Improvements in Sugeno-Yasukawa Modelling Algorithm

Amir H. Hadad, B. S. U. Mendis, Tom D. Gedeon

***Abstract*— A modified version of Sugeno-Yasukawa (SY) modelling algorithm is presented. We have employed a new method for parameter identification phase based on genetic algorithms (GA). Moreover, we have modified the modelling sequence by applying parameter identification on intermediate models. Models created with this method had lower mean square errors (MSE) compared to original algorithm. A case study on breast cancer survival prediction is also presented that demonstrates a thorough comparison of the new modelling algorithm with several other methods such as SVM, C5 decision tree, ANFIS and the original SY method. The modified SY method had the highest average of accuracies among all models. Moreover, it had significantly higher accuracy compared to the original SY method and ANFIS. 10-fold cross validation approach was employed for all evaluations.**

I. INTRODUCTION

Fuzzy logic was introduced by Zadeh in 1965 as a way of expressing uncertainty [1]. Later in 1976, Mamdani and Assilian [2], [3] presented a fuzzy logic controller based on Zadeh's calculus [4]. In this study, control rules that were stated by an expert human operator were converted into an automatic controller. The controlling rules employed in the system were called Mamdani fuzzy model in future studies. Takagi and Sugeno introduced a fuzzy model in 1985 (TS fuzzy model)[5]. Both TS and Mamdani fuzzy models are described by fuzzy IF-THEN rules. However, unlike Mamdani fuzzy model, where both antecedents and consequences are fuzzy propositions, in TS fuzzy model only the antecedents are fuzzy propositions and the consequences are linear functions. In [6], [7], [8] the literature it has shown that Mamdani and TS fuzzy models are both universal approximators.

Most of these early studies were focused on defining fuzzy models, however, they did not introduce a process to identify fuzzy models automatically and the identification process was manual. For complex systems manual fuzzy model identification is not possible. Therefore, in later studies automatic model identification was considered [7], [9], [10], [11], [12]. One of these early studies was Sugeno-Yasukawa fuzzy modelling algorithm that was a clustering based method to identify Mamdani fuzzy model [12].

In this study, Sugeno-Yasukawa modelling is described in Section II. Then, AH fuzzy modelling (the modified modelling algorithm is refereed to as AH fuzzy modelling algorithm in the text) algorithm and the evolutionary parameter identification algorithm is presented in Section III. In the experimental result section, the AH algorithm is employed for modelling of a non-linear system in subsection IV-A and then in a case study for breast cancer survival prediction. The purpose of this case study was to classify patients into two groups of surviving patients and non-surviving patients. The survival threshold was considered 5 years similar to other studies on breast cancer survival prediction [13], [14], [15]. The original SY fuzzy modelling algorithm along with common machine learning techniques such as SVM [16], C5 decision tree [17] and ANFIS [11] were employed in this case study as a benchmark for AH algorithm. Predictive classification accuracy (accuracy and predictive classification accuracy are referring to the same measure in this text) was employed for comparing the performance of the classifiers.

II. SUGENO YASUKAWA MODELLING

In their study, Sugeno and Yasukawa [12] defined two high level steps: structure identification and parameter identification. Two types of structure identification was described as type I that is dealing with finding input variables and type II that is dealing with finding input-output relations. After the structure of a model is identified, parameter identification, that can be interpreted as fine tuning of the model parameters, takes place. For Mamdani fuzzy model, model parameters are fuzzy membership functions and parameter identification is optimizing the initially defined membership functions.

A. Fuzzy Model

The Mamdani fuzzy model used in SY modelling is developed for a multi−input and single−output system [12]:

$$R^i : if\, x_1\, is\, A_1^i\, and\, x_2\, is\, A_2^i\, ...and\, x_n\, is\, A_n^i\, then\, y\, is\, B^i, \quad (1)$$

where R^i is the *i*th rule ($1 \leq i \leq m$), x_j ($1 \leq j \leq n$) are input variables, y is the output, and A_j^i and B^i are fuzzy variables.

