



# Implementation of Takagi Sugeno Kang Fuzzy with Rough Set Theory and Mini-Batch Gradient Descent Uniform Regularization

Sugiyarto Surono <sup>1</sup>, Zani Anjani Rafsanjani Hsm <sup>1</sup>, Deshinta Arrova Dewi <sup>2\*</sup>,  
Annisa Eka Haryati <sup>3</sup>, Tommy Tanu Wijaya <sup>4</sup>

<sup>1</sup> Department of Mathematics, Ahmad Dahlan University, Yogyakarta, Indonesia.

<sup>2</sup> Faculty of Data Science and Information Technology, INTI International University, Nilai, Malaysia.

<sup>3</sup> Master of Mathematics Education, Ahmad Dahlan University, Yogyakarta, Indonesia.

<sup>4</sup> School of Mathematical Sciences, Beijing Normal University, China.

## Abstract

The Takagi Sugeno Kang (TSK) fuzzy approach is popular since its output is either a constant or a function. Parameter identification and structure identification are the two key requirements for building the TSK fuzzy system. The input utilized in fuzzy TSK can have an impact on the number of rules produced in such a way that employing more data dimensions typically results in more rules, which causes rule complexity. This issue can be solved by employing a dimension reduction technique that reduces the number of dimensions in the data. After that, the resulting rules are improved with MBGD (Mini-Batch Gradient Descent), which is then altered with uniform regularization (UR). UR can enhance the classifier's fuzzy TSK generalization performance. This study looks at how the rough sets method can be used to reduce data dimensions and use Mini Batch Gradient Descent Uniform Regularization (MBGD-UR) to optimize the rules that come from TSK. 252 respondents' body fat data were utilized as the input, and the mean absolute percentage error (MAPE) was used to analyze the results. Jupyter Notebook software and the Python programming language are used for data processing. The analysis revealed that the MAPE value was 37%, falling into the moderate area.

## Keywords:

Rough Set;  
Takagi Sugeno Kang Fuzzy;  
Mini Batch Gradient Descent;  
Uniform Regularization.

## Article History:

<b>Received:</b>	05	October	2022
<b>Revised:</b>	26	January	2023
<b>Accepted:</b>	11	March	2023
<b>Available online:</b>	03	May	2023

## 1- Introduction

Each data set contains a variety of dimensions, ranging in size from small to large, with large dimensions frequently being observed to have a high level of complexity, which has an impact on the number of rules produced when employed in a fuzzy inference system. This is demonstrated by the fact that using large data dimensions as input typically results in the creation of more rules. The Takagi Sugeno Kang (TSK) fuzzy used in one of the studies [1] was found to have limitations when the input used was large. This TSK, which Takagi and Sugeno first developed in 1985 [2], is a form of fuzzy inference that is frequently employed for prediction or classification [3–5]. TSK was utilized by Shaheen et al. [6] in order to carry out the AP-TSK-PID method while dealing with stochastic and non-stochastic uncertainties in nonlinear dynamic systems. In the field of medicine, it has also been used by Du et al. [7] to forecast how well hemodialysis patients will respond to their treatment and by Pan et al. [8] to address the issue of type 1 diabetes's difficulty controlling their blood glucose levels.

\* **CONTACT:** [deshinta.ad@newinti.edu.my](mailto:deshinta.ad@newinti.edu.my)

**DOI:** <http://dx.doi.org/10.28991/ESJ-2023-07-03-09>

© 2023 by the authors. Licensee ESJ, Italy. This is an open access article under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<https://creativecommons.org/licenses/by/4.0/>).

To achieve improved operating time performance, TSK fuzzy can be tuned using Gradient Descent (GD) [9]. This GD is one of the techniques typically employed in optimization to reduce the cost function in machine learning, and it is typically carried out by updating each parameter depending on previous steps [10, 11]. GD can be divided into three categories: batch, stochastic, and mini-batch. Prior to this, several studies employed the batch type and stochastic type, respectively, for classification [12–16]. Due to its tendency to have a lesser computing load and a faster convergence because only the data from a batch is used in each iteration, the Mini-Batch Gradient Descent (MBGD) method was adopted in this work [17, 18].

