IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 13, No. 2, June 2024, pp. 2433~2440
ISSN: 2252-8938, DOI: 10.11591/ijai.v13.i2.pp2433-2440

Advancing machine learning for identifying cardiovascular disease via granular computing

Ku Muhammad Naim Ku Khalif[1,2], Noryanti Muhammad[1,2], Mohd Khairul Bazli Mohd Aziz[1,2], Mohammad Isa Irawan[3], Mohammad Iqbal[3], Muhammad Nanda Setiawan[3]

[1]Centre for Mathematical Sciences, Universiti Malaysia Pahang Al-Sultan Abdullah, Pahang, Malaysia
[2]Centre of Excellence for Artificial Intelligence and Data Science, Universiti Malaysia Pahang Al-Sultan Abdullah, Pahang, Malaysia
[3]Department of Mathematics, Faculty of Sciences and Data Analytics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

Article Info

Article history:

Received Jul 30, 2023
Revised Oct 2, 2023
Accepted Jan 6, 2024

Keywords:

Cardiovascular
Fuzzy numbers
Granular computing
Machine learning
Z-numbers

ABSTRACT

Machine learning in cardiovascular disease (CVD) has broad applications in healthcare, automatically identifying hidden patterns in vast data without human intervention. Early-stage cardiovascular illness can benefit from machine learning models in drug selection. The integration of granular computing, specifically z-numbers, with machine learning algorithms, is suggested for CVD identification. Granular computing enables handling unpredictable and imprecise situations, akin to human cognitive abilities. Machine learning algorithms such as Naïve Bayes, k-nearest neighbor, random forest, and gradient boosting are commonly used in constructing these models. Experimental findings indicate that incorporating granular computing into machine learning models enhances the ability to represent uncertainty and improves accuracy in CVD detection.

Corresponding Author:

Ku Muhammad Naim Ku Khalif
Centre for Mathematical Sciences, Universiti Malaysia Pahang Al-Sultan Abdullah
Pahang, Malaysia
Email: kunaim@umpsa.edu.my

1. INTRODUCTION

Cardiovascular disease (CVD), involving heart and blood vessel disorders often caused by fatty deposits in arteries, is a major cause of global disability. Preventable through healthy lifestyle choices, CVD is predicted by World Health Organization (WHO) to claim 17.9 million lives globally in 2019. Addressing risk factors like smoking, poor diet, obesity, and inactivity can prevent most cases. Early diagnosis is key for effective management and reducing fatalities, although challenges exist in analyzing symptoms in medical data due to duplication, multi-attribution, incompleteness, and time correlation. In addition, it is difficult to provide the correct drug therapy to a patient after manually reviewing vast amounts of cardiac disease data [1]. To solve this issue, machine learning techniques facilitate the creation of predictive models for accurately discerning the presence or absence of CVD in patients by processing vast, complex medical datasets. In machine learning, the software is programmed to do a specific task to learn from experience and anticipate the result of testing data based on training data [1]. Machine learning models, utilizing anonymous data, expedite accurate CVD predictions, enhancing healthcare efficiency and potentially saving lives. Their improved accuracy in classification tasks underscores their transformative impact in medical diagnostic [2]. Machine learning, encompassing predictive analytics and statistical learning, aims to unearth knowledge and patterns from data. A key focus is classification under supervised learning, which involves segregating objects into distinct, mutually exclusive classes based on predefined labels, ensuring each instance uniquely belongs to a single class [3]. Suppose the information gathered is used to describe a set of objects. Without loss of the generality or the nature of knowledge, assumptions are made where there is a unique attribute class taking class labels as its value [4].

Numerous experts have applied diverse machine learning techniques to predict CVD. Nawaz *et al.* [5] developed a gradient descent optimization (GDO)-based CVD prediction system, addressing challenges in heart disease detection. Weng *et al.* [6] compared traditional risk prediction methods with machine learning algorithms, utilizing electronic health data for accurate CVD case predictions. Cai *et al.* [7] reviewed existing CVD risk models, focusing on validation and comparison rather than creating new models. Hagan *et al.* [8] examined machine learning methods, including support vector machines (SVM), multi-layer perception (MLP) neural networks, and ensemble methods for CVD classification. Damen *et al.* [9] investigated bias in machine learning systems for CVD risk assessment, analyzing 117 studies using the preferred reporting items for systematic reviews and meta-analyses (PRISMA) model. It highlighted issues in traditional logistic regression analysis (LRA) and the limitations of well-known models like the Framingham heart risk scores. Granular computing, a human-centric problem-solving approach, is increasingly used in machine learning for complex applications. It involves breaking down wholes into parts and has become a significant paradigm for developing efficient models, especially in machine learning, to optimize information granule utilization.