B. Structure Identification

As mentioned structure identification were divided into two types. Firstly structure identification type II will be discussed. In order to find input-output relations, number of rules and partitioning of input space are the two main phases. An incremental algorithm based on fuzzy c-means (FCM) clustering [18] is applied on output values to identify the optimal number of rules. In this algorithm, number of output clusters created based on FCM increases from two to more until SC criterion (see Equation 2) reaches its minimum

Amir H. Hadad is a PhD candidate at School of Computer Science, Australian National University (phone: +61 261 259665; email: amir.hadad@anu.edu.au).

B.S.U. Mendis is a research fellow at School of Computer Science, Australian National University (phone: +61 261 259664; email: sumudu@cs.anu.edu.au).

Tom Gedeon is a professor at School of Computer Science, Australian National University (phone: +61 261 2514216; email: tom.gedeon@anu.edu.au).

978-1-4244-8126-2/10/$26.00 ©2010 IEEE

value.

$$SC_m = \sum_{i=1}^{N} \sum_{j=1}^{nc} u_{ij}^m \left(\|x_i - v_j\|^2 - \|v_j - \overline{x}\|^2 \right), \quad (2)$$

where,
N: number of data to be clustered;
nc: number of clusters ($nc \geq 2$);
x_i: i'th data;
v_j: center of j'th cluster;
$\overline{x}$: average of data $x_1,\ x_2,\ ..., x_n$;
$\|\|$: norm;
u_{ij}: grade of i'th data belonging to j'th cluster;
m: adjustable weight (usually between 1.5 and 3)

Afterwards, process of selecting effective inputs takes place. Firstly, the training data is divided into two groups, A and B. Then, by using these two groups of the training data, two intermediate fuzzy models are created for each of the input variables that have one input and one output (For example, if there are five input parameters in the training data, 5 pairs of intermediate models will be created). The best input is chosen based on the RC values of these models (see Equation 3). Smaller RC values implies that the input(s) has higher effect on the output. As the next stage, new pairs of intimidate models with two inputs are created. The first input of the later intermediate models is the same as the best intermediate model from the first stage. This process continues until all RC values of stage n are more than stage n best RC or all inputs are used. Meanwhile, if in stage n, RC value of the pair of intermediate model is less than the global minimum RC value, the n'th input variable is considered ineffective. The reason is to avoid creating too many intermediate models.

$$RC = \left[\sum_{i=1}^{k_A} \frac{(y_i^A - y_i^{AB})^2}{k_A} + \sum_{i=1}^{k_B} \frac{(y_i^B - y_i^{BA})^2}{k_B} \right], \quad (3)$$

where,
k_A and k_B: the number of data of the groups A and B;
y_A and y_B: the output data of the groups A and B;
y_{AB}: the output data of intermediate model identified for group A for input data of group B;
y_{BA}: the output data of intermediate model identified for group B for input data of group A;

The membership values calculated by FCM algorithm [18] is employed to create trapezoid type membership function for inputs and outputs by employing a trapezoidal approximation algorithm (Studies presented in [12], [19], [20] are examples of trapezoidal approximation algorithms). By the end of this stage a Mamdani fuzzy rule base presented in 1 is created.

C. Parameter Identification

The Mamdani model that is created by the structure identification process, can be further optimized. In this study trapezoidal shaped fuzzy membership functions (tmf) are used to represent fuzzy variables:

$$A : \{p^1, p^2, p^3, p^4\}, \quad (4)$$

where A is a fuzzy variable and $\{p^1, p^2, p^3, p^4\}$ are trapezoidal membership function parameters.