MBGD has been the subject of numerous studies, including its application by Gou and Yu [19] to effectively train the ANN equalizer, Messaoud et al. [20] to maximize the IoT 4.0 market, and Hu et al. [21] to combine it with MMLDA to predict lncRNA illness associations. Additionally, the regularization strategy was employed to prevent overfitting and boost generalization because it can help the algorithm become more universal by preventing coefficients from being used to match the training sample data [22]. Regularization is described by Kukačka et al. [23] as "any change made to a learning system intended to reduce generalization mistakes and not training errors." It is thought that this method is necessary to stabilize numerical calculations [24].

The rough set is a set theory extension that Pawlak first proposed in 1982. It is a subset of the universe that is characterized by two original sets known as the upper and lower approximation sets. Equivalence relations, specifically reflexive, symmetrical, and transitive relations, are the primary component of the rough set model [25]. This was carried out by Wang et al. [26] to carry out feature selection in the genetic algorithm, and it was observed that the picked features produced good results. This technique for making decisions was also employed in Zhan et al. [27]'s study, and Jothi et al. [28] used it to categorize leukemia by identifying prominent traits.

This study builds on earlier work by Cui et al. [1] on the fuzzy TSK system with optimization based on Mini Batch Gradient Descent (MBGD) on classification issues. To enhance the TSK fuzzy classification's generalization performance and prevent overfitting, researchers utilize uniform regularization (UR), as in Cui et al. [1]. Further, the researcher suggests applying the rough set method. The following are the primary contributions of this study:

1. The rough set technique is utilized to minimize the data's dimensions. The experimental findings demonstrate that the rough set enhances the rules produced by TSK fuzzy classification.
2. Additional researchers who used the rough set and UR approach found that the model falls into the reasonable category.

## 2- Research Methodology

### 2-1- Rough Set

The rough set is one of the dimension reduction techniques developed by Pawlak in 1982 with its principle associated with a reflexive, symmetrical, and transitive equivalence relation [25]. In the rough set, there is an information system that can be represented in the form of a table  $S = (U, A)$ , where  $U$  is a non-empty finite set of objects and  $A$  is a non-empty finite set of attributes [29]. If the information table is added with the output of the classification, it will become a decision system table denoted by  $S = (U, A \cup \{d\})$ , where  $d \notin A$  is the decision attribute. Indiscernibility relation is a relationship that cannot be separated because an object can have the same value for a condition attribute. This can happen at the time of the decision system. Suppose  $S = (U, A)$  is an information system and  $B \subseteq A$ . The indiscernibility relation of objects according to attribute  $B$  is denoted by  $IND_B(B)$  can be defined as follows:

$$IND_B(B) = \{(X, X') \in U^2 | \forall a \in B, a(x) = a(x')\} \quad (1)$$

Attributes in the rough set can be removed without losing their true value by using core and reduct. Reduct is the set of attributes that can produce the same classification as if all attributes were used. While attributes that are not reducts are attributes that are not useful in the classification process [29]. Core is the intersection of all reducts, so the core is in every reduction, that is, every core attribute is included in every reduction. Suppose  $B \subseteq A$  and core of  $B$  is the set of all dispensable attributes of  $B$  then core can be defined as follows [30]:

$$Core(B) = \cap Red(B) \quad (2)$$

### 2-2- Fuzzy Set

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

$$\mu_A(x): X \rightarrow [0,1] \quad (3)$$

Definition 2: The fuzzy set  $A$  in universe  $X$  can be defined as a set of ordered pairs as indicated in the following equation:

$$A = \{(x, \mu_A(x)) | x \in X\} \quad (4)$$

where,  $\mu_A(x)$  is the membership function  $x$  in the fuzzy set  $A$  which lies on the interval  $[0,1]$  [31].

### 2-3-Takagi Sugeno Kang (TSK) Fuzzy

TSK fuzzy system with one input  $x_1$  and  $x_2$  as well as output  $y$  is described by fuzzy inference rules as follows [32]:

$$R_j = \text{IF } x_1 \in F_j(x_1) \text{ AND } x_2 \in G_j(x_2) \text{ THEN } y = P_j(x_1, x_2) \quad (5)$$

where,  $j = 1, 2, \dots, r$ ,  $F_j(x_1)$ ,  $G_j(x_2)$  is a fuzzy set and  $P_j(x_1, x_2)$  is a degree polynomial  $d$ .

