact fuzzy rule-based systems based on the 2-tuples linguistic representation. *International Journal of Approximate Reasoning*, 44(1), 45-64.
5. Tan, C.H.; Yap, K.S.; Ishibuchi, H.; Nojima, Y.; and Yap, H.J. (2014). Application of fuzzy inference rules to early semi-automatic estimation of activity duration in software project management. *IEEE Transactions on Human-Machine Systems*, 44(5), 678-688.
6. Filipič, B.; and Juričić, Đ. (1996). A genetic algorithm to support learning fuzzy control rules from examples.
7. Simon, D. (2002). Sum normal optimization of fuzzy membership functions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(04), 363-84.
8. Casillas, J.; Cordón, O.; Herrera, F.; and Villar, P. (2004). A hybrid learning process for the knowledge base of a fuzzy rule-based system. *Proceedings of the 2004 international conference on information processing and management of uncertainty in knowledge-based systems.* Perugia, Italy, 2189-2196.
9. Cordón, O.; Herrera, F.; and Villar, P. (2001). Generating the knowledge base of a fuzzy rule-based system by the genetic learning of the data base. *IEEE Transactions on fuzzy systems*, 9(4), 667-74.
10. Cordón, O.; Herrera, F.; Magdalena, L.; and Villar, P. (2001). A genetic learning process for the scaling factors, granularity and contexts of the fuzzy rule-based system data base. *Information Sciences*, 136(1-4), 85-107.
11. Pedrycz, W. (2004). Associations and rules in data mining: A link analysis. *International Journal of Intelligent Systems*, 19(7), 653-70.
12. Teng, Y.W.; and Wang, W.J. (2004). Constructing a user-friendly ga-based fuzzy system directly from numerical data. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 34(5), 2060-2070.
13. Tan, C.H.; Yap, K.S.; and Yap, H.J. (2012). Application of genetic algorithm for fuzzy rules optimization on semi expert judgment automation using Pittsburg approach. *Applied Soft Computing*, 12(8), 2168-77.
14. Smith, S.F. (1980). *A learning system based on genetic adaptive algorithms*. Ph.D. Dissertation. ACM, University of Pittsburgh, Pittsburgh, PA, USA. AAI8112638.
15. Santhanam, T.; and Ephzibah, E.P. (2015). Heart disease prediction using hybrid genetic fuzzy model. *Indian Journal of Science and Technology*, 8(9), 797-803.

