

Brought to you by [INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA](#)



Scopus



[Back](#)

Performance Metrics Analysis for Diabetes Prediction Using Machine Learning Algorithm

[ICETAS 2024 - 9th IEEE International Conference on Engineering Technologies and Applied Sciences](#) • Conference Paper • 2024 •

DOI: [10.1109/ICETAS62372.2024.11120010](#)

[Said, Saida O.](#) ; [Binti Mohamad Zulkufli, Nurul Liyana](#) ; [Ahmad Puzi, Asmarani Binti](#) ; [Shah, Asadullah](#)

International Islamic University Malaysia (IIUM), Kulliyah of Information & Communication Technology, Kuala Lumpur, 53100, Malaysia

[Show all information](#)

0
Citations

[View PDF](#) [Full text](#) [Export](#) [Save to list](#)

[Document](#) [Impact](#) [Cited by \(0\)](#) [References \(14\)](#) [Similar documents](#)

Abstract

Diabetes is the among of the chronic diseases which is every year the number of deaths increases to the diabetes patients. Despite the different ages, diabetes can either be inherited genetically from the family or the person can get due to the imbalance of balance diet. Therefore, to reduce the speed of being affected inside the body, prediction techniques such as Machine learning should be used. The main objective of this study is to predict diabetes using two different datasets with different features using Machine learning algorithm. The first dataset had a large number of instances and was divided into the different age groups that include Pediatrics, Early Adulthood, Middle age and Geriatric. While the second dataset was coming through Sylhet Diabetes Hospital in Bangladesh with the small number of instances. To see the behaviour of these dataset, each dataset used three machine learning algorithm which were Naïve Bayes, Random Forest and Decision tree (J48). In the experiments various performance metrics were used including accuracy, precision, recall, F1 score, ROC, Kappa statistics, Mean Absolute Error (MAE), Root Mean Squares Error (RMSE) and Relative Absolute Error (RAE). The experimental results show that in the accuracy that the larger datasets, Naïve Bayes had higher accuracy of 99.78% and 98.78% on the age group of Pediatrics and Early Adulthood respectively, while in the age group of Middle age and Geriatric, Decision Tree had high accuracy of 95.55% and 93.25%, but in term of other performance metrics Naïve byes performed well. As for the second dataset the Random Forest had the highