Journal of Theoretical and Applied Information Technology
30th April 2022. Vol.100. No 8

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

EFFECTIVENESS OF ENHANCED TAKAGI SUGENO KANG'S FUZZY INFERENCE MODEL

SUGIYARTO SURONO[1], GOH KHANG WEN[2*], CHOO WOU ONN[2], YOSZA BIN DASRIL[3], ASIH YULI ASTUTI[1], NITHYA PERIASAMY[4]

[1]Department of Mathematics, Ahmad Dahlan University, Yogyakarta, Indonesia

[2]Faculty of Data Science and Information Technology, INTI International University, Nilai, Malaysia

[3]Faculty of Technology Management and Business, Universiti Tun Husseion Onn Malaysia, Batu Pahat, Johor, Malaysia

[4]Department of Mathematics, Science and Computer, Ungku Omar Polytechnic, Perak, Malaysia

E-mail: [1]sugiyarto@math.uad.ac.id , [2]khangwen.goh@newinti.edu.my, [2]wouonn.choo@newinti.edu.my, [3]yosza@uthm.edu.my, [1]asih1800015019@webmail.uad.ac.id, [4]nithya@puo.edu.my

*Correspondence: khangwen.goh@newinti.edu.my

ABSTRACT

Fuzzy logic is one of the components from soft computing and it is useful as a method to map the problem from input into the desired output. The main feature of fuzzy logic is membership function. Takagi Sugeno Kang fuzzy system appears to be more preferred and easier to use, thanks to its simple structure and high proximity. The use of learning rate is extremely important in Mini-Batch Gradient Descent to improve the quality and the rate of training convergence. However, choosing the right learning rate is one of the shortcomings of the MBGD method, so AdaBound is used to optimize the selection of learning rates. AdaBound is an adaptive method that applies dynamic boundaries on learning rate. The said dynamic boundaries has the definition of learning rate limitation, both from up and down, to make sure the learning rate is neither too big nor too small. Aside from that, the boundaries become tighter as iteration goes and forcing the learning rate close to constant value. The optimization of AdaBound is intended not only to improve the convergence rate, but also to optimize the convergence into global minimum value in the end of training period. The method used in this study is Takagi Sugeno Kang (TSK) method. The rule of Takagi Sugeno Kang fuzzy system will be optimized using Mini-Batch Gradient Descent that has been modified by AdaBound. The goal of this study is to observe the accuracy of classification done by TSK with MBGD-A. The data used in this study is obesity data obtained from Kaggel dataset, while the variables are composed of three independent variables and one dependent variable. There are 32 rules in the TSK inference model based on data, however only 24 of them can be applied. To obtain firm numbers, the defuzzification value is derived using the weighted average approach from the 24 rules obtained. MAD was utilized as the evaluation model, and the error value was 14,549, indicating that the approach was effective and efficient.

Keywords: *Fuzzy TSK, Optimization, MBGD, AdaBound*

1. INTRODUCTION

Machine learning algorithms are divided into 3 types, namely supervised learning, unsupervised learning, and reinforcement learning. An example of a supervised learning algorithm is classification. Classification is a process to find a model or a function that describes and differentiates data into several classes. It includes a process in which the characteristic of an object is evaluated and is grouped into previously defined class [1]. While an example of an unsupervised learning algorithm is clustering. Clustering is a method of grouping data by dividing the data set into several groups according to similarities. Look for patterns in a collection of data by grouping data into several groups that are the goal of clustering [2], [3].

Fuzzy rule-based classification is usually encountered in machine learning construction because of its ability to build models based on linguistic variables. There are three important components in fuzzy rule-based classification, including database, rule base, and fuzzy logic [4].

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Fuzzy logic can be used to map the problem from the input to the expected output. Fuzzy logic can be considered as a black box where the black box contains a method or method that can be used to process input data into output in the form of good information [5], [6]. Several reasons on why fuzzy logic is widely used [7] such as the concept of fuzzy logic is easier to understand, fuzzy logic is very flexible which makes it adaptable in changes and uncertainty, fuzzy logic has a tolerance over incorrect data, fuzzy logic is capable in modelling complex nonlinear function. Additionally, fuzzy model has been approved as important architect and effective in tackling uncertainty in system modelling.

Fuzzy inference system is a framework based on the concept of fuzzy sets, if-then rules, and fuzzy logic. One of the fuzzy inference systems that is often used is the Takagi Sugeno Kang (TSK) fuzzy inference system. In this method, the reasoning is almost the same as the Mamdani fuzzy inference system, the difference is the output produced. If the Mamdani method produces a fuzzy set, then the TSK method produces a constant or linear function [8].

TSK fuzzy system, which was proposed by Takagi, Sugeno, and Kang, has become more preferred and better in practicality because of its simple structure and high level approach ability [8]. Two things are needed in the construction of TSK fuzzy system that is structure identification and parameter estimation [9]. Structure identification is related in finding the correct partition from input room, while parameter estimation is related in finding optimal value from all of parameter rule in TSK fuzzy system [10], [11]. Some of the advantages of the TSK method include that the TSK method is more efficient in solving computational problems, works best for linear and adaptive optimization techniques, and can guarantee the continuity of the output surface [12].