16. Dennis, B.; and Muthukrishnan, S. (2014). AGFS: Adaptive Genetic Fuzzy System for medical data classification. *Applied Soft Computing*, 25, 242-52.
17. Shi, Y.; Eberhart, R.; and Chen, Y. (1999). Implementation of evolutionary fuzzy systems. *IEEE Transactions on fuzzy systems*, 7(2), 109-19.
18. Das, R. (2010). A comparison of multiple classification methods for diagnosis of Parkinson disease. *Expert Systems with Applications*, 37(2), 1568-72.
19. Wisconsin Breast Cancer dataset (1992). Retrieved November 30, 2017, from http://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer-wisconsin/breast-cancer-wisconsin.data.
20. Pima Indians Diabetes dataset (1990). Retrieved February 15, 2017, from http://archive.ics.uci.edu/ml/machine-learning-databases/pima-indians-diabetes/pima-indians-diabetes.data.
21. UCI Machine Learning Repository (2008). Parkinson Disease dataset. Retrieved February 15, 2017, from https://archive.ics.uci.edu/ml/datasets/ parkinsons
22. Chang, S.W.; Abdul-Kareem, S.; Merican, A.F.; and Zain, R.B. (2013). Oral cancer prognosis based on clinicopathologic and genomic markers using a hybrid of feature selection and machine learning methods. *BMC bioinformatics*, 14(1), 170.
23. Tan, M.S.; Tan, J.W.; Chang, S.W.; Yap, H.J.; Kareem, S.A.; and Zain, R.B. (2016). A genetic programming approach to oral cancer prognosis. *PeerJ*, 4, e2482.
24. Efron, B. (1981). Nonparametric standard errors and confidence intervals. *canadian Journal of Statistics*, 9(2), 139-58.
25. Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7(1), 1-26.
26. Quteishat, A.; and Lim, C.P. (2008). A modified fuzzy min–max neural network with rule extraction and its application to fault detection and classification. *Applied Soft Computing*, 8(2), 985-95.
27. Tan, S.C. (2008). Classification and rule extraction using artificial neural network. *M.Sc*. University Sains Malaysia, Penang, Malaysia.
28. Lekkas, S.; and Mikhailov, L. (2010). Evolving fuzzy medical diagnosis of Pima Indians diabetes and of dermatological diseases. *Artificial Intelligence in Medicine*, 50(2), 117-26.
29. Yap, K.S.; and Lim, CP., Mohamad-Saleh J. (2010). An enhanced generalized adaptive resonance theory neural network and its application to medical pattern classification. *Journal of Intelligent and Fuzzy Systems*, 21(1, 2), 65-78.
30. Lekkas, S.; and Mikhailov, L. (2009). Breast cancer diagnosis based on evolvable fuzzy classifiers and feature selection. *Applications and Innovations in Intelligent Systems XVI*. Springer, London, 185-195.
31. Eklund, P.W.; and Hoang, A. (2006). A comparative study of public domain supervised classifier performance on the UCI database. University of Wollongong. Australian *Journal of Intelligent Systems*, 9(1),1–47.
32. Hühn, J.; and Hüllermeier, E. (2009). FURIA: an algorithm for unordered fuzzy rule induction. *Data Mining and Knowledge Discovery*, 19(3), 293-319.
33. Alcala-Fdez, J.; Alcala, R.; and Herrera, F. (2011). A fuzzy association rule-based classification model for high-dimensional problems with genetic rule selection and lateral tuning. *IEEE Transactions on Fuzzy Systems*, 19(5), 857-72.

Supplementary Table S1

Interpretation of Rule 1 and the respective fuzzy equation extracted from GAFL (WBC dataset)

IF	Clump Thickness (CT) is from extremely low to very low or nominal to very high
and	Uniformity of Cell Size (UCS) is very low or nominal or very high
and	Uniformity of Cell if Cell Shape (UCCS) is extremely low to very low or high to very high
and	Marginal Adhesion (MA) is extremely low to low or very high
and	Single Epithelial Cell Size (SPCS) is from very low or high to very high
and	Bare Nuclei (BN) is from extremely low
and	Bland Chromatin (BC) is from very low or high or extremely high
and	Normal Nucleoli (NN) is extremely low or low to nominal
and	Mitoses (MT) is extremely low or low
THEN	Benign

Firing strength of Rule 1 is

$w1 = [\mu_1(CT) \vee \mu_2(CT) \vee \mu_4(CT) \vee \mu_5(CT) \vee \mu_6(CT)] \wedge [\mu_2(UCS) \vee \mu_4(UCS) \vee \mu_6(UCS)]$

$\wedge [\mu_1(UCCS) \vee \mu_2(UCCS) \vee \mu_4(UCCS) \vee \mu_5(UCCS)] \wedge [\mu_1(MA) \vee \mu_2(MA) \vee \mu_3(MA) \vee \mu_6(MA)]$

$\wedge [\mu_2(SPCS) \vee \mu_5(SPCS) \vee \mu_6(SPCS)] \wedge [\mu_1(BN)] \wedge [\mu_2(BC) \vee \mu_5(BC) \vee \mu_7(BC)]$

$\wedge [\mu_1(NN) \vee \mu_3(NN) \vee \mu_4(NN)] \wedge [\mu_1(MT) \vee \mu_3(MT)]$

Constant for the output variable of Rule 1 is $f1 = -1$

Supplementary Table S2

Interpretation of Rule 1 and the respective fuzzy equation extracted from GAFL (PID dataset)