Parameter identification is responsible for this process. Sugeno and Yasukawa demonstrated a parameter identification algorithm in their study [12] that is as follows:

1) Set the value f of adjustment.
2) Assume that the kth parameter of the jth fuzzy set is p_j^k.
3) Calculate $p_j^k + f$ and $p_j^k - f$. If $k = 2, 3, 4$ and $p_j^k - f$ is less than p_j^{k-1}, then $\hat{p}_j^k = p_j^{k-1}$; else $\hat{p}_j^k = p_j^k - f$. Also if $k = 1, 2, 3$ and $p_j^k + f$ is bigger than p_j^{k+1}, then $\hat{\hat{p}}_j^k = p_j^{k+1}$; else $\hat{\hat{p}}_j^k = p_j^k + f$.
4) Choose the parameter which shows the best performance among $\{\hat{p}_j^k, p_j^k, \hat{\hat{p}}_j^k\}$ and replace p_j^k with it.
5) Go to step 2 while unadjusted parameters exist.
6) Repeat step 2 until we are satisfied with the result.

The value of f was set into 5% of the width of universe of discourse and steps 1 to 6 were ran six times.

III. AH FUZZY MODELLING

The high level flowchart of SY modelling algorithm and the its modified version AH modelling algorithm are presented in Figures 1 and 2. We have employed a GA parameter identification instead of the original sequential algorithm employed by Sugeno and Yasukawa. Furthermore, in AH, the parameter identification process is applied for intermediate fuzzy models as well as the final fuzzy model.

The introduced parameter identification algorithm in [12] is a sequential algorithm. Therefore, it is possible to calculate MSE values for one of the readjusted membership parameters at a time. On the other hand, in this algorithm there are 2 new values for each of the trapezoidal membership function parameters. As a result, the search space for tuning the fuzzy model is very restricted and there is no possibility of parallel change of parameters for tuning. This problem increases the chance of choosing a local minimum instead of the global minimum. By employing GA for parameter identification, parallel readjustment of membership functions becomes possible. On the other hand, for each of the membership function parameters, more values will be tested for fine tuning.

Our earlier study showed that by applying parameter identification for intermediate fuzzy models, the RC values decreases [21]. This decrease in RC values implies that the intermediate models have a lower error rate. As a result of having more accurate intermediate fuzzy models, the process of selecting effective input parameters (structure identification type I or feature selection) is more reliable. In this study by facilitating a way of using GA for parameter identification of intermediate models we have achieved lower RC values compared to both studies [21] and [12] for the non-linear system presented in Equation 8.

A. Evolutionary Parameter Identification

Evolutionary algorithms have been employed for different stages of fuzzy modelling (GA based parameter identification and evolutionary parameter identification are referring to the

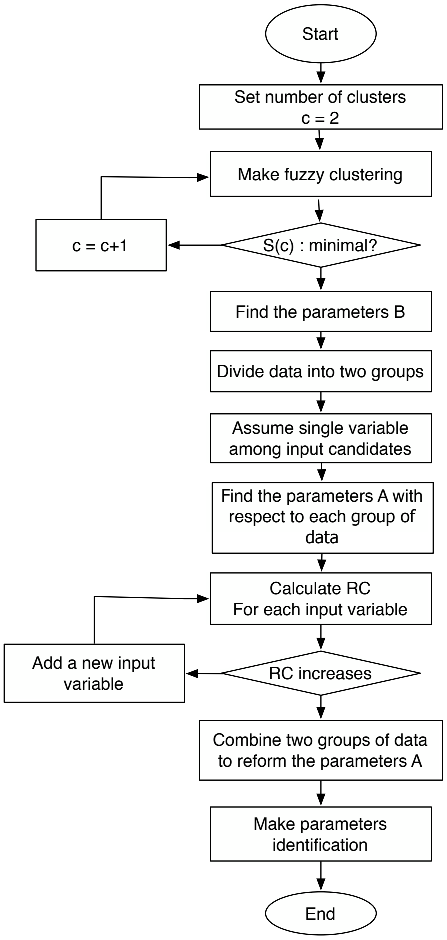

Fig. 1. Algorithm flowchart for SY modelling

same algorithm). Cordon and his colleagues in [22] present two general categories of employing genetic algorithms in fuzzy modelling: genetic tuning processes and learning with genetic algorithms. In this study genetic algorithm is employed in the first category and more specifically for tuning of membership functions of the fuzzy model (parameter identification). Aim of a genetic tuning process is to adopt an existing fuzzy model to create a fuzzy model with a better performance.

In [22] definition of chromosome C_j was presented for a rule base with triangular shaped membership function. We have presented a similar definition for a Mamdani rule base with trapezoidal membership functions. First, we need to define the partial chromosome C_{ji}, $i = 1, ..., m$, as follows:

$$C_{ji} = \left(p^1_{i1}, p^2_{i1}, p^3_{i1}, p^4_{i1}, ..., p^1_{in}, p^2_{in}, p^3_{in}, p^4_{in}, p^1_i, p^2_i, p^3_i, p^4_i\right), \tag{5}$$

where,
$(p^1_{ik}, p^2_{ik}, p^3_{ik}, p^4_{ik})$: are the four parameters of the k'th input fuzzy set in the i'th rule antecedent;
$(p^1_i, p^2_i, p^3_i, p^4_i)$: are parameters of the output fuzzy set in the i'th rule consequent;
j: index of the chromosome in the whole population;
i: index of the rule in the fuzzy rule base $(1 \leq i \leq m)$;
m: number of rules;
n: number of inputs.

Therefore, the j'th chromosome C_j definition is concatenation of rule chromosomes C_{ij}:

$$C_j = C_{j1}\ C_{j2}\ ...\ C_{jm} \tag{6}$$

The first chromosome that is created based on the fuzzy model is referred to as the initial chromosome. Given that size of the chromosomes affects the search space, in our simulations the population of the chromosomes is set based on the length of the initial chromosome:

$$L = Ratio \times ||C_j||, \tag{7}$$

where,
L: Initial chromosomes population,
$Ratio$: a value that defines what percentage of the chromosome length is the initial population $(0 < Ratio \leq 1)$,
$||C_j||$: length of a chromosome.

For example for a Mamdani rule base with 3 rules that each rule has two inputs, the size of a chromosome equals 36 (3 rules × (2 inputs × 4 trapezoidal membership function parameters (tmf params) + 1 output * 4 tmf params) = 36). If the $Ratio$ equals 50% (0.50), therefore the size of the initial population equals 18 chromosomes (from Equation 7, 36 multiplied by 0.50 equals 18). Note that the initial population is created based on the initial chromosome (see Equation 6) that was created from the fuzzy model (similar to GA tuning process in [22], [23]).

IV. EXPERIMENTAL RESULTS

In this section we present the modified algorithm results for a non-linear system firstly. Then a case study on employing the modified algorithm (AH [1]) along with other machine learning algorithms for breast cancer survival prediction is presented. In both simulations for the nonlinear system and the case study for breast cancer survival predication, Ratio's value was 60%. The MSE function was employed as the GA's fitness function, number of iterations was set to 40 iteration. For other GA parameters such as mutation, crossover functions, etc. the default values of the MATLAB GA toolbox version 2.3 were used.

A. Non-Linear System

In [12] a sample non-linear system was presented (see Equation 8). We have used the evolutionary parameter identification for this system and the results are compared (see Figure IV-A). The best MSE achieved Sugeno and Yasukawa study was 0.01 which is presented by dotted line in Figure IV-A. Results show that the MSE values of all 10 runs of 10-fold cross-validation are lower than the best SY MSE value that is a significant improvement.

$$y = (1 + x_1^{-2} + x_2^{-1.5})^2, 1 \leq x_1, x_2 \leq 5. \tag{8}$$

---

[1] AH stands for Amir Hadad's modelling algorithm and SY stands for Sugeno-Yasukawa's modelling algorithm

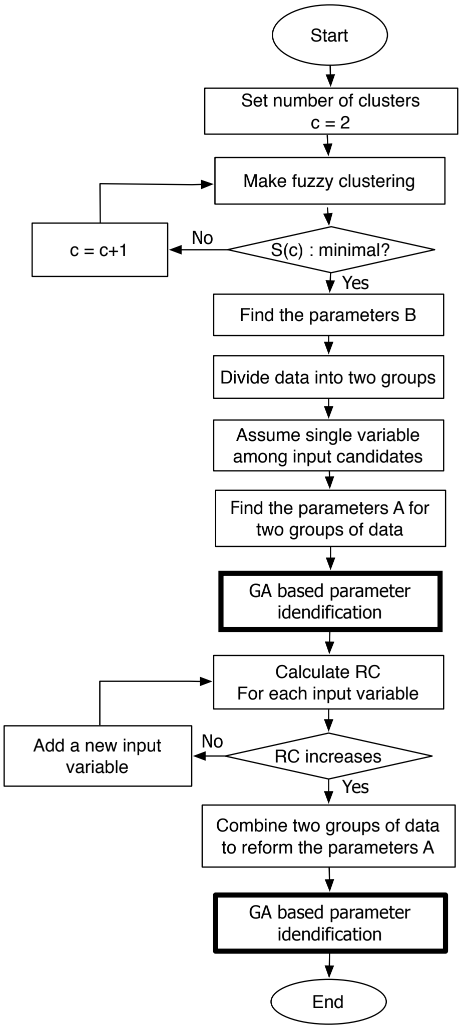

Fig. 2. Algorithm flowchart for AH modelling

TABLE I

COMPARISON OF MINIMUM RC VALUES FOR INTERMEDIATE MODELS

	One input	Two inputs
AH	*0.1450*	*0.0680*
Case (I)	0.6300	0.4240
Case (II)	0.4705	0.2905

The minimum RC values of the intermediate models created by AH method in this experiment is compared to two other cases in Table I. In case (I) no parameter identification was applied for intermediate models and in case (II) SY parameter identification is applied.

B. A Case Study: Breast Cancer Survival Prediction

The purpose of this case study was to investigate the applicability of the improved SY method for a real world problem. The problem investigated was survival prediction for breast cancer patients. Moreover, a group of different machine learning techniques were employed as a benchmark for the AH modeler (modified SY modeler) results.

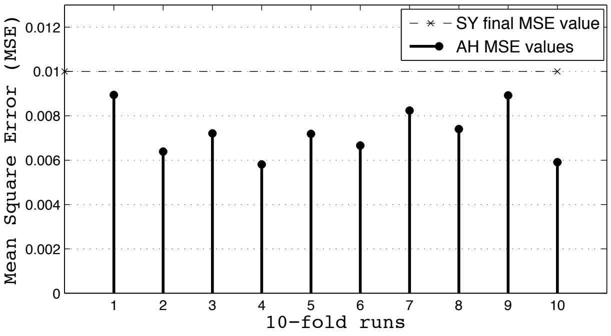

Fig. 3. Comparison of AH and SY performance for a sample nonlinear system (see Equation 8)

A classification problem was defined on the Australian National Territory and surrounding regions Breast Cancer Treatment Dataset (ABCTD) (1997-2009). There were 12 prognosis factors for 814 patients in the dataset. Additionally, for each patient, the number of survival days was recorded in the dataset. In the classification problem, a five year survival threshold were used. The aim was to predict surviving and non-surviving patients. The five year threshold used for patients' survival in the first classification problem, is the same as studies parented in [13], [14], [15].

1) The ACT and surrounding regions Breast Cancer Treatment Dataset: ABCTD dataset consists of prognosis records of 814 patients. The eligibility criteria for entry into the registry was primarily found on early breast cancer stage, although it included patients with single primary and multiple primary tumours. It has been collected over twelve years of follow-up from 1997 to 2009 as a part of the Breast Cancer Treatment Project[2] [24]. The breast Cancer Treatment Project was started in May 1997. As a result, an ongoing community based audit of breast cancer treatment was developed. The clinical indicators used in this audit were agreed by the ACT and SE NSW Breast Cancer Treatment Group. The aim of this project is to collect and examine data on treatment and outcomes for women with breast cancer. There are over 80 prognosis factors in the dataset as well as invaluable survival information. For this case study, based on the experience of breast cancer expertise a subset of thirteen prognosis factors were selected for survival perdition purpose (Table II).

Unlike SEER [25] that has unifocal tumour size information (only the largest tumour size is provided), ABCTD provides data about whether the patient has multiple tumours or a single tumour (this factor is referred to as multi_single prognosis factor in the text).

2) Experiment Design: C5 decision tree [17], SVM with RBF kernel [16], ANFIS [11], SY and AH were employed to classify patients. In the classification problem, 10-fold cross validation of the ABCTD dataset was performed for training and testing of the classifiers. Predictive classification accuracy [26], [27] was measured for all classifiers (see

---

[2] http://health.act.gov.au/c/healtha=da&did=10048779&pid=1060038141

TABLE II

PATIENTS' PROGNOSIS FACTORS IN CANBERRA BREAST CANCER DATASET(1997-2007)

No	Prognosis Factor	Description
1	Age	Age of the patient
2	Nodes removed	Number of Lymphatic nodes removed from the patients breasts
3	Nodes involved	Number of cancerous lymphatic nodes
4	Progression Status	Progression Status
5	Multi_single	Whether the patient has got multiple or single tumours
6	Histology	Histology of the patient
7	LargTum	Size of the largest patient's tumour
8	Nodes	Wether lymphatic nodes are cancerous or not
9	LVI	Lymphovascular Invasion
10	Grade	Tumour grade (TNM staging)
11	ER	Receptor status: oestrogen
12	PR	Receptor status: progesterone
	Survival	Which survival group patient belongs to: Less than 5 years or more than 5 years

TABLE III

CLASSIFICATION CONFUSION MATRIX (REDRAWN FROM [27])

True Class	Predicted $^{-}$ve	Class $^{+}$ve	
$^{-}$ve	T$_n$	F$_p$	C$_n$
$^{+}$ve	F$_n$	T$_p$	C$_p$
	R$_n$	R$_p$	N

TABLE IV

ACCURACY VALUES FOR 5 YEARS SURVIVAL IN ABCTD

Method	Average of Accuracies %	Variance of Accuracies %	Highest Accuracy %	Lowest Accuracy %
ANFIS	60.20	3.45	66.67	48.72
C5	61.86	1.26	64.81	60.13
SVM	63.27	1.45	66.08	60.90
SY	53.75	4.00	56.15	46.15
AH	*63.57*	2.30	69.62	59.49

TABLE V

PAIRWISE STUDENT T-TEST RESULTS

Method	p-value
AH vs SY	*0.00028*
AH vs ANFIS	*0.03387*
AH vs SVM	0.40236
AH vs C5	0.08743

Equation 9). Furthermore, the pairwise student t-test was employed to investigate whether the AH algorithm had improved the accuracy for the ABCTD survival prediction significantly.

Two types of errors can be defined for binary classification problems: type I error is false positives (F_p) and type II error is false negatives (F_n). False positives occur when a record that is actually negative is classified as positive and false negatives occur when a record is acutely positive and is classified as negative. True positives (T_p) and true negatives (T_n) is when the records are classified correctly as either positive or negative. C_n and C_p are total number true negatives and positive records, and R_n and R_p are predicted positive and negative records (see Table III). As mentioned before, for measuring the performance of different classifiers, predictive accuracy was calculated. It is a measure that takes into consideration all correctly detected cases and also all cases that were classified correctly as not belonging to the class (T_p and T_n respectively). It is a very common performance measure and most of the recent studies used it for analysis of classifiers performance [28], [29], [15].

$$Accuracy = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} = \frac{T_p + T_n}{C_p + C_n}, \qquad (9)$$

3) Case study results: LIBSVM toolbox [30] was used to train both a C5 decision tree (modified version of C4.5 [17], [31]) and an RBF Kernel SVM classifier. For the C5 decision three, the training was performed with different parameter settings. For each of the training runs, confidence parameter were changed from 0.1 to 0.5 with step size of 0.1 and the best confidence value was selected. This process was performed ten times for each fold. Therefore, there was 50 (5 different confidence value * 10 times) runs of training for each fold and 500 runs for all ten folds. Grid search for C and γ parameters was performed to create the most optimized SVM classifier (C $\in$ [1,10] with steps of size 1.0 and γ $\in$ [-5,5] with steps of size 1.0). ANFIS and SY classifiers were the fuzzy based methods used to create classifiers for the binary classification problem. For ANFIS the default parameter values in MATLAB ANFIS modeler [32] were used. The f value for SY algorithm was 5% of universe of discourse and SC criterion was employed to measure the optimal number of clusters. The values of predictive accuracy for all classifiers are presented in Table IV. AH method had the highest accuracy compared to all methods (63.57%). The student pairwise t-test were employed for AH algorithm and all other algorithm (see Table V). Results shows that AH algorithm accuracy is significantly higher than SY and ANFIS algorithms. But AH algorithm has not achieved a significant improvement compared to SVM and C5 algorithms. By comparing the variance values in Table IV, AH average of accuracies shows a higher reliability compared to ANFIS and SY algorithms average of accuracies. The most reliable accuracies' average belongs to C5 and the SVM is the second most reliable. In 10 runs, AH highest accuracy is the highest accuracy among all classifiers in this table (69.62%).

V. CONCLUSION

A modified version of SY modeler with an evolutionary parameter identification algorithm was presented. Two experiments were performed: a non-linear system modelling and breast cancer survival prediction. In both experiment, we demonstrated that by employing the evolutionary algorithm instead of the original parameter identification presented by Sugeno and Yasukawa, the tuned fuzzy model was performing significantly better compared to SY fuzzy model. The final fuzzy model created by AH algorithm in the case study had the highest accuracy value. Although the accuracy was not significantly higher than SVM and C5 accuracy values, we need to consider that the created classifier based on Mamdani fuzzy model has the advantage of linguistic interpretability [33], [23].

Additionally, we modified SY modelling process, by applying the GA based parameter identification for intermediate fuzzy models in the modelling process. As shown in nonlinear modelling experiment, this change resulted in lower RC values compared to the original SY algorithm [12] and compared to a case where SY parameter identification algorithm was applied [21]. Therefore, AH modelling has a more reliable effective inputs selection (feature selection) compared to SY modelling.

In the case study that was performed for breast cancer survival prediction, AH algorithm average of accuracies was moderately reliable compared to the other algorithms. One might consider improving the evolutionary parameter identification (or the GA tuning process) in order to decrease the variance the fuzzy modelling algorithm as future work.

REFERENCES

[1] L. A. Zadeh. Fuzzy sets [j]. *Information and Control*, 8(3):338–353, 1965.

[2] E. H. Mamdani. Application of fuzzy logic to approximate reasoning using linguistic synthesis. In *Proceedings of the sixth international symposium on Multiple-valued logic*, pages 196–202. IEEE Computer Society Press Los Alamitos, CA, USA, 1976.

[3] E. H. Mamdani and S. Assilian. An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies*, 7(1):1–13, 1975.

[4] L.A. Zadeh. Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Transactions on Systems, Man, and Cybernetics*, 3(1):28–44, 1973.

[5] T. Takagi and M. Sugeno. Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man, and Cybernetics*, 15(1):116–132, 1985.

[6] H. Ying. Sufficient conditions on uniform approximation of multivariate functions by general Takagi-Sugeno fuzzy systems with linear rule consequent. *IEEE Transactions on Systems, Man and Cybernetics, Part A*, 28(4):515–520, 1998.

[7] H. Ying. Sufficient conditions on general fuzzy systems as function approximators. *Automatica (Journal of IFAC)*, 30(3):521–525, 1994.

[8] K. Tanaka and H.O. Wang. *Fuzzy control systems design and analysis: a linear matrix inequality approach.* Wiley-Interscience, 2001.

[9] S.L. Chiu. Fuzzy model identification based on cluster estimation. *Journal of intelligent and Fuzzy systems*, 2(3):267–278, 1994.

[10] C.T. Sun. Rule-base structure identification in an adaptive-network-based fuzzy inference system. 1994.

[11] JS Jang. ANFIS: Adaptive-Network-Based Fuzzy Inference System. *IEEE Transactions on Systems, Man and Cybernetics*, 23(3):665–685, 1993.