Machine learning, data mining, and uncertainty reasoning with uncertainty will benefit from granular computing in the age of massive data. In the context of granular computing, granulation generally means decomposing a whole into several parts [10]. Perceptions are intrinsically imprecise, reflecting a fundamental limitation on the cognitive ability of humans to resolve detail and store information [11]. Granular computing-based machine learning is now reasonable and timely to construct broad theoretical models and practical methods. Granular computing concepts, which may be found in a plethora of representations and computational models, also need to be extracted. There are a lot of studies that have been done from the literature where rapid development in the utilisation of machine learning in granular computing contexts such as [10], [12]–[14]. In many fields of application in science and engineering, such as those that are often not crisp, phenomena such as uncertainty and fuzziness are well-known. Nevertheless, several grades of membership nearly appear on their own whenever a difficulty requires a resolution. When uncertainty or approximate reasoning has to be modelled, type-2 fuzzy sets are the proper tools. It has a greater power, which enables it to provide more latitude when depicting the unpredictability of human-based decision making issues. According to Klir *et al.* [15] interpretation, type-1 fuzzy sets are best used to convey imprecision rather than uncertainty. The motivation for further interest in type-2 fuzzy sets is that it provides a better scope for modelling uncertainty than type-1 fuzzy sets [16]. Karnik and Mendel [17] claim that type-2 fuzzy sets can be characterised as fuzzy membership functions value for type-2 fuzzy sets are in interval form [0,1], unlike type-1 fuzzy sets where the membership value is crisp in [0,1].

The z-numbers have also been utilised with granular computing for enhanced uncertainty handling [18]. In a world where most of the information on which judgments are based is unknown, there is a great likelihood that such decisions will be fraught with uncertainty [19]. In the body of work devoted to fuzzy sets, Zadeh [20] is credited for introducing the fuzzy set theory as a means of mathematically describing vagueness or imprecision. To accomplish this, the primary justification for using fuzzy sets is the capacity to deal effectively with fuzzy sets, instead of dealing inappropriately with approximate numerical amounts and the preferences subjectively held by decision-makers [21]. The involvement of higher-level uncertainty and reliability in z-numbers provides an additional degree of freedom to represent the uncertainty and fuzziness of real-world problems. The treatment of uncertainty and the consideration of reliability in analysing any computerised or mathematical model is essential to understanding possible ranges of scenario implications. The capability of quantifying the impact of uncertainty in the decision-making context is critical. In literature, there are two types of uncertainty: inter and intra-personal. This is also supported by Wallsten and Budescu [22] where there are supposedly two kinds of uncertainties related to linguistic characteristics: intra-personal and inter-personal. In particular, a lot of researchers and practitioners have applied z-numbers in diversified environments [23]–[27]. Due to implementing it, the way to handle z-numbers is different and much more unique and complex compared to type-1 and type-2 fuzzy sets.

This paper focuses on developing machine learning models for CVD detection within the framework of granular computing, specifically incorporating z-numbers in classification tasks. Attributes are treated as information granules, with their membership degrees aiding in accurate class determination, as supported in [3]. It independently assesses the membership degree of instances to each class in generative classification. The paper also compares theoretical and empirical results of these machine learning models. It is structured as follows: section 2 outlines the theoretical background of granular computing and z-numbers, followed by machine learning algorithms in section 3. Section 4 discusses the implementation and evaluation of these models in the CVD context, culminating in a conclusion in section 5.

2. MACHINE LEARNING MODELS VIA GRANULAR COMPUTING APPROACH

This section elaborates on the development of advanced machine learning systems within a granular computing framework, particularly for identifying CVD in binary classification problems. These systems utilize human indicators as input to diagnose illnesses. This study indicates that contemporary machine learning models primarily rely on procedural algorithms, diverging from traditional methods. Efforts are underway to create more sophisticated machine learning models using granular computing, especially employing z-numbers. These advanced models undergo several critical phases before generating results. The structure of these proposed machine learning models, integrated with granular computing, is illustrated in Figure 1.

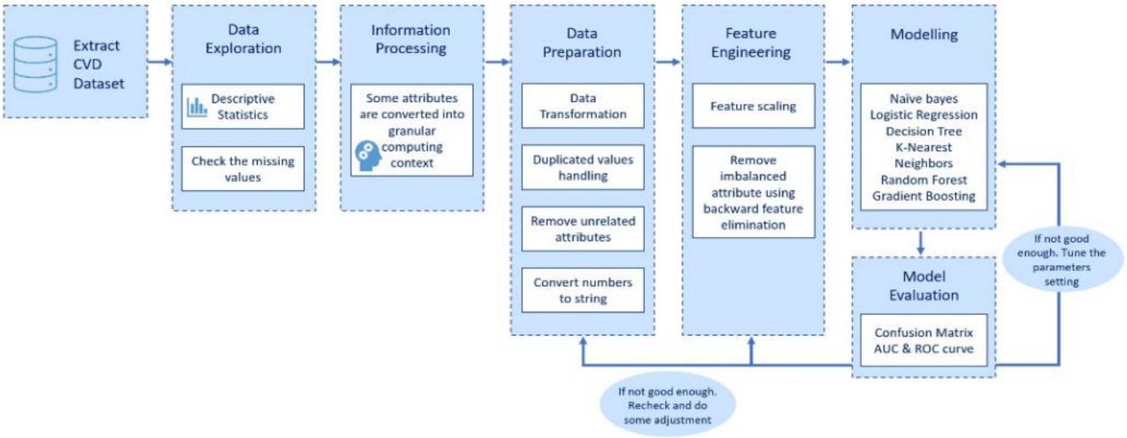


Figure 1. Machine learning architecture models development via granular computing

To enhance the accuracy of CVD detection, it is proposed to develop an advanced machine learning model using granular computing architecture. The Naive Bayes, k-nearest neighbor, random forest, and gradient boosting are applied to the same dataset to evaluate diverse algorithms. Each model is analyzed for improved accuracy, sensitivity, and specificity.

2.1. Phase 1: data extraction and exploration

This study utilized a dataset comprising 70,000 records with 11 characteristics, sourced from Kaggle. These data, documented as part of medical records, were collected during patient examinations. Table 1 details each characteristic of the dataset, along with the necessary statistical computations.

Table 1. CVD dataset attributes description with some statistical [1]

Serial number	Description of variable
1	calculation age-int (days); Min: 10798, Max: 23713, Mean: 19468.866, StdDev: 2467.252
2	Height-int (cm); Min: 55, Max: 250, Mean: 164.359, StdDev: 8.21
3	Weight-float (kg); Min: 10, Max: 200, Mean: 74.206, StdDev: 14.396
4	gender-categorical code; (f = female, m = male)
5	ap_hi-int; Min: -150, Max: 16020, Mean: 128.817, StdDev: 154.011
6	ap_lo-int; Min: -70, Max: 11000, Mean: 96.63, StdDev: 188.473
7	Cholesterol; (1 = normal, 2 = above normal, 3 = well above normal)
8	gluc; (1 = normal, 2 = above normal, 3 = well above normal)
9	Smoke-binary; (1 = smoker, 0 = non-smoker)
10	Alco-binary; (1 = yes, 0 = no)
11	Target- binary; (1 = Presence = 1, 0 = absence of cardiovascular disease)

2.2. Phase 2: information processing

2.2.1. Stage 1: linguistic terms are utilised for some related attributes

The authors briefly review some concepts of granular notion using z-numbers to represent the fuzzy nature of certain attributes. As previously mentioned, the primary motivation for employing fuzzy sets is their effectiveness in handling fuzzy concepts, and their ability to manage approximate numerical quantities and subjective expert preferences appropriately. The first component, $\tilde{A}$ is referred to as the restriction component, and it has the form of a real-valued uncertainty. The second component, $\tilde{B}$, is quantified in terms of reliability from $\tilde{A}$[20], [28].

$$\mu_{\tilde{A}}(x) = (a_1, a_2, a_3, a_4) = \begin{cases} \frac{(x-a_1)}{(a_2-a_1)} & if \quad a_1 \le x \le a_2 \\ 1, & a_2 \le x \le a_3 \\ \frac{(a_4-x)}{(a_4-a_3)} & if \quad a_3 \le x \le a_4 \\ 0 & otherwise \end{cases} \tag{1}$$

$$\mu_{\tilde{B}}(x) = (b_1, b_2, b_3, b_4) = \begin{cases} \frac{(x-b_1)}{(b_2-b_1)} & if \quad b_1 \le x \le b_2 \\ 1, & b_2 \le x \le b_3 \\ \frac{(b_4-x)}{(b_4-b_3)} & if \quad b_3 \le x \le b_4 \\ 0 & otherwise \end{cases}$$

2.2.2. Stage 2: convert the z-numbers into type-1 fuzzy numbers and aggregate them

Through a reduction method that makes use of an intuitive vectorial centroid, all of the z-numbers in the attributes are changed into type-1 fuzzy numbers. An extension of the traditional vectorial centroid techniques for fuzzy numbers, the intuitive vectorial centroid was suggested [29] and is an example of an extension. The approach to finding the centroid value is more intelligent, simple to calculate, produces more balanced, and takes into account all plausible scenarios of fuzzy numbers. This is in comparison to other centroid methods that can be found in the published research. For the dataset, the age-int attribute is presented by days, which the authors need to transform into years. Then, some measurement attributes are implemented using z-numbers such as age (after converted to year), height, weight, ap_hi, and ap_lo. The reduction process of z-numbers into type-1 fuzzy sets using intuitive vectorial centroid can be computed using [26], [29].

2.3. Phase 3: data preparation and feature extraction

This phase illustrates the process of implementing data preparation and feature extraction with some particular stages. It involves cleansing the dataset to ensure accuracy and consistency. Subsequently, the data undergoes normalisation or standardisation, which modifies the range of values to a uniform scale, facilitating the processing of algorithms. The feature extraction, which involves extracting pertinent information from the unprocessed data and converting it into a format that is better suited for machine learning models development.

2.3.1. Stage 1: preparing the data using extract, transform and load (ETL)

In this stage, the dataset is prepared and cleaned. There is no missing value and no replicated data. Some unrelated attributes are removed considering the rationale to compute for modelling purposes, such as id, age, height, weight, ap_hi, ap_lo, and age_yr. Some features, as mentioned, are converted into z-numbers in implementing granular computing. The original dataset presents the types of cardio and gender in numerical representation. To avoid the issue for the classification stage, these attributes are converted into string types.

2.3.2. Stage 2: feature engineering

The features in the data provided will directly influence the predictive models used and the results we can achieve. Some feature engineering techniques will be applied to fit with the data provided. Some attributes need rescale in terms of normalising the range of features in a dataset. They are computationally exclusive and often used in greedy search strategies; forward selection and backward elimination are fast and avoid overfitting to get the best-nested subset of features [30]. Scaling may be accomplished by the use of feature normalisation, also known as min-max normalisation, or feature scaling. To rescale a range between an arbitrary set of values [a, b], the (2) becomes:

$$x' = a + \frac{(x-min(x))(b-a)}{max(x)-min(x)}, a \text{ and } b \text{ are the min-max values} \tag{2}$$

To avoid scarcity issues in training the machine learning models, the backward elimination technique is applied by reducing some unfavourable or insignificant features involved.

2.4. Phase 4: modelling

Before we go for modelling, the data partitioning process is crucial to improve scalability, reduce contention, and optimise performance. It can also provide a mechanism for dividing data by usage patterns. This study’s dataset is split into 90% for training and 10% for testing. The authors utilise some machine learning algorithms, Naïve Bayes, k-nearest neighbor, random forest, and gradient boosting, for the modelling phase. Here, the authors considered two conditions of dataset environments which are: i) via granular computing and ii) without granular computing. The performance of machine learning models is measured by comparing simulations of the above mentioned two environments.

2.5. Phase 5: model evaluation

The evaluation of the model's performance utilizes a confusion matrix, generating four outcomes: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). Additionally, the receiver operating characteristic (ROC) curve assesses the robustness of the classification model at various thresholds, plotting the true positive rate (TPR) against the false positive rate (FPR), with TPR on the y-axis and FPR on the x-axis. The performance evaluation of the advanced machine learning models developed through granular computing is detailed in Tables 2 and 3, considering these metrics.

Table 2. Simulation results of proposed advanced machine learning models via granular computing (z-numbers) classifiers

Algorithms	Accuracy (%)	Sensitivity (%)	Specificity (%)
Z - Naïve Bayes	80.13	80	89
Z - K-nearest neighbor	91.29	91	89
Z - Random forest	92	92	90
Z - Gradient boosting	91	91	89

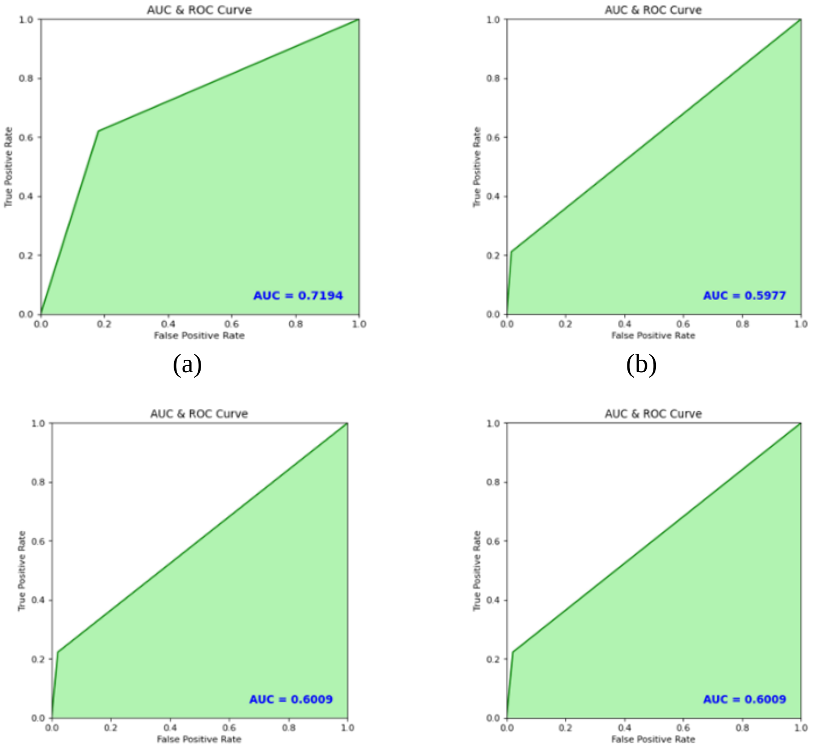
Table 3. Simulation results of conventional machine learning model classifiers

Algorithms	Accuracy (%)	Sensitivity (%)	Specificity (%)
Naïve Bayes	58.37	58	64
K-nearest neighbor	63.43	63	63
Random forest	70.69	71	71
Gradient boosting	72.13	72	73

When considering granular computing information, the classification results provide a comprehensive view of how well the classifier performs in each class. As a consequence of this, it is efficient in imbalanced classification, which enables us to effortlessly achieve a high degree of accuracy by only categorising every sample as belonging to the dominant class. However, the minority class has much worse accuracy, recall, and f1 score measures than the majority class. Regarding multiclass classification, precision, recall, and the f1 score are also applicable. We may consider the class that interests us to be a positive case, while all other courses might be considered negative cases. As depicted in Tables 2 and 3, the simulation results of proposed advanced machine learning models via granular computing classifiers show better accuracy, precision, recall, and f1 score metrics. This fusion from the proposed advanced machine learning models via granular computing has a better capability of handling uncertainty and vagueness arising from the lack of information during the computational process. This will avoid significant loss of information contained and can be implemented to solve data science problems.

During the process of tuning a binary classifier, the adjustment of various combinations of hyperparameters (such as the smoothing factor in our Naive Bayes classifier, for instance) would be ideal if there were a set of parameters in which the highest averaged and class individual f1 scores could be achieved simultaneously. However, in most situations, this is not the case. There are scenarios where two models have the same average f1 score, but one has a higher f1 score for one course and a lower score for another model has a higher f1 score for one course and a lower score for another course. There are also scenarios where a model has a significantly lower f1 score for a particular class despite having a higher overall average f1 score than another model. Because it is a consolidated measurement typically used in binary classification, the area under the curve (AUC) of the ROC is considered typically used in binary classification, the AUC of the ROC is something that is taken into consideration. The simulated ROC curve and AUC for advanced machine learning models that were generated using granular computing classifiers are shown in Table 3.

The area under the ROC curve is a graph that compares the TPR against the FPR at different probability levels ranging from 0 to 1. If the chance of a positive class is higher than the threshold for a testing sample, a positive class is assigned; otherwise, a negative class is provided. To summarise, the real positive rate is the same as recall, while the FPR is the fraction of negatives that are incorrectly recognised as positives. In contrast, the actual positive rate is the same as recall. A model that has a more prominent AUC is more likely to forecast the value 0 as 0 and the value 1 as 1. An AUC near one indicates that a classifier has a high degree of separability. Referring to Figures 2(a)-(d), the Naive Bayes model's AUC value is at its highest when the data set is split by the ratio of 72:28. A result of 71.94 for the AUC shows that the Naive Bayes model has a probability of 71.94 percent of successfully discriminating between absence and presence classes when applied to the prediction of CVD. The diagonal line that should be given as the default should reflect the AUC of the model that we have provided. A model with a more prominent AUC has a better probability of correctly forecasting that 0s will be 0s and 1s will be 1s. When the AUC is close to 1, the classifier has a high degree of separability.