Definition 3: TSK system in line with rules is defined as follows [32]:

- The order is zero if  $P_j(x_1, x_2) = b_j$ , where  $b_j \in \mathbb{R}$ , and this means the consequent function is a constant (degree polynomial  $d$  is equal to zero).
- The order is one if  $P_j(x_1, x_2) = w_{1j}x_1 + v_{1j}x_2 + b_j$ , where  $w_{1j}, v_{1j} \in \mathbb{R}$ , and this means the consequent functions are linear (a degree polynomial  $d$  is equal to one).
- The order is high if  $P_j(x_1, x_2) = w_{mj}x_1^m + \dots + w_{1j}x_1 + v_{mj}x_2^m + \dots + v_{1j}x_2 + b_j$ , where  $m \geq 2$ ,  $w_{kj}, v_{kj} \in \mathbb{R}$  and  $k = 2, 3, \dots, m$  and this means the consequent function is nonlinear (a degree polynomial  $d$  is greater than one).

Defuzzification is a fuzzy process aimed at converting fuzzy numbers to crisp numbers. Therefore, the defuzzification value ( $Y^*$ ) was calculated using the following equation:

$$Y^* = \frac{\sum_{i=1}^N a_i y_i}{\sum_{i=1}^N a_i}; i = 1, 2, \dots, N \quad (6)$$

where,  $a_i$  is output value in the  $i$ -th rule and  $y_i$  is output value in the  $i$ -th rule.

### 2-4-Mini Batch Gradient Descent (MBGD)

MBGD is Gradient Descent (GD) method that uses the concept of Mini-Batch to update parameters. Meanwhile, the updated parameter can be defined as follows [17]:

$$\theta = \theta - \eta \cdot \nabla_{\theta} J(\theta; x^{(i:i+n)}; y^{(i:i+n)}) \quad (7)$$

where  $\eta > 0$  is the learning rate (step size) [33].

### 2-5-Uniform Regularization (UR)

UR is a regularization method that forces the rules to have firing levels by minimizing losses [1]. It can be calculated as follows:

$$\ell_{UR} = \sum_{r=1}^R \left( \frac{1}{N} \sum_{n=1}^N \bar{f}_r(x_n) - \tau \right)^2 \quad (8)$$

where  $N$  is the number of training samples and  $\tau$  is the firing level of each rule. Furthermore,  $\ell_{UR}$  is added to the loss function in MBGD-based TSK classification training using Mini-Batch with  $N$  training samples and this is represented as follows:

$$\mathcal{L} = \ell + \alpha \ell_2 + \lambda \sum_{r=1}^R \left( \frac{1}{N} \sum_{n=1}^N \bar{f}_r(x_n) - \frac{1}{R} \right)^2 \quad (9)$$

### 2-6-Mean Absolute Percentage (MAPE)

MAPE is one of the methods normally used to evaluate a model and its value can be determined using the following equation [34]:

$$MAPE = \frac{\sum_{i=1}^n \frac{|y_i - y'_i|}{y_i}}{n} \times 100\% \quad (10)$$

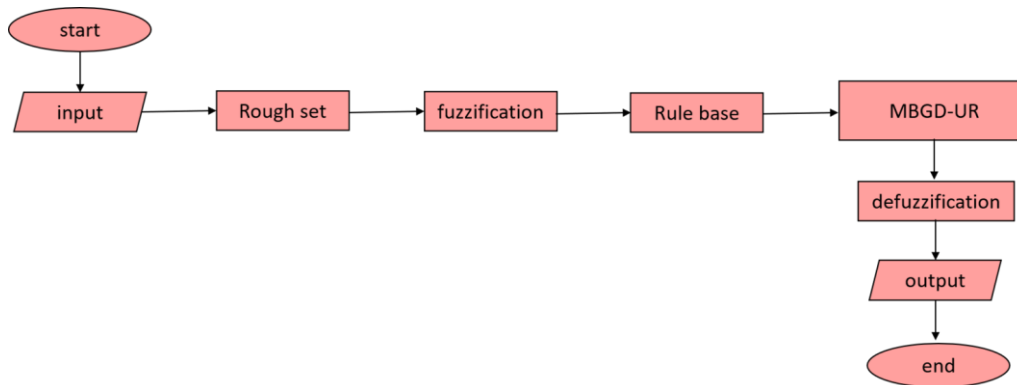
where,  $y_i$  is the  $i$ -th data,  $y'_i$  is the  $i$ -th data for forecasting, and  $n$  is the total data. The prediction criteria for MAPE as indicated by Rohmah et al. [34] are as follows (Table 1):

**Table 1. MAPE Criteria**

MAPE	Prediction Criteria
<10%	Excellent
10% - 20%	Good
20% - 50%	Reasonable
>50%	Bad

### 2-7-Flowchart

The flowchart in this study can be seen in Figure 1.



**Figure 1. Flowchart Rough Set TSK-MBGDUR**

First, a rough set method is utilized to process the data in Figure 1. Afterward, a method known as "fuzzification" is used to transform the rough set's results into fuzzy numbers. The creation of the fundamental rules, IF-THEN, where IF is the antecedent and THEN is the consequent, comes next once the fuzzification phase has been completed. The established rules will be improved using Mini Batch Gradient Descent Uniform Regularization (MBGD-UR). The next and last stage is the defuzzification procedure, which involves converting the fuzzy set back to the crisp set.

### 3- Results

The body fat data from a database known as Kaggle (<https://www.kaggle.com/datasets/fedesoriano/body-fat-prediction-dataset>) was used in this study. It consists of data for 252 respondents with 14 independent variables and 1 dependent variable as indicated in the following Table 2.

**Table 2. Data from PCA Dimension Reduction**

$X_1$	$X_2$	$X_3$	...	$Y$
23	154.25	67.75	...	12.3
22	173.25	72.25	...	6.1
⋮	⋮	⋮	...	⋮
72	190.75	70.5	...	26
74	207.5	70	...	31.9

The dimensions of the data set were reduced using a rough set and the results are presented in Table 3.

**Table 3. Body fat data reduction results**

$X_1$	$X_2$	$X_3$	$Y$
154.25	67.75	1.0708	12.3
173.25	72.25	1.0853	6.1
⋮	⋮	⋮	⋮
190.75	70.5	1.0399	26
207.5	70	1.0271	31.9

The data in Table 3 were further converted into fuzzy numbers, and the results are presented in the following table (Table 4):

**Table 4. Body fat membership value**

$X_1$	$X_2$	$X_3$	$Y$
0	0.2713	0	0
0.3437	0.9823	0	0
⋮	⋮	⋮	⋮
0.4965	0.7462	0	0
0.6493	0.6105	0	0

The input in Table 4 was used to obtain the following rules:

[R1] If  $X_1$  is nonstandard2,  $X_2$  is high nonstandard2,  $X_3$  is high, then  $Y$  is athletic.

[R2] If  $X_1$  is standard,  $X_2$  is high nonstandard2,  $X_3$  is high, then  $Y$  is athletic.

[R3] If  $X_1$  is nonstandard2,  $X_2$  is high nonstandard2,  $X_3$  is high, then  $Y$  is good.

[R4] If  $X_1$  is nonstandard2,  $X_2$  is high nonstandard2,  $X_3$  is high, then  $Y$  is normal.

[R5] If  $X_1$  is nonstandard2,  $X_2$  is high nonstandard2,  $X_3$  is high, then  $Y$  is overweight.

[R6] If  $X_1$  is standard,  $X_2$  is high nonstandard2,  $X_3$  is high, then  $Y$  is good.

[R7] If  $X_1$  is nonstandard1,  $X_2$  is high nonstandard2,  $X_3$  is high, then  $Y$  is athletic.