Takagi Sugeno Kang (TSK) is widely applied in various fields. Research conducted by [13] is applied to the health sector to detect EEG Epilepsy with fuzzy mukti-view TSK. Based on the experimental results, it was found that the proposed method has a better classification performance compared to other EEG detections. Takagi Sugeno Kang was used by Raj and Mohan [14] for modeling and analysis of PID controllers with modified rules. Du, et al [15] also used Takagi Sugeno Kang but an efficient graph-based predictor of dialysis adequacy of hemodialysis patients. In addition Meng and Zhang [16] using TSK and deep features to automatically identify the anxiety of college students and experimental results show the TSK fuzzy system has good classification and generalization performance.

Gradient descent optimization is often used to find the minimum value of cost function [17]–[20]. Gradient Descent is the first introduced method and applies constant point method in forming the first derivative from cost function into zero [21]. There are three methods of gradient descent including Batch Gradient Descent (BGD), Stochastic Gradient Descent (SGD), and Mini-Batch Gradient Descent (MBGD). Nakasima et al. [22] compared the performance of BGD, SGD, and MBGD on a Mamdani based neuro fuzzy system using center-of-sets defuzzification and obtained the result that the MBGD learning method achieved the best performance compared to BGD and SGD.

Mini-Batch Gradient Descent (MBGD) is the improved version which is a combination between Batch Gradient Descent and Stochastic Gradient Descent shows the best result in improving performance and minimalizing error in cost function on machine learning algorithm [22]–[24]. Among them, the mini-batch gradient descent MBGD method is also commonly used in the model training of large-scale data in machine learning [25]. Several studies using MBGD include Gou and Yu [26] using MBGD to train an Artificial Neural Network (ANN) equalizer efficiently. The Mini Batch Gradient Descent was also used by Jing Li et al [27] to overcome the high computational costs of large-scale real hyperspectral images. Kodali and Ramani [28] also use Mini Batch Gradient Descent to detect mental disorders of social media users.

Learning rate is very important in MBGD to improve the quality and speed of convergence in training. But in MBGD there are shortcomings in choosing the learning rate so that an optimizer is needed to optimize the learning rate. Many optimizer can be used include AdaGrad [29], RMSProp [30], and Adam [17].

AdaBound is the updated version of Adam has shown faster convergence [31]. Aside from that, AdaBound can optimize so that the convergence can be transformed into global minimum in the end of the training [32]. In addition, Adabound also shows a faster generalization than Adam. Generalization means that the model must adapt well to new, previously unseen/unobserved data. Liu et al [33] proved that, with AdaBound iterations, the cost function converges to a finite value and the corresponding gradient converges to zero.

The data that is used in this study is obesity data obtained from Kaggle Dataset. The data has 3 independent variables and one dependent variable. The level of obesity is classified based on BMI value that is divided into 6 classes (extremely weak, weak, normal, overweight, obesity and extreme obesity).

2. METHOD

The method in this study is Mini-Batch Gradient Descent that has been modified with AdaBound on Takagi Sugeno Kang fuzzy inference. The next step is evaluation of the model using MAD.

2.1 Fuzzy Set

A fuzzy set is a grouping of things based on a linguistic variable which is expressed by a membership function in the universe X. Membership of a value in the set is expressed by the degree of membership whose value is in the interval [0,1].

Definition 1 Let X represent the universe of discourse, x is a member of the universe, X and A represent fuzzy sets. So, the fuzzy set with membership function $\mu_A(x)$ is

$$\mu_A(x): X \rightarrow [0,1] \quad (1)$$

Definition 2 If X is a group of objects symbolized by x, the fuzzy group A in X can be defined as a set of ordered pair.

$$A = \{(x, \mu_A(x)) | x \in X\} \quad (2)$$

with $\mu_A(x)$ as membership degree of x in fuzzy set A that is located between [0,1] [34].

2.2 Membership Function

Membership function is a graph that defines the size of membership degree of each input variable that ranges in 0 to 1. Fuzzy logic uses several type of membership function that is triangular, trapezoid, sigmoid, gaussian, etc. the determination of membership function is depended on the parameter or data that is used [35].

The formula of triangular membership function is:

$$\mu(x; a, b, c) = \begin{cases} \frac{(x-a)}{(b-a)}, & a \le x \le b \\ \frac{(c-x)}{(c-b)}, & b \le x \le c \\ 0, & others \end{cases} \quad (3)$$

The alternative expression of said equation can be presented using min and max:

$$\mu(x; a, b, c) = \max\left(\left(min\left(\frac{x-a}{b-a} - \frac{c-x}{c-b}\right)\right), 0\right) \quad (4)$$

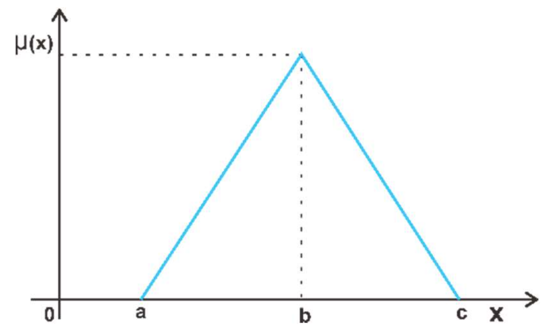

Figure 1: The representstion of trisngulsr membership function

2.3 Takagi Sugeno Kang

TSK is a fuzzy inference method that has fuzzy logic control which is more efficient than the Tsukamoto and Mamdani fuzzy inference methods because the output produced is a constant or linear function. The reasoning using the TSK method is to build an input-output relationship based on the if-then rule with the max or min method. There are several steps used to get the output of the TSK method as follows.

The first step is fuzzification. Fuzzification is the first step in the TSK method to convert numeric variables into linguistic variables. Numerical variables is a value (number) that indicates the size of a variable. While the linguistic variable is the name of a group that represents a certain condition, for example high and low. In changing numeric variables to linguistic variables, membership function curves are used.

The second step is to form the implication function. The relation between input and output in Takagi Sugeno Kang (TSK) model can be explained by fuzzy rule of IF-THEN which is represented such as follows:

$$\begin{aligned} &Rule\ R^k : IF\ x_1\ is\ A_1^k \wedge x_2\ is\ A_2^k, \\ &THEN\ f^k(x) = P_j(x_1, x_2) \end{aligned} \quad (5)$$

with x_1, x_2 is an input variable, A_i^k is a fuzzy set, $\wedge$ is a conjugate operator, and $P_j(x_1, x_2)$ is a polynomial degree d.

Definition 3 Takagi Sugeno Kang (TSK) system in equation (5) can be defined as [36]:

1. Zero order if $P_j(x_1, x_2) = b_j$, where $b_j \in \mathbb{R}$, which means the consequence function is a constant (polynomial degree $d = 0$).
2. First order if $P_j(x_1, x_2) = w_{1j}x_1 + v_{1j}x_2 + b_j$, where $w_{1j}, v_{1j} \in \mathbb{R}$, which means the consequence function is a linier (polynomial degree $d = 1$).
3. High order if $P_j(x_1, x_2) = w_{mj}x_1^m + \ldots + w_{1j}x_1 + v_{mj}x_2^m + \ldots + v_{1j}x_2 + b_j$, where $m \geq 2, w_{kj}, v_{kj} \in \mathbb{R}$ and $k = 2,3,\ldots,m$, which means the consequence function is non linier (polynomial degree $d > 1$).

The last step of TSK fuzzy is defuzzification. The defuzzification is done by calculation average value of fuzzy rules composition and the output of this process is a crisp number.

$$Y = \frac{\sum_{i=1}^{n} \alpha_i y_i}{\sum_{i=1}^{n} \alpha_i}, \qquad i = 1,2,3,\ldots,n \qquad (6)$$

with α_i is a degree of membership of the output value in the i-th rule, y_i is a output value in the i-th rule, and n is the number of rules used.

2.4 Mini-Batch Gradient Descent

Mini-Batch Gradient Descent (MBGD) is the improved version which is a combination between Batch Gradient Descent and Stochastic Gradient Descent to overcome the problems that exist in both methods. Stochastic Gradient Descent has a problem with unstable weight update changes because it updates too often and on Batch Gradient Descent it takes a long time to update, so the solution is to use Mini-Batch Gradient Descent [37].

Optimized Mini-Batch Gradient Descent (MBGD) is extremely popular in deep learning, it can also be the solution to train TSK fuzzy system in large dataset and high dimension. MBGD algorithm calculate the gradient of a randomly chosen group of small data in each of its iteration [38], [39]. The mathematic equation of MGBD algorithm is written as follows:

$$\theta_{t+1} = \theta_t - \eta \times \nabla_\theta MSE(z_t, \theta_t) \qquad (7)$$

with η is a learning rate and z_t is a randomly chosen mini-batch.

2.5 AdaBound

AdaBound is the improved version of Adam optimization method, it has faster convergence compared to Adam and gives a solution with satisfying generalization performance [31]. AdaBound applies dynamic boundaries on learning rate by limiting both from up and bottom so that the learning rate is neither too big nor too small [40]. In determining upper limit and lower limit, the function that is used is stated below:

$$l(k) = 0.01 - \frac{0.01}{(1-\beta_2)k + 1}$$
$$u(k) = 0.01 - \frac{0.01}{(1-\beta_2)k} \qquad (8)$$

In the initial step of training ($k = 0$), the limit is $[0, +\infty)$. During the training ($k \to \infty$), the limit is close to $[0.01, 0.01]$ [40] [41].

2.6 Mean Absolute Deviation (MAD)

Mean Absolute Deviation (MAD) is a method used to evaluate forecasting methods using the number of absolute errors. Mean Absolute Deviation (MAD) is used to measure prediction accuracy by averaging the predicted error (absolute value of each error). MAD is useful for measuring prediction error in the same units as the data real [42], [43]. The formula for calculating MAD value is as follows:

$$MAD = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{n} \qquad (9)$$

with y_i is a real data, $\hat{y}_i$ is a prediction data, and n is the amount of data.

3. EXPERIMENT AND RESULT

The data that is used in this study is an obesity data extracted from Kaggle Dataset (https://www.kaggle.com/yersever/500-person-gender-height-weight-bodymassindex). There are 500 data sets with 3 independent variables and 1 dependent variable, but only 486 data can be used. There are 6 BMI classification indexes, including:

i. Extremely weak is denoted by 0
ii. Weak is denoted by 1
iii. Normal is denoted by 2
iv. Overweight is denoted by 3
v. Obesity is denoted by 4
vi. Extreme obesity is denoted by 5

Table 1: Obesity Data.

Gender	Height (cm)	Weight (kg)	BMI
Male	174	96	4
Male	189	87	2
Female	185	110	4

Journal of Theoretical and Applied Information Technology
30th April 2022. Vol.100. No 8

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

⋮	⋮	⋮	⋮
Male	150	95	5
Male	173	131	5

The content of the BMI variable contained in the kaggle is a classification index so it is necessary to calculate the BMI value using the following formula:

$$BMI = \frac{Weight(kg)}{Height(m^2)} \quad (10)$$

After being calculated using the formula, the following results were obtained.

Table 1: Obesity Data.

Gender	Height (cm)	Weight (kg)	BMI
Male	174	96	31.708
Male	189	87	24.355
Female	185	110	32.14
⋮	⋮	⋮	⋮
Male	150	95	42.22
Male	173	131	43.77

Shown below are the steps used in the method of Mini-Batch Gradient Descent that has been modified with AdaBound on Takagi Sugeno Kang inference fuzzy model:

3.1. Fuzzification

Fuzzification is a process of changing numeric variable (nonfuzzy variable) into linguistic variable (fuzzy variable) using the fuzzification formula so that fuzzy values are obtained. The fuzzification formula used is the triangular membership function as written in equation (3). The membership function curve for each variable can be seen in the image below.

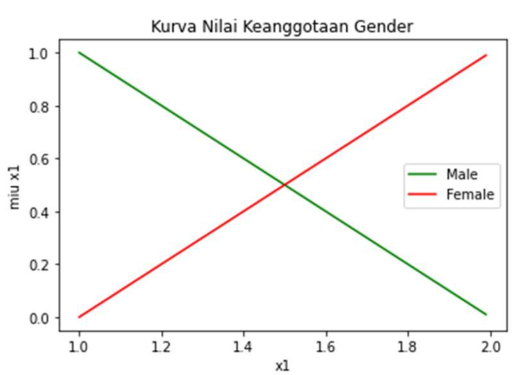

Figure 2: Gender variable membership function curve

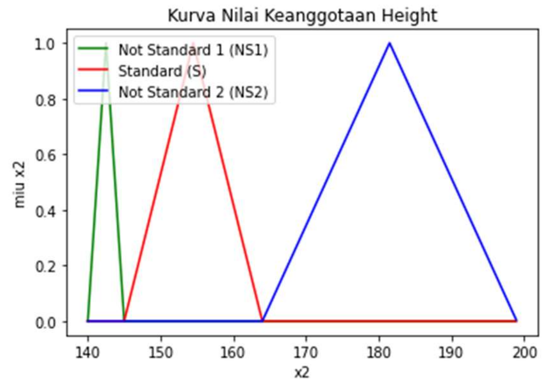

Figure 3: Height variable membership function curve

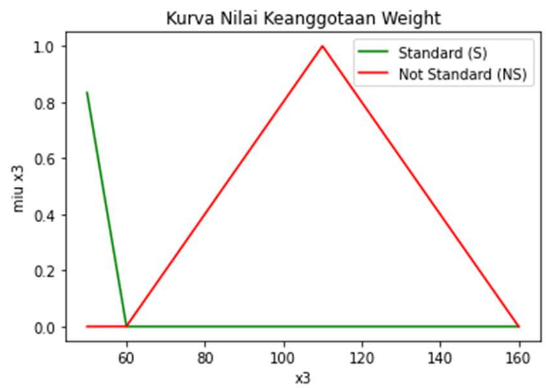

Figure 4: Weight variable membership function curve

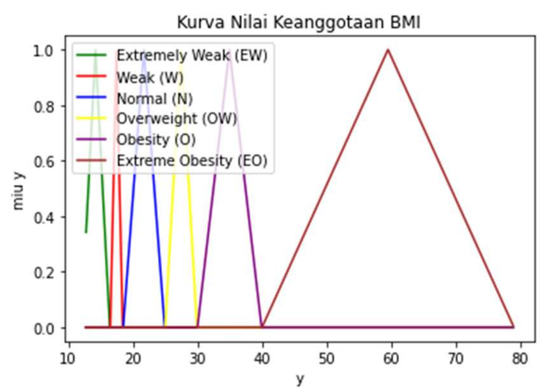

Figure 5: Gender variable membership function curve

The formula for calculating the membership value based on the membership function curve above is as follows.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

a. Gender Variable

$$\mu_{men[x]} = \begin{cases} 1, & x \leq 1 \\ \frac{2-x}{2-1}, & 1 < x < 2 \\ 0, & x \geq 2 \end{cases}$$

$$\mu_{woman[x]} = \begin{cases} 0, & x \leq 1 \\ \frac{x-1}{2-1}, & 1 < x < 2 \\ 1, & x \geq 2 \end{cases}$$

b. Height Variable

$$\mu_{NS\ [x]} = \begin{cases} \frac{x-140}{142.5-140}, & 140 < x < 142.5 \\ \frac{145-x}{145-142.5}, & 142.5 \leq x < 145 \\ 0, & others \end{cases}$$

$$\mu_{S[x]} = \begin{cases} \frac{x-145}{154.5-145}, & 145 < x < 154.5 \\ \frac{164-x}{164-154.5}, & 154.5 \leq x < 164 \\ 0, & others \end{cases}$$

$$\mu_{NS\ [x]} = \begin{cases} \frac{x-164}{181.5-164}, & 164 < x < 181.5 \\ \frac{199-x}{199-181.5}, & 181.5 \leq x < 199 \\ 0, & others \end{cases}$$

c. Weight Variable

$$\mu_{S[x]} = \begin{cases} 1, & x \leq 36 \\ \frac{60-x}{60-36}, & 36 < x < 60 \\ 0, & x \geq 60 \end{cases}$$

$$\mu_{NS[x]} = \begin{cases} \frac{x-60}{110-60}, & 60 < x < 110 \\ \frac{160-x}{160-110}, & 110 \leq x < 160 \\ 0, & others \end{cases}$$

d. BMI Variable

$$\mu_{EW[x]} = \begin{cases} \frac{x-12}{14.2-12}, & 12 < x < 14.2 \\ \frac{16.4-x}{16.4-14.2}, & 14.2 \leq x < 16.4 \\ 0, & others \end{cases}$$

$$\mu_{W[x]} = \begin{cases} \frac{x-16.5}{17.45-16.5}, & 16.5 < x < 17.45 \\ \frac{18.4-x}{1684-17.45}, & 17.45 \leq x < 18.4 \\ 0, & others \end{cases}$$

$$\mu_{N[x]} = \begin{cases} \frac{x-18.5}{21.7-18.5}, & 18.5 < x < 21.7 \\ \frac{24.9-x}{24.9-21.7}, & 21.7 \leq x < 24.9 \\ 0, & others \end{cases}$$

$$\mu_{OW[x]} = \begin{cases} \frac{x-25}{27.45-25}, & 25 < x < 27.45 \\ \frac{29.9-x}{29.9-27.45}, & 27.45 \leq x < 29.9 \\ 0, & others \end{cases}$$

$$\mu_{O[x]} = \begin{cases} \frac{x-30}{134.95-30}, & 30 < x < 34.95 \\ \frac{39.9-x}{39.9-34.95}, & 34.95 \leq x < 39.9 \\ 0, & others \end{cases}$$

$$\mu_{EO[x]} = \begin{cases} \frac{x-40}{59.5-40}, & 40 < x < 59.5 \\ \frac{79-x}{79-59.5}, & 59.5 \leq x < 79 \\ 0, & others \end{cases}$$

The results of the membership value obtained after all data are calculated using the above formula are shown in Table 2.

Table 2: Obesity Membership Value.

Gender	Height (cm)	Weight (kg)	BMI
1	0.571	0.72	0.202
1	0.571	0.54	0.281
0	0.8	1	0.404
⋮	⋮	⋮	⋮
1	0.526	0.7	0.103
1	0.514	0.58	0.154

3.2. Implication Function

According to the formed rule, there are total of 24 rules that can be used to process the data, which are:

[R1] If X_1 female, X_2 height not standard, X_3 weight not standard, so Y obesity
[R2] If X_1 male, X_2 height standard, X_3 weight not standard, so Y extreme obesity

[R3] If X_1 female, X_2 height standard, X_3 weight not standard, so Y extreme obesity

⋮

[R23] If X_1 male, X_2 height standard, X_3 weight tidak standar, so Y overweight

[R24] If X_1 female, X_2 height not standard, X_3 weight standard, so Y weak

3.3. Mini-Batch Gradient Descent with AdaBound

The obtained consequence value in each rule is then calculated using Mini-Batch Gradient Descent optimized with AdaBound. To get the results of the linear regression equation is calculated using equation (7). This calculation uses 1000 iterations.

The initial values are known as $b_0 = 0$ and $b_1 = b_2 = b_3 = 1$. In the first iteration, the values of b_0, b_1, b_2 and b_3 can be calculated using the initial b_0, learning rate, and gradient values. An example of the calculation in rule 1 can be seen as below:

$$\begin{aligned} b_0 &= b_0 - learning\ rate \times gradien\ b_0 \\ &= 0 - 0.01 \times (-0.07827) \\ &= 0.0007827 \end{aligned}$$

$$\begin{aligned} b_1 &= b_1 - learning\ rate \times gradien\ b_1 \\ &= 1 - 0.01 \times (0) \\ &= 1 \end{aligned}$$

$$\begin{aligned} b_2 &= b_2 - learning\ rate \times gradien\ b_2 \\ &= 1 - 0.01 \times (-0.24011) \\ &= 1.0024011 \end{aligned}$$

$$\begin{aligned} b_3 &= b_3 - learning\ rate \times gradien\ b_3 \\ &= 1 - 0.01 \times (-0.14686) \\ &= 1.0014686 \end{aligned}$$

For the next iteration, use the same method until the 1000th iteration so that the values for $b_0 = 0.1451, b_1 = 1,\ b_2 = 0.7781$ and $b_3 = 1.6123$ are obtained. So that the linear regression equation of rule 1 to rule 24 is obtained as follows:

$$y_1 = 0.1451 + X_1 + 0.7781X_2 + 1.6123X_3$$
$$y_2 = -0.1993 + X_1 + 0.1909X_2 + 0.8382X_3$$
$$y_3 = -0.1424 + X_1 + 0.2746X_2 + 0.7704X_3$$
$$\vdots$$
$$y_{23} = 0.0382 + X_1 + 0.1473X_2 + 0.4367X_3$$
$$y_{24} = 0.1367 + X_1 + 0.2224X_2 + 0.4367X_3$$

3.4. Defuzzification

Defuzzification is a fuzzy process with a goal to transform the fuzzy number into crisp number. The defuzzification value is calculated using equation (6), with the results obtained as follows:

Table 2: Defuzzification.

Real Data (y_i)	Prediction ($\hat{y}_i$)
31.708	35.704
24.355	24.427
32.14	45.424
⋮	⋮
42.22	41.045
43.77	64.827

The results of defuzzification in Table 3 are being evaluated using MAD. The MAD calculation is being shown below:

$$\begin{aligned} \text{MAD} &= \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{n} \\ &= \frac{|31.708 - 35.704| + \cdots + |43.77 - 64.827|}{486} \\ &= 14.54 \end{aligned}$$

Based on MAD using the formula [42], the resulting error value is 14.549 which means the predictive ability is effective.

Figure 6 below shows the results of the comparison of each data between the data real and the predicted data. The red line shows the data real and the blue line shows the predicted data.

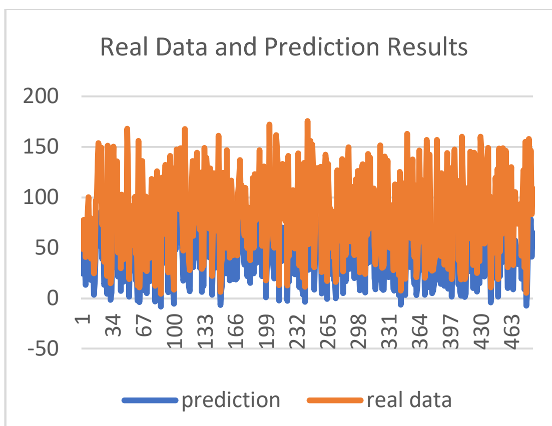

Figure 6: The representstion of prediction and real data

4. CONCLUSION

The model used in this study is the first order fuzzy Takagi Sugeno Kang inference model modified using Mini-Batch Gradient Descent with AdaBound. Mini-Batch Gradient Descent is used to modify the consequent on TSK which is a linear function of each obtained rules. Adabound is an optimizer that is used to overcome the problem of choosing a learning rate in MBGD by using a dynamic limit on the learning rate because in MBGD the selection of a learning rate is very important to improve the quality and speed of training.

The data used is Body Mass Index data which has 6 classification indices. The rules obtained in the TSK inference model based on the data are 32 rules, but only 24 rules can be used. So that from the 24 rules obtained, the defuzzification value is calculated using the weighted average method to obtain firm numbers. The evaluation model used is MAD and the error value is 14,549, which means the method used is effective and efficient. Comparing this method with other optimization methods such as AdaGrad and RMSProp can be done in the future research.

REFERENCES:

[1] H. D. Wijaya and S. Dwiasnati, "Implementasi Data Mining dengan Algoritma Naive Bayes pada Penjualan Obat," *J. Inform.*, vol. 7, no. 1, pp. 1–7, 2020, doi: 10.31311/ji.v7i1.6203.

[2] A. Eka Haryati and A. Dahlan Yogyakarta, "Implementation of Minkowski-Chebyshev Distance in Fuzzy Subtractive Clustering," *EKSAKTA journal.uii.ac.id*, vol. 2, no. 2, pp. 82–87, 2021, doi: 10.20885/EKSAKTA.vol2.iss1.art.

[3] A. E. Haryati, S. Surono, M. P. Matematika, U. A. Dahlan, and U. A. Dahlan, "Comparative study of distance measures on fuzzy subtractive clustering 1," vol. 14, no. 2, pp. 137–145, 2021, doi: 10.14710/medstat.14.2.137-145.

[4] C. Fu, W. Lu, W. Pedrycz, and J. Yang, "Fuzzy Granular Classification Based on the Principle of Justifiable Granularity," *Knowledge-Based Syst.*, vol. 170, pp. 89–101, 2019, doi: 10.1016/j.knosys.2019.02.001.

[5] N. Gelley and R. Jang, *Fuzzy Logic Toolbox*. USA: Mathwork, Inc, 2000.

[6] S. Kusumadewi and H. Purnomo, *Aplikasi Logika Fuzzy untuk Pendukung Keputusan*, Edisi 2. Yogyakarta: Graha Ilmu, 2013.

[7] R. Taber, *The Fuzzy Systems Handbook: A Practitioner's Guide to Building, Using, and Maintaining Fuzzy Systems (Earl Cox)*, vol. 37, no. 2. Massachusetts: Academic Press, Inc, 1995.

[8] S. Batubara, "Analisis Perbandingan Metode Fuzzy Mamdani Dan Fuzzy Sugeno Untuk Penentuan Kualitas Cor Beton Instan," *It J. Res. Dev.*, vol. 2, no. 1, pp. 1–11, 2017, doi: 10.25299/itjrd.2017.vol2(1).644.

[9] W. Zou, C. Li, and N. Zhang, "A T-S Fuzzy Model Identification Approach Based on a Modified Inter Type-2 FRCM Algorithm," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 3, pp. 1104–1113, 2018, doi: 10.1109/TFUZZ.2017.2704542.

[10] X. Gu and S. Wang, "Bayesian Takagi-Sugeno-Kang fuzzy model and its joint learning of structure identification and parameter estimation," *IEEE Trans. Ind. Informatics*, vol. 14, no. 12, pp. 5327–5337, 2018, doi: 10.1109/TII.2018.2813977.

[11] X. li Wang, W. xin Xie, and L. qun Li, "Structure Identification of Recursive TSK Particle Filtering via Type-2 Intuitionistic Fuzzy Decision," *Int. J. Fuzzy Syst.*, vol. 23, no. 5, pp. 1294–1312, 2021, doi: 10.1007/s40815-020-01021-6.

[12] J. Wang, Z. Yi, J. M. Zurada, B.-L. Lu, and H. Yin, *Advances in Neural Network - ISNN 2006*, Part 1. 2006.

[13] Y. Li, P. Qian, S. Wang, and S. Wang, "Novel multi-view Takagi–Sugeno–Kang fuzzy system for epilepsy EEG detection," *J. Ambient Intell. Humaniz. Comput.*, vol. 5, no. 0123456789, 2021, doi: 10.1007/s12652-021-03189-7.

[14] R. Raj and B. M. Mohan, "Modeling and analysis of the simplest fuzzy PID controller of Takagi–Sugeno type with modified rule base," *Soft Comput.*, vol. 22, no. 15, pp. 5147–5161, 2018, doi: 10.1007/s00500-017-2674-8.

[15] A. Du *et al.*, "Assessing the Adequacy of Hemodialysis Patients via the Graph-Based Takagi-Sugeno-Kang Fuzzy System," *Comput. Math. Methods Med.*, vol. 2021, no. Ml, 2021, doi: 10.1155/2021/9036322.

[16] X. Meng and J. Zhang, "Anxiety Recognition of College Students Using a Takagi-Sugeno-Kang Fuzzy System Modeling Method and Deep Features," *IEEE Access*, vol. 8, pp. 159897–159905, 2020, doi: 10.1109/ACCESS.2020.3021092.

[17] D. P. Kingma and J. Lei Ba, "ADAM: A Method for Stochastic Optimization," *ICLR*, pp. 1–15, 2015, [Online]. Available:

https://arxiv.org/pdf/1412.6980.pdf %22 entire document.

[18] M. J. Kochenderfer and T. A. Wheeler, *Algorithms for Optimization*. The MIT Press Cambridge, UK, 2019.

[19] G. K. WEN, Y. BIN DASRIL, N. BUJANG, M. MOHAMAD, M. D. H. GAMAL, and L. C. SOON, "Hybridization gradient descent search with artificial bees colony algorithm in general global optimization problems," *J. Theor. Appl. Inf. Technol.*, vol. 99, no. 4, pp. 999–1008, 2021.

[20] G. K. Wen, Y. Dasril, and A. R. Othman, "A class of ABCED conjugate gradient method in solving general global optimization problems," *J. Theor. Appl. Inf. Technol.*, vol. 96, no. 23, pp. 7984–7995, 2018.

[21] D. Yi, J. Ahn, and S. Ji, "An Effective Optimization Method for Machine Learning Based on ADAM," *Appl. Sci.*, vol. 10, no. 3, 2020, doi: 10.3390/app10031073.

[22] S. Nakasima-López, J. R. Castro, M. A. Sanchez, O. Mendoza, and A. Rodríguez-Díaz, *An approach on the implementation of full batch, online and mini-batch learning on a Mamdani based neuro-fuzzy system with center-of-sets defuzzification: Analysis and evaluation about its functionality, performance, and behavior*, vol. 14, no. 9. 2019.

[23] E. Bisong, "Building Machine Learning and Deep Learning Models on Google Cloud Platform," *Build. Mach. Learn. Deep Learn. Model. Google Cloud Platf.*, pp. 203–207, 2019, doi: 10.1007/978-1-4842-4470-8.

[24] P. Murugan and S. Durairaj, "Regularization and Optimization Strategies in Deep Convolutional Neural Network," pp. 1–15, 2017, [Online]. Available: http://arxiv.org/abs/1712.04711.

[25] H. Xie, F. Yang, M. Hua, S. Liu, J. Hu, and Y. He, "Grounding grid corrosion detection based on mini-batch gradient descent and greedy method," *AIP Adv.*, vol. 11, no. 6, 2021, doi: 10.1063/5.0051678.