(a) (b)

(c) (d)

Figure 2. Simulation ROC curve and AUC for proposed advanced machine learning models via granular computing classifiers: (a) Z - Naïve Bayes, (b) Z - Random Forest, (c) Z - K-Nearest Neighbor, and (d) Z - Gradient Boosting

3. CONCLUSION

This study presents an innovative approach to enhancing machine learning systems for CVD diagnosis using a granular computing environment. The proposed systems offer an efficient and economical computational method for handling imprecise, random, uncertain, and voluminous data, integrating modest formulas based on analytic and geometric principles. By merging machine learning algorithms with granular computing, a versatile computational technique addressing various information scenarios is developed. The study improvises multiple stages of computing machine learning algorithms with hyperparameter adjustments, ensuring thorough integration of both methods. CVD research, spurred by its severity and WHO data indicating 17.9 million deaths in 2017, has always faced challenges in prediction and detection accuracy. The findings here show significant performance improvements in CVD diagnosis, with the suggested system exhibiting higher accuracy, precision, sensitivity, and specificity than most existing models. This study also documents comparative investigations, noting these metrics. The research utilizes data on 11 clinical characteristics but future advancements could involve deep extreme machine learning with larger datasets, encompassing demographics and social life factors, to further enhance diagnosis. Adapting machine learning models and granular computing to address subjective aspects is crucial. Therefore, the proposed advanced machine learning models for CVD detection via granular computing aim to build a robust and reliable model yielding promising results and resource optimization. This model could extend its applicability to various fields involving human-centric data science challenges.

ACKNOWLEDGEMENTS

The authors would like to thank Universiti Malaysia Pahang Al-Sultan Abdullah for laboratory facilities as well as additional financial support under the UMP research grant, No. RDU230375 and Institut Teknologi Sepuluh Nopember grant 1297/PKS/ITS/2021.

REFERENCES

[1] M. N. Uddin and R. K. Halder, “An ensemble method based multilayer dynamic system to predict cardiovascular disease using machine learning approach,” *Informatics in Medicine Unlocked*, vol. 24, 2021, doi: 10.1016/j.imu.2021.100584.

[2] M. R. Ahmed, S. M. H. Mahmud, M. A. Hossin, H. Jahan, and S. R. H. Noori, “A cloud based four-tier architecture for early detection of heart disease with machine learning algorithms,” *2018 IEEE 4th International Conference on Computer and Communications, ICCC 2018*, pp. 1951–1955, 2018, doi: 10.1109/CompComm.2018.8781022.

[3] H. Liu and L. Zhang, “Fuzzy rule-based systems for recognition-intensive classification in granular computing context,” *Granular Computing*, vol. 3, no. 4, pp. 355–365, 2018, doi: 10.1007/s41066-018-0076-7.

[4] Y. Yao, “Perspectives of granular computing,” *2005 IEEE International Conference on Granular Computing*, vol. 2005, pp. 85–90, 2005, doi: 10.1109/GRC.2005.1547239.

[5] M. S. Nawaz, B. Shoaib, and M. A. Ashraf, “Intelligent cardiovascular disease prediction empowered with gradient descent optimization,” *Heliyon*, vol. 7, no. 5, pp. 1–8, 2021, doi: 10.1016/j.heliyon.2021.e06948.

[6] S. F. Weng, J. Reps, J. Kai, J. M. Garibaldi, and N. Qureshi, “Can machine-learning improve cardiovascular risk prediction using routine clinical data?,” *PLoS ONE*, vol. 12, no. 4, pp. 1–14, 2017, doi: 10.1371/journal.pone.0174944.

[7] R. Cai, X. Wu, C. Li, and J. Chao, “Prediction models for cardiovascular disease risk in the hypertensive population: a systematic review,” *Journal of Hypertension*, vol. 38, no. 9, pp. 1632–1639, 2020, doi: 10.1097/HJH.0000000000002442.

[8] R. Hagan, C. J. Gillan, and F. Mallett, “Comparison of machine learning methods for the classification of cardiovascular disease,” *Informatics in Medicine Unlocked*, vol. 24, pp. 1–10, 2021, doi: 10.1016/j.imu.2021.100606.

[9] J. A. A. G. Damen *et al.*, “Prediction models for cardiovascular disease risk in the general population: systematic review,” *BMJ*, vol. 353, pp. 1–11, 2016, doi: 10.1136/bmj.i2416.

[10] H. Liu, A. Gegov, and M. Cocea, “Rule-based systems: a granular computing perspective,” *Granular Computing*, vol. 1, no. 4, pp. 259–274, 2016, doi: 10.1007/s41066-016-0021-6.

[11] L. A. Zadeh, “Toward a perception-based theory of probabilistic reasoning with imprecise probabilities,” *Intelligent Systems for Information Processing: From Representation to Applications*, pp. 3–34, 2003, doi: 10.1016/B978-044451379-3/50001-7.

[12] Q. Hu, J. Mi, and D. Chen, “Granular computing-based machine learning in the era of big data,” *Information Sciences*, vol. 378, pp. 242–243, 2017, doi: 10.1016/j.ins.2016.10.048.

[13] H. Liu and L. Zhang, “Advancing ensemble learning performance through data transformation and classifiers fusion in granular computing context,” *Expert Systems with Applications*, vol. 131, pp. 20–29, 2019, doi: 10.1016/j.eswa.2019.04.051.

[14] H. C. Yan, Z. R. Wang, J. Y. Niu, and T. Xue, “Application of covering rough granular computing model in collaborative filtering recommendation algorithm optimization,” *Advanced Engineering Informatics*, vol. 51, 2022, doi: 10.1016/j.aei.2021.101485.

[15] G. J. Klir, U. St. Clair, and B. Yuan, *Fuzzy set theory: foundations and applications*, New Jersey, USA: Prentice Hall, 1997.

[16] C. Wagner and H. Hagras, “Uncertainty and type-2 fuzzy sets and systems,” *2010 UK Workshop on Computational Intelligence, UKCI 2010*, 2010, doi: 10.1109/UKCI.2010.5625603.

[17] N. N. Karnik and J. M. Mendel, “Centroid of a type-2 fuzzy set,” *Information Sciences*, vol. 132, no. 1–4, pp. 195–220, 2001, doi: 10.1016/S0020-0255(01)00069-X.

[18] S. Banerjee and T. Kumar Roy, “Arithmetic operations on generalized trapezoidal fuzzy number and its applications,” *An Official Journal of Turkish Fuzzy Systems Association*, vol. 3, no. 1, pp. 16–44, 2012.

[19] Z. Q. Xiao, “Application of z-numbers in multi-criteria decision making,” *ICCSS 2014 - Proceedings: 2014 International Conference on Informative and Cybernetics for Computational Social Systems*, pp. 91–95, 2014, doi: 10.1109/ICCSS.2014.6961822.

[20] L. A. Zadeh, “A note on z-numbers,” *Information Sciences*, vol. 181, no. 14, pp. 2923–2932, 2011, doi: 10.1016/j.ins.2011.02.022.

[21] H. Deng, “Comparing and ranking fuzzy numbers using ideal solutions,” *Applied Mathematical Modelling*, vol. 38, no. 5–6, pp. 1638–1646, 2014, doi: 10.1016/j.apm.2013.09.012.

[22] T. S. Wallsten and D. V. Budescu, “A review of human linguistic probability processing: General principles and empirical evidence,” *The Knowledge Engineering Review*, vol. 10, no. 1, pp. 43–62, 1995, doi: 10.1017/S0269888900007256.

[23] A. S. A. Bakar and A. Gegov, “Multi-layer decision methodology for ranking z-numbers,” *International Journal of Computational Intelligence Systems*, vol. 8, no. 2, pp. 395–406, 2015, doi: 10.1080/18756891.2015.1017371.

[24] R. A. Aliev, A. V. Alizadeh, and O. H. Huseynov, “The arithmetic of discrete Z-numbers,” *Information Sciences*, vol. 290, pp. 134–155, 2015, doi: 10.1016/j.ins.2014.08.024.

[25] R. A. Krohling, A. G. C. Pacheco, and G. A. D. Santos, “TODIM and TOPSIS with z-numbers,” *Frontiers of Information Technology and Electronic Engineering*, vol. 20, no. 2, pp. 283–291, 2019, doi: 10.1631/FITEE.1700434.

[26] K. M. N. K. Khalif, A. Gegov, and A. S. A. Bakar, “Hybrid fuzzy MCDM model for z-numbers using intuitive vectorial centroid,” *Journal of Intelligent and Fuzzy Systems*, vol. 33, no. 2, pp. 791–805, 2017, doi: 10.3233/JIFS-161973.

[27] A. M. Yaakob and A. Gegov, “Interactive TOPSIS based group decision making methodology using z-numbers,” *International Journal of Computational Intelligence Systems*, vol. 9, no. 2, pp. 311–324, 2016, doi: 10.1080/18756891.2016.1150003.

[28] B. Kang, D. Wei, Y. Li, and Y. Deng, “Decision making using z-numbers under uncertain environment,” *Journal of Computational Information Systems*, vol. 8, no. 7, pp. 2807–2814, 2012.

[29] K. M. N. K. Khalif, “Generalised hybrid fuzzy multi criteria decision making based on intuitive multiple centroid,” Ph.D. Theses, School of Computing, University of Portsmouth, Portsmouth, United Kingdom, 2016.

[30] M. B. de Moraes and A. L. S. Gradvohl, “A comparative study of feature selection methods for binary text streams classification,” *Evolving Systems*, vol. 12, no. 4, pp. 997–1013, 2021, doi: 10.1007/s12530-020-09357-y.

BIOGRAPHIES OF AUTHORS

Dr. Ku Muhammad Naim Ku Khalif is a senior lecturer at the Centre for Mathematical Sciences, Universiti Malaysia Pahang, Malaysia. He holds a Ph.D. in Computational Intelligence from the University of Portsmouth, United Kingdom. His research focuses on developing computational intelligence methods for decision-making support systems and modeling complex systems under fuzzy/uncertain environments. He is also interested in machine learning/deep learning with a focus on probabilistic models and their applications. Recently, he has been working on deploying machine learning/deep learning technology in data science. Additionally, he has supervised Ph.D. students, participated in various research projects, and served as a reviewer for several journals and conferences, including Atlantis Press, IOS Press, IEEE Transactions on Fuzzy Systems, PLOS ONE, FSDM, SCML, ICMSCT2019, ICoAIMS, SKSM, and IACE. He can be contacted at email: kunaim@umpsa.edu.my.

Dr. Noryanti Muhammad is a senior lecturer of Mathematics (specialisation on statistics). In research, she is interested in developing statistical methodologies for a variety of real applications, including medical applications, finance, and reliability. Besides, she was involved in analysing data using a variety of statistical tools (now popular with data analytics). Her main research is about nonparametric predictive inference (NPI), focusing on NPI for bivariate data with considering dependence structure using copula. At the same time, she did research in the modelling of cardiovascular diseases (CVD) and development of predictive heart risk scores; and developed a parametric model for grouped and ungrouped line transect data. She can be contacted at email: noryanti@umpsa.edu.my.

Dr. Mohd Khairul Bazli Mohd Aziz focusing on applied statistics and operational research. His main past research contributions are designing rain gauge network systems using geostatistics and simulated annealing hybrid with particle swarm optimization as the optimization method. This research uses rainfall, humidity, elevation, solar radiation, wind speed and temperature data to determine the optimal number and locations for the rain gauge network based on the estimated variance calculated. His past research also involves stochastic modelling on estimating the clostridium acetobutylicum solvent production acetone, butanol, and ethanol during the fermentation process. He can be contacted at email: khairulbazli@umpsa.edu.my.

Prof. Dr. Mohammad Isa Irawan is a professor at Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia. He holds a Ph.D. in Software Engineering & Interactive System Group, from the Department of Informatics, Vienna University of Technology, Austria. His research focuses on developing mathematical modelling systems in natural science and mathematical science. Recently, he has been working on deploying machine learning technology in data science problems. He has conducted research by developing an intelligent irrigation water requirement system based on artificial neural networks and profit optimization for planning time decision making of crops in Lombok Island, Indonesia. He can be contacted at email: mii@its.ac.id.

Mohammad Iqbal is an asst. professor in the Deptartment of Mathematics, Institut Teknologi Sepuluh Nopember (ITS), Indonesia. He received a Ph.D. degree in Machine Learning and Data Analytics from the Deptartment of Computer Science and Information Engineering, National Taiwan University of Science and Technology (NTUST), Taiwan. He joined as artificial intelligence engineer in KaiKutek Corp. Taiwan for 2021 – 2022. In 2023, he was invited as a visiting Asst. Prof. to teach summer course on natural language processing. His interest in research includes data science, machine learning, data analytics, data mining, and deep learning. He can be contacted at email: iqbal@its.ac.id

Muhammad Nanda Setiawan is a data scientist at a private company in Indonesia. He holds a Bachelor's degree in Mathematics with a specialization in Computer Science from Institut Teknologi Sepuluh Nopember (ITS). Previously, he was a member of the Machine Learning and Big Data Group at Mathematics ITS. His research interests lie in the application of machine learning and data analysis techniques. He can be contacted at email: m.nanda98@hotmail.com