The similarities in each rule were later determined using MBGD-UR, and the results are indicated as follows:

$$\begin{aligned}
 y_1 &= -0.1515 + 0.9664X_1 + 0.9201X_2 + 0.6203X_3 \\
 y_2 &= -0.2981 + 0.8369X_1 + 0.8939X_2 + 0.7895X_3 \\
 y_3 &= -0.5750 + 0.6356X_1 + 0.6385X_2 + 0.6492X_3 \\
 y_4 &= -0.6431 + 0.8585X_1 + 0.6853X_2 + 0.5158X_3 \\
 y_5 &= -0.5017 + 0.7990X_1 + 0.4268X_2 + 0.6014X_3 \\
 y_6 &= -0.3812 + 0.7074X_1 + 1.1365X_2 + 0.4482X_3 \\
 y_7 &= -0.3093 + 0.9108X_1 + 0.6590X_2 + 0.7610X_3
 \end{aligned} \tag{11}$$

Defuzzification was conducted on the rules obtained, and the results are indicated in Table 5.

**Table 5. Body fat defuzzification results**

$Y$	$Y'$
12.3	18.2045
6.1	19.7421
$\vdots$	$\vdots$
26	20.0780
31.9	20.6161

MAPE value was determined as follows:

$$MAPE = \frac{\sum_{i=1}^n \frac{|y_i - y'_i|}{y_i}}{n} \times 100\% = \frac{\sum_{i=1}^n \frac{|12.3 - 18.2045|}{12.3} + \frac{|6.1 - 19.7421|}{6.1} + \dots + \frac{|31.9 - 20.6161|}{31.9}}{252} \times 100\% = 36.5510\% \approx 37\% \tag{12}$$

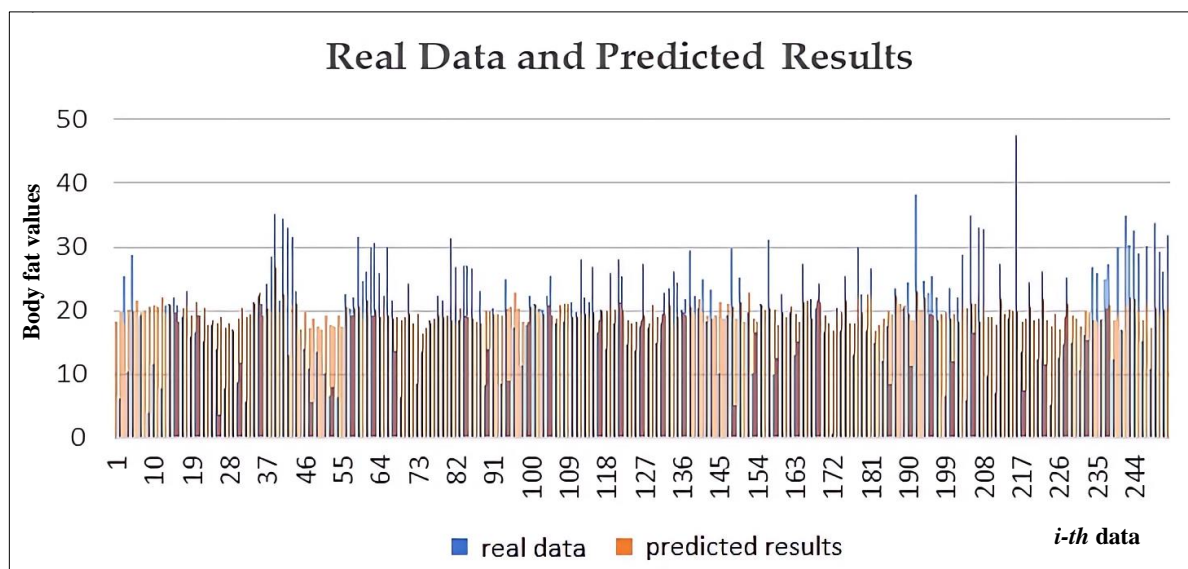
#### 4- Discussion

This study used 252 data with 14 independent variables and 1 dependent variable, and the results of the dimension reduction using the rough set method are presented in Table 3. The 14 variables were discovered to be reduced to 4 variables including weight ( $X_1$ ), height ( $X_2$ ), density ( $X_3$ ), and body fat ( $Y$ ).

Rough set results were subsequently used as input in TSK with each variable subjected to a fuzzification process to convert the data to fuzzy numbers using membership functions. This led to the generation of 7 rules which are in the form of IF-THEN as in Equation 5. Moreover, the consequences for each rule were optimized using MBGD-UR, and the constants generated for each rule were arranged into Equation 8.

The defuzzification process was later used to obtain output in the form of firm numbers. It is important to know that the defuzzification value was determined by multiplying the  $y$  value with the predicate alpha in each rule and dividing it by the total predicate alpha. The defuzzification ( $Y'$ ) was calculated using Equation 6 and the results are shown in Table 5.

Figure 2 shows the results of the representation of Table 5 which compares the predicted defuzzification data with the actual data. In addition, the MAPE value is calculated to determine the accuracy of the model obtained, and a value of 37% is obtained which belongs to the fair category as shown in Table 1.



**Figure 2. Real data and predicted results**

## 5- Conclusion

In order to transform data with big enough dimensions into those with small enough dimensions, the rough set method is employed in this study. The rules are then enhanced in the fuzzy TSK classification process using MBGD modification and UR to get better results. The classification prediction result after the MBGD-UR model modification is more reliable. This is so that the MBGD method's parameters can be updated using the mini-batch idea. By reducing losses, the role of UR in this study can push the rules to have a firing level. Additionally, MAPE (Mean Absolute Percent Error), which is a crucial component in assessing the forecast's accuracy, is employed based on the size of the forecasting variable. The MAPE value is recorded at 37%, indicating that the model is included in the reasonable category. MAPE measures how large the forecast error is in comparison to the actual value of the series. Further study is advised to compare other reduction techniques, such as using a decision tree and the Raw Classification Accuracy (RCA) accuracy, and to determine the minimum variable limit for dimensional reduction based on the analysis findings and conclusions.

## 6- Declarations

### 6-1-Author Contributions

Conceptualization, S.S. and A.E.H.; methodology, S.S. and D.A.D.; software, A.E.H. and Z.A.R.; validation, S.S., D.A.D., and Z.A.R.; formal analysis, S.S.; investigation, S.S.; resources, A.E.H.; data curation, A.E.H. and T.T.W.; writing—original draft preparation, S.S. and T.T.W.; writing—review and editing, S.S., D.A.D., and T.T.W.; visualization, A.E.H.; supervision, S.S.; project administration, Z.A.R. All authors have read and agreed to the published version of the manuscript.

### 6-2-Data Availability Statement

The data presented in this study are available on request from the corresponding author.

### 6-3-Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

### 6-4-Institutional Review Board Statement

Not applicable.

### 6-5-Informed Consent Statement

Not applicable.

### 6-6-Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies have been completely observed by the authors.



## 7- References

- [1] Cui, Y., Wu, D., & Huang, J. (2020). Optimize TSK Fuzzy Systems for Classification Problems: Minibatch Gradient Descent with Uniform Regularization and Batch Normalization. *IEEE Transactions on Fuzzy Systems*, 28(12), 3065–3075. doi:10.1109/TFUZZ.2020.2967282.
- [2] Suryatini, F., Maimunah, M., & Fauzandi, F. I. (2019). Implementation of a Drip Irrigation Control System Using the IoT Concept Based on Fuzzy Takagi-Sugeno Logic. *JTERA (Jurnal Teknologi Rekayasa)*, 4(1), 115. doi:10.31544/jtera.v4.i1.2019.115-124.
- [3] Muhajirah, A., Safitri, E., Mardiana, T., Hartina, H., & Setiawan, A. (2019). Analysis of the Accuracy Level of the Neuro Fuzzy Method in Predicting HDI Data in NTB. *JTAM | Journal of Mathematical Theory and Application*, 3(1), 58. doi:10.31764/jtam.v3i1.769. (In Indonesian).
- [4] Hajar, S., Badawi, M., Setiawan, Y. D., Siregar, M. N. H., & Windarto, A. P. (2020). Calculation Prediction of Mahanda Tofu Production with Fuzzy Sugeno Technique. *J-SAKTI (Journal of Computer Science and Informatics)*, 4(1), 210-219. doi:10.30645/j-sakti.v4i1.200. (In Indonesian).
- [5] Hu, Z., Wang, J., Zhang, C., Luo, Z., Luo, X., Xiao, L., & Shi, J. (2022). Uncertainty Modeling for Multicenter Autism Spectrum Disorder Classification Using Takagi-Sugeno-Kang Fuzzy Systems. *IEEE Transactions on Cognitive and Developmental Systems*, 14(2), 730–739. doi:10.1109/TCDS.2021.3073368.
- [6] Shaheen, O., El-Nagar, A. M., El-Bardini, M., & El-Rabaie, N. M. (2020). Stable adaptive probabilistic Takagi–Sugeno–Kang fuzzy controller for dynamic systems with uncertainties. *ISA Transactions*, 98, 271–283. Doi:10.1016/j.isatra.2019.08.035.
- [7] Du, A., Shi, X., Guo, X., Pei, Q., Ding, Y., Zhou, W., Lu, Q., & Shi, H. (2021). Assessing the Adequacy of Hemodialysis Patients via the Graph-Based Takagi-Sugeno-Kang Fuzzy System. *Computational and Mathematical Methods in Medicine*, 2021. doi:10.1155/2021/9036322.
- [8] Pan, J., Kang, Z., Zhang, W., Zhang, B., & Zhang, P. (2020). A Delay-dependent Glucose Control Approach for Type-1 Diabetes via Takagi-Sugeno Fuzzy Models. *2020 Chinese Control and Decision Conference (CCDC)*. doi:10.1109/ccdc49329.2020.9163827.
- [9] Buele, J., Ríos-Cando, P., Brito, G., Moreno-P., R., Salazar, F.W. (2020). Temperature Controller Using the Takagi-Sugeno-Kang Fuzzy Inference System for an Industrial Heat Treatment Furnace. *Computational Science and Its Applications – ICCSA 2020. ICCSA 2020, Lecture Notes in Computer Science*, 12254, Springer, Cham, Switzerland. doi:10.1007/978-3-030-58817-5\_27.
- [10] Bisong, E. (2019). *Building Machine Learning and Deep Learning Models on Google Cloud Platform*. Apress, Berkeley, United States. doi:10.1007/978-1-4842-4470-8.
- [11] Xie, H., Yang, F., Hua, M., Liu, S., Hu, J., & He, Y. (2021). Grounding grid corrosion detection based on mini-batch gradient descent and greedy method. *AIP Advances*, 11(6), 1–11. doi:10.1063/5.0051678.
- [12] Si, Z., Wen, S., & Dong, B. (2019). NOMA Codebook Optimization by Batch Gradient Descent. *IEEE Access*, 7, 117274–117281. doi:10.1109/ACCESS.2019.2936483.
- [13] Chen, Y., & Shi, C. (2019). Network Revenue Management with Online Inverse Batch Gradient Descent Method. *SSRN Electronic Journal*. doi:10.2139/ssrn.3331939.
- [14] Hanifa, A., Fauzan, S. A., Hikal, M., & Ashfiya, M. B. (2021). Comparison of LSTM and GRU (RNN) Methods for Classification of Fake News in Indonesian. *Engineering Dynamics*, 17(1), 33. doi:10.20884/1.dr.2021.17.1.436. (In Indonesian).
- [15] Hsueh, B.-Y., Li, W., & Wu, I.-C. (2019). Stochastic Gradient Descent with Hyperbolic-Tangent Decay on Classification. *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*. doi:10.1109/wacv.2019.00052.
- [16] Mignacco, F., Krzakala, F., Urbani, P., & Zdeborová, L. (2021). Dynamical mean-field theory for stochastic gradient descent in Gaussian mixture classification. *Journal of Statistical Mechanics: Theory and Experiment*, 2021(12), 124008. doi:10.1088/1742-5468/ac3a80.
- [17] Ruder, S. (2016). An overview of gradient descent optimization algorithms. *arXiv preprint. arXiv:1609.04747*. doi:10.48550/arXiv.1609.04747
- [18] Khirirat, S., Feyzmahdavian, H. R., & Johansson, M. (2017). Mini-batch gradient descent: Faster convergence under data sparsity. *2017 IEEE 56<sup>th</sup> Annual Conference on Decision and Control (CDC)*. doi:10.1109/cdc.2017.8264077.
- [19] Gou, P., & Yu, J. (2018). A nonlinear ANN equalizer with mini-batch gradient descent in 40Gbaud PAM-8 IM/DD system. *Optical Fiber Technology*, 46, 113–117. doi:10.1016/j.yofte.2018.09.015.
- [20] Messaoud, S., Bradai, A., & Moulay, E. (2020). Online GMM Clustering and Mini-Batch Gradient Descent Based Optimization for Industrial IoT 4.0. *IEEE Transactions on Industrial Informatics*, 16(2), 1427–1435. doi:10.1109/TII.2019.2945012.
- [21] Hu, J., Gao, Y., Li, J., & Shang, X. (2019). Deep Learning Enables Accurate Prediction of Interplay Between lncRNA and Disease. *Frontiers in Genetics*, 10. doi:10.3389/fgene.2019.00937.