IF	Number of Times of Pregnant (NTP) is from very low to high or extremely high
and	Plasma Glucose Concentration (PGC) is nominal
and	Diastolic Blood Pressure (DBP) is low to high
and	Triceps Skin Fold Thickness (TSFT) is very low to nominal
and	Two Hours Serum Insulin (THSI) is from extremely low to very low or high to extremely high
and	Body Mass Index (BMI) is from extremely low to low
and	Diabetes Pedigree Function (DPF) is from extremely low to very low or extremely high
and	Age (AGE) is very low or very high to extremely high
THEN	Diagnose of diabetes is negative (non-diabetic)

Firing strength of Rule 1 is

$w1 = [\mu_2(NTP) \vee \mu_3(NTP) \vee \mu_4(NTP) \vee \mu_5(NTP) \vee \mu_7(NTP)] \wedge [\mu_4(PGC)] \wedge [\mu_3(DBP) \vee \mu_4(DBP) \vee \mu_5(DBP)]$

$\wedge [\mu_2(TSFT) \vee \mu_3(TSFT) \vee \mu_4(TSFT)] \wedge [\mu_1(THSI) \vee \mu_2(THSI) \vee \mu_5(THSI) \vee \mu_6(THSI) \vee \mu_7(THSI)]$

$\wedge [\mu_1(BMI) \vee \mu_2(BMI) \vee \mu_3(BMI)] \wedge [\mu_1(DPF) \vee \mu_2(DPF) \vee \mu_7(DPF)] \wedge [\mu_2(AGE) \vee \mu_6(AGE) \vee \mu_7(AGE)]$

Constant for the output variable of Rule 1 is $f1 = -1$

Supplementary Table S3

Interpretation of Rule 4 and the respective fuzzy equation extracted from GAFL (Parkinson dataset)

IF		MDVP:Fo(Hz) is extremely low or low or high
	and	MDVP:Fhi(Hz) is extremely low or low or from high to extremely high
	and	MDVP:Flo(Hz) is extremely low or low or nominal or very high
	and	MDVP:Jitter(%) is extremely low or nominal or very high
	and	MDVP:Jitter(Abs) is high or extremely high
	and	MDVP:RAP is from extremely low to very high
	and	MDVP:PPQ is nominal or high
	and	Jitter:DDP is from low to very high
	and	MDVP:Shimmer is low or from high to extremely high
	and	MDVP:Shimmer(dB) is low or high or extremely high
	and	Shimmer:APQ3 is from extremely low to nominal
	and	Shimmer:APQ5 is very low or low
	and	MDVP:APQ is extremely low or very high or extremely high
	and	Shimmer:DDA is from very low to high
	and	NHR is low or nominal or very high or extremely high
	and	HNR is low or very high
	and	RPDE is extremely low or nominal
	and	DFA is from very low to high or extremely high
	and	Spread1 is extremely low or low or high or very high
	and	Spread2 is extremely low or low or very high
	and	D2 is from very low to nominal
	and	PPE is from low to extremely high
THEN		Presence with Parkinson Disease

