Development of rough-TOPSIS algorithm as hybrid MCDM and its implementation to predict diabetes

Shampa Sengupta

Department of Information Technology,
MCKV Institute of Engineering,
Liluah, Howrah-711 204, West Bengal, India
Email: shampa2512@gmail.com

Debabrata Datta*

Department of Information Technology,
Heritage Institute of Technology Kolkata,
West Bengal, India
Email: debabrata.datta@heritageit.edu
*Corresponding author

S. Suman Rajest

Department of Research and Development (R&D) and International Student Affairs (ISA),
Dhaanish Ahmed College of Engineering,
Chennai-601301, Tamil Nadu, India
Email: sumanrajest414@gmail.com

P. Paramasivan

Department of Research and Development,
Dhaanish Ahmed College of Engineering,
Chennai-601301, Tamil Nadu, India
Email: paramasivanchem@gmail.com

T. Shynu

Department of Biomedical Engineering,
Agni College of Technology,
Chennai, Tamil Nadu, India
Email: shynu469@gmail.com

R. Regin

Department of Computer Science and Engineering,
SRM Institute of Science and Technology,
Ramapuram, Chennai-89, Tamil Nadu, India
Email: regin12006@yahoo.co.in

Abstract: In this work, an innovative approach of multi-criteria decision-making method guided by rough set theory is researched to predict diabetes. Diabetes is the root cause of various deadly diseases. Designing an expert diabetes prediction model can solve the health monitoring issue with preventive measures beforehand. The proposed work has mainly two phases. In the first phase, the ensemble classification method develops the classification model, and rough set theory is implemented as a feature selection technique. In the second phase, TOPSIS, a multi-criteria decision-making method, is implemented for optimising classification models. Ensemble classification methods used here in this work: Bagging, AdaBoost, M1, Logit Boost, attributed selected classifier, random subspace, and multi-class classifier. The technique for order preference by similarity to ideal solution, the so-called TOPSIS, a multi-criteria decision-making method, has been used to select the optimised prediction model. Experimental diabetes data are collected from the UCI repository. Results obtained for predicting diabetes agree with those obtained from clinical practitioners.

Keywords: rough set theory; RST; ensemble classification; diabetes prediction; TOPSIS; algorithm as hybrid; Adaboost; logit boost; multiclass classifier.

Reference to this paper should be made as follows: Sengupta, S., Datta, D., Rajest, S.S., Paramasivan, P., Shynu, T. and Regin, R. (2023) 'Development of rough-TOPSIS algorithm as hybrid MCDM and its implementation to predict diabetes', *Int. J. Bioinformatics Research and Applications*, Vol. 19, No. 4, pp.252–279.

Biographical notes: Shampa Sengupta is an Associate Professor at MCKV Institute of Engineering, India. She completed her PhD (CST) from IIEST, Shibpur. She received her BTech (Hons.) degree in Instrumentation from the Haldia Institute of Technology and her MTech degree in Information Technology from Bengal Engineering and Science University. Her research interests include data mining, pattern recognition, and big data analytics. She has published several book chapters, conference papers, and journal papers in

P. Paramasivan has been working as a Director at Dhaanish Ahmed College of Engineering, Chennai, India. He is an accomplished professor with extensive experience, and he has dedicated his career to mentoring and educating students with a passion for engineering and technology. His dedication to advancing knowledge is evident through the publication of ten utility patents and one design patent. He has received research funding from various sources, including government bodies, industries, and non-governmental organisations. Government entities such as MSME, PMKVY, DST, and AICTE have recognised his research's value and supported his endeavours. Additionally, he has secured funding from industry players like Haarish Info Tech and Techno, further validating the impact and relevance of his work.

T. Shynu is a young researcher in the Master of Engineering, Department of Biomedical Engineering, Agni College of Technology, Chennai, Tamil Nadu, India. She has been an instrumental developer and researcher for three years and has published in 17 international journals. She is interested in the block diagram of IoT-based, automated oxygen pumping systems, CNN, biomedical, instrumental, etc.

R. Regin is an Assistant Professor at the Department of Computer Science and Engineering at SRM Institute of Science and Technology in Ramapuram, Chennai, India. He has been awarded PhD in Computer Science and Engineering at Anna University. He has been a Software Developer and researcher for ten years and has published 40 international journals, 20 international conferences, 15 national conferences, and ten Springer book chapters. He is on Ilmiah Teunuleh's advisory board and editor of the *International Journal of Technology Information and Computer*, and *Growing Scholar USA*. He edits the *ICT-based Framework for Data Science and Machine Learning*. He explores VANET, cloud computing, and infosec. He reviews for Springer and others.

1 Introduction

Diabetes (2021) occurs due to the increased level of blood glucose. Glucose is obtained from food, and that is the source of energy. Insulin, a hormone produced by the pancreas, helps extract glucose from food intake and get into our cells. Sometimes our body does not produce or utilise the insulin properly, which causes glucose to stay in the bloodstream, and people become diabetic patients (Zannah et al., 2023). Generally, three different types of diabetes are seen, and those are called type 1, type 2, and gestational diabetes (Priscila et al., 2023). In Type 1 diabetes, insulin is not produced due to the bad immune system of the body, so needed insulin injection and the case is known as insulin-dependent diabetes mellitus (IDDM). Type 2 diabetes grows due to improper utilisation of insulin and is known as nom-insulin-dependent diabetes mellitus (NIDDM). Gestational diabetes is seen during pregnancy due to hormonal changes (Al-Khasawneh et al., 2022).

The proposed diabetes prediction model for the prediction of diabetes disease in the map-reduce framework (Shafqat et al., 2018) is provided in Figure 1. During the map step of the process, the entire training dataset is split up into many decision subsystems, each of which is then allocated to a specific slave node (Nirmala et al., 2023; Al-Khasawneh

et al., 2018). Each slave node independently executes a feature subset selection algorithm based on rough set theory (RST) on the corresponding decision data subset and generates classification rules using a base classifier (such as Bagging, AdaboostM1, LogitBoost, attributed selected classifier (Bishop, 2010). Random subspace, and multi-class classifier) (Siva and Murugan, 2005). Then, in the reduction phase, the final set of classification rules for a given classifier across the entire dataset is generated by combining the rules acquired from the various slave nodes using a majority voting mechanism (Rani et al., 2021). Each of the investigated base classifiers then receives its separate prediction model (Pandit, 2023; Jeba et al., 2023). In the second phase, the optimal prediction model is selected based on seven classification parameters such as TP Rate, FP Rate, Accuracy, Precision, Recall, F1-Measure, and ROC Area (Bishop, 2010). Using the TOPSIS method (Aruldoss et al., 2013; Alarood et al., 2022). Figure 1 represents the steps of developing the classification model in the map-reduce framework.

Figure 1 Proposed diabetes prediction model development