[26] P. Gou and J. Yu, "A nonlinear ANN equalizer with mini-batch gradient descent in 40Gbaud PAM-8 IM/DD system," *Opt. Fiber Technol.*, vol. 46, no. September, pp. 113–117, 2018, doi: 10.1016/j.yofte.2018.09.015.

[27] J. Li, X. Li, and L. Zhao, "HYPERSPECTRAL UNMIXING VIA PROJECTED MINI-BATCH GRADIENT DESCENT," vol. 0, no. 2, pp. 0–3.

[28] K. Rahul, "Detection of Mental Disorders From Online Social Behaviors Using Mini Batch Gradient," vol. 25, no. 10, pp. 12–16, 2018.

[29] J. Duchi, E. Hazan, and Y. Singer, "Adaptive Subgradient Methods for Online Learning and Stochastic Optimization," *J. Mach. Learn. Res.*, vol. 12, pp. 2121–2159, 2011.

[30] T. Tieleman and G. Hinton, "Lecture 6.5-RMSProp: Divide the Gradient by a Running Average of its Recent Magnitude," *COURSERA Neural Netw. Mach. Learn.*, vol. vol.4, pp. 26–31, 2012.

[31] L. Luo, Y. Xiong, Y. Liu, and X. Sun, "Adaptive Gradient Methods With Dynamic Bound of Learning Rate," no. 2018, pp. 1–19, 2019.

[32] R. Wang, H. Chen, and C. Guan, "Random convolutional neural network structure: An intelligent health monitoring scheme for diesel engines," *Meas. J. Int. Meas. Confed.*, vol. 171, no. December 2020, p. 108786, 2021, doi: 10.1016/j.measurement.2020.108786.

[33] J. Liu, J. Kong, D. Xu, M. Qi, and Y. Lu, "Convergence analysis of AdaBound with relaxed bound functions for non-convex optimization," *Neural Networks*, vol. 145, pp. 300–307, 2022, doi: 10.1016/j.neunet.2021.10.026.

[34] H.-J. Zimmermann, *Fuzzy Set Theory and Its Applications*. 2001.

[35] M. H. Azam, M. H. Hasan, S. Hassan, and S. J. Abdulkadir, "Fuzzy Type-1 Triangular Membership Function Approximation Using Fuzzy C-Means," *2020 Int. Conf. Comput. Intell. ICCI 2020*, no. October, pp. 115–120, 2020, doi: 10.1109/ICCI51257.2020.9247773.

[36] K. Wiktorowicz and T. Krzeszowski, "Approximation of two-variable functions using high-order Takagi–Sugeno fuzzy systems, sparse regressions, and metaheuristic optimization," *Soft Comput.*, vol. 24, no. 20, pp. 15113–15127, 2020, doi: 10.1007/s00500-020-05238-3.

[37] M. W. Khan, M. Zeeshan, and M. Usman, "Traffic scheduling optimization in cognitive radio based smart grid network using mini-batch gradient descent method," *Iber. Conf. Inf. Syst. Technol. Cist.*, vol. 2019-June, no. June, pp. 19–22, 2019, doi: 10.23919/CISTI.2019.8760693.

[38] Y. Cui, D. Wu, and J. Huang, "Optimize TSK Fuzzy Systems for Classification Problems: Minibatch Gradient Descent with Uniform Regularization and Batch Normalization,"

IEEE Trans. Fuzzy Syst., vol. 28, no. 12, pp. 3065–3075, 2020, doi: 10.1109/TFUZZ.2020.2967282.

[39] A. Mustapha, L. Mohamed, and K. Ali, *An Overview of Gradient Descent Algorithm Optimization in Machine Learning: Application in the Ophthalmology Field*, vol. 1207 CCIS. Springer International Publishing, 2020.

[40] D. Wu, Y. Yuan, J. Huang, and Y. Tan, "Optimize TSK Fuzzy Systems for Regression Problems: Minibatch Gradient Descent with Regularization, DropRule, and AdaBound (MBGD-RDA)," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 5, pp. 1003–1015, 2020, doi: 10.1109/TFUZZ.2019.2958559.

[41] D. Wu, "MBGD-RDA Training and Rule Pruning for Concise TSK Fuzzy Regression Models," pp. 1–11, 2020, [Online]. Available: http://arxiv.org/abs/2003.00608.

[42] U. Khair, H. Fahmi, S. Al Hakim, and R. Rahim, "Forecasting Error Calculation with Mean Absolute Deviation and Mean Absolute Percentage Error," *J. Phys. Conf. Ser.*, vol. 930, no. 1, pp. 0–6, 2017, doi: 10.1088/1742-6596/930/1/012002.

[43] A. Krisma, M. Azhari, and P. P. Widagdo, "Perbandingan Metode Double Exponential Smoothing Dan Triple Exponential Smoothing Dalam Parameter Tingkat Error Mean Absolute Percentage Error (MAPE) dan Means Absolute Deviation (MAD)," *Pros. Semin. Nas. Ilmu Komput. dan Teknol. Inf.*, vol. 4, no. 2, pp. 81–87, 2019.