[12] M. Sugeno and T. Yasukawa. A fuzzy-logic-based approach to qualitative modeling. *In 5th Proceeding of Fuzzy System Symposium*, 1:7–31, 1993.

[13] Harry B. Burke, Philip H. Goodman, and David B. Rosen. Neural network for measuring cancer outcomes. *Instrumentation and Measurement Technology Conference, IEEE/94*, 1994.

[14] H. Seker, M. Odetayo, D. Petrovic, RNG Naguib, C. Bartoli, L. Alasio, MS Lakshmi, and GV Sherbet. A fuzzy measurement-based assessment of breast cancer prognosticmarkers. In *2000 IEEE EMBS International Conference on Information Technology Applications in Biomedicine, 2000. Proceedings*, pages 174–178, 2000.

[15] D Delen, G Walker, and A Kadam. Predicting breast cancer survivability: a comparison of three data mining methods. *Artificial Intelligence In Medicine*, 34(2):113–127, 2005.

[16] V. Vapnik, S.E. Golowich, and A. Smola. Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing. *Advances in Neural Information Processing Systems 9*, 1997.

[17] J.R. Quinlan. C4. 5: Programs for machine learning. *Morgan Kaufmann Publishers Inc. San Francisco, CA, USA*, 1993.

[18] J. C. Bezdek and R. Ehrich. FCM: the Fuzzy C-Means Clustering. *COMP. GEOSCI.*, 10(2):191–203, 1984.

[19] D Tikk, G Biro, T Gedeon, and L Koczy. Improvements and critique on Sugeno's and Yasukawa's qualitative modeling. *Fuzzy Systems*, pages 596–606, Jan 2002.

[20] A. H. Hadad, T. D. Gedeon, and B. S. U. Mendis. Finding Input Sub-Spaces for Polymorphic Fuzzy Signatures. *Proceedings of the 18Th International Conference on Fuzzy Systems*, pages 1089–1094, 2009.

[21] A Hadad, S Ghidary, and S Shouraki. An Improvement in Sugeno-Yasukawa Modeler. *Proceedings of the International Conference on Computational Intelligence*, Jan 2006.

[22] O. Cordon, F. Herrera, F. Hoffmann, and L. Magdalena. *Genetic fuzzy systems*. World Scientific, 2001.

[23] J.A. Roubos, M. Setnes, and J. Abonyi. Learning fuzzy classification rules from labeled data. *Information Sciences*, 150(1-2):77–93, 2003.

[24] Breast Cancer Treatment Project (last accessed on 27th December 2009), 1997-2009.

[25] Surveillance, Epidemiology and End Results (SEER). Program (www.seer.cancer.gov) SEER*Stat Database: Mortality – All COD, Public-Use With State, Total U.S. (1969–2004), National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, released April 2007. Underlying mortality data provided by NCHS, 2007.

[26] KJ Cios and G William Moore. Uniqueness of medical data mining. *Artificial Intelligence In Medicine*, 26(1-2):1–24, 2002.

[27] Andrew P. Bradly. Roc curves and the x^2 test. *Pattern Recognition Letters*, 17(3):287–294, 1996.

[28] D.J. Sargent. Comparison of artificial neural networks with other statistical approaches. *Cancer*, 91, 2001.

[29] J.A. Cruz and D.S. Wishart. Applications of machine learning in cancer prediction and prognosis. *Cancer Informatics*, 2:59–78, 2006.

[30] C.C. Chang and C.J. Lin. Libsvm: a library for support vector machines, 2001.

[31] I.H. Witten and E. Frank. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann Publishers Inc. San Francisco, CA, USA, 2005.

[32] The MathWorks, Inc. MATLAB®. 1998-2009.

[33] J. Abonyi, J.A. Roubos, and F. Szeifert. Data-driven generation of compact, accurate, and linguistically sound fuzzy classifiers based on a decision-tree initialization. *International journal of approximate reasoning*, 32(1):1–21, 2003.