- [22] Murugan, P., & Durairaj, S. (2017). Regularization and optimization strategies in deep convolutional neural network. arXiv preprint arXiv:1712.04711. doi:10.48550/arXiv.1712.04711.
- [23] Kukačka, J., Golkov, V., & Cremers, D. (2017). Regularization for deep learning: A taxonomy. arXiv:1710.10686. doi:10.48550/arXiv.1710.10686.
- [24] Gong, Y. (2016). Bernstein filter: A new solver for mean curvature regularized models. 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China. doi:10.1109/icassp.2016.7471967.
- [25] Davvaz, B., Mukhlash, I., & Soleha, S. (2021). Himpunan Fuzzy DAN Rough Sets. Limits: Journal of Mathematics and Its Applications, 18(1), 79. doi:10.12962/limits.v18i1.7705.
- [26] Wang, Q., Qian, Y., Liang, X., Guo, Q., & Liang, J. (2018). Local neighborhood rough set. Knowledge-Based Systems, 153, 53–64. doi:10.1016/j.knosys.2018.04.023.
- [27] Zhan, J., Liu, Q., & Herawan, T. (2017). A novel soft rough set: Soft rough hemirings and corresponding multicriteria group decision making. Applied Soft Computing Journal, 54, 393–402. doi:10.1016/j.asoc.2016.09.012.
- [28] Jothi, G., Inbarani, H. H., Azar, A. T., & Devi, K. R. (2019). Rough set theory with Jaya optimization for acute lymphoblastic leukemia classification. Neural Computing and Applications, 31(9), 5175–5194. doi:10.1007/s00521-018-3359-7.
- [29] Wang, B., Li, J., & Sun, W. (2023). CD - polytomous knowledge spaces and corresponding polytomous surmise systems. British Journal of Mathematical and Statistical Psychology, 76(1), 87-105. doi:10.1111/bmsp.12283.
- [30] Yao, Y., & Zhang, X. (2017). Class-specific attribute reducts in rough set theory. Information Sciences, 418–419, 601–618. doi:10.1016/j.ins.2017.08.038.
- [31] Zimmermann, H. J. (2011). Fuzzy set theory—and its applications. Springer Science & Business Media. doi:10.1007/978-94-010-0646-0.
- [32] Wiktorowicz, K., & Krzeszowski, T. (2020). Approximation of two-variable functions using high-order Takagi–Sugeno fuzzy systems, sparse regressions, and metaheuristic optimization. Soft Computing, 24(20), 15113–15127. doi:10.1007/s00500-020-05238-3.
- [33] Wu, D., Yuan, Y., Huang, J., & Tan, Y. (2020). Optimize TSK Fuzzy Systems for Regression Problems: Minibatch Gradient Descent with Regularization, DropRule, and AdaBound (MBGD-RDA). IEEE Transactions on Fuzzy Systems, 28(5), 1003–1015. doi:10.1109/TFUZZ.2019.2958559.
- [34] Rohmah, M. F., Ardiantoro, L., Putra, I. K. G. D., & Hartati, R. S. (2019). Forecasting the District Consumer Price Index in East Java with the Support Vector Regression Data Mining Method. Seminar Nasional Aplikasi Teknologi Informasi (SNATI), 3 August, 2019, Yogyakarta, Indonesia.