Firing strength of Rule 4 is

$w4 = [\mu_1(\text{MDVP:Fo(Hz)}) \vee \mu_3(\text{MDVP:Fo(Hz)}) \vee \mu_5(\text{MDVP:Fo(Hz)})]$
$\wedge [\mu_1(\text{MDVP:Fhi(Hz)}) \vee \mu_3(\text{MDVP:Fhi(Hz)}) \vee \mu_5(\text{MDVP:Fhi(Hz)}) \vee \mu_6(\text{MDVP:Fhi(Hz)}) \vee \mu_7(\text{MDVP:Fhi(Hz)})]$
$\wedge [\mu_1(\text{MDVP:Flo(Hz)}) \vee \mu_3(\text{MDVP:Flo(Hz)}) \vee \mu_4(\text{MDVP:Flo(Hz)}) \vee \mu_6(\text{MDVP:Flo(Hz)})]$
$\wedge [\mu_1(\text{MDVP:Jitter(\%)}) \vee \mu_4(\text{MDVP:Jitter(\%)}) \vee \mu_6(\text{MDVP:Jitter(\%)})]$
$\wedge [\mu_5(\text{MDVP:Jitter(Abs)}) \vee \mu_7(\text{MDVP:Jitter(Abs)})]$
$\wedge [\mu_1(\text{MDVP:RAP}) \vee \mu_2(\text{MDVP:RAP}) \vee \mu_3(\text{MDVP:RAP}) \vee \mu_4(\text{MDVP:RAP}) \vee \mu_5(\text{MDVP:RAP}) \vee \mu_6(\text{MDVP:RAP})]$
$\wedge [\mu_4(\text{MDVP:PPQ}) \vee \mu_5(\text{MDVP:PPQ})]$
$\wedge [\mu_3(\text{Jitter:DDP}) \vee \mu_4(\text{Jitter:DDP}) \vee \mu_5(\text{Jitter:DDP}) \vee \mu_6(\text{Jitter:DDP})]$
$\wedge [\mu_2(\text{MDVP:Shimmer}) \vee \mu_5(\text{MDVP:Shimmer}) \vee \mu_6(\text{MDVP:Shimmer}) \vee \mu_7(\text{MDVP:Shimmer})]$
$\wedge [\mu_2(\text{MDVP:Shimmer(dB)}) \vee \mu_5(\text{MDVP:Shimmer(dB)}) \vee \mu_7(\text{MDVP:Shimmer(dB)})]$

∧ [μ_1 (Shimmer:APQ3) ∨ μ_2 (Shimmer:APQ3) ∨ μ_3 (Shimmer:APQ3) ∨ μ_4(Shimmer:APQ3)]
∧ [μ_2(Shimmer:APQ5) ∨ μ_3(Shimmer:APQ5)]
∧ [μ_1(MDVP:APQ) ∨ μ_6(MDVP:APQ) ∨ μ_7(MDVP:APQ)]
∧ [μ_2 (Shimmer:DDA) ∨ μ_3 (Shimmer:DDA) ∨ μ_4 (Shimmer:DDA) ∨ μ_5(Shimmer:DDA)]
∧ [μ_3(NHR) ∨ μ_4(NHR) ∨ μ_6(NHR) ∨ μ_7(NHR)]
∧ [μ_3(HNR) ∨ μ_6(HNR)]
∧ [μ_1(RPDE) ∨ μ_4(RPDE)]
∧ [μ_2(DFA) ∨ μ_3(DFA) ∨ μ_4(DFA) ∨ μ_5(DFA) ∨ μ_7(DFA)]
∧ [μ_1(Spread1) ∨ μ_3(Spread1) ∨ μ_5(Spread1) ∨ μ_6(Spread1)]
∧ [μ_1(Spread2) ∨ μ_2(Spread2) ∨ μ_6(Spread2)]
∧ [μ_2(D2) ∨ μ_3(D2) ∨ μ_4(D2)]
∧ [μ_3(PPE) ∨ μ_4(PPE) ∨ μ_5(PPE) ∨ μ_6(PPE) ∨ μ_7(PPE)]

Constant for the output variable of Rule 4 is $f4 = +1$

Supplementary Table S4

Interpretation of Rule 1 and the respective fuzzy equation extracted from GAFL (oral cancer dataset)

IF		**Age is within 1-3**
	and	Ethnicity is 2 or 3
	and	Site is within 2 to 4
	and	PT is 1 or 3
	and	PN is 1 or 3
	and	Stage is 2 or 3
	and	Size is 3
	and	Treatment is 1
THEN		It is of survival.

If A is within scale of 1-3, B is of scale 2 or 3, C is within scale of 2-4, D is of scale 2 or 4, E is of scale 2 or 4, F is of scale 2 or 3, G is of scale 3, and H is of scale 1, then the predicted survival rate is survive.

$W_1 = [\mu_{1\text{-}3}(A) \wedge \mu_{2\text{-}3}(B) \wedge \mu_{2\text{-}4}(C) \wedge \{\mu_2(D) \vee \mu_4(D)\} \wedge \{\mu_2(E) \vee \mu_4(E)\} \wedge \mu_{2\text{-}3}(F) \wedge \mu_3(G) \wedge \mu_1(H)]$
$F_1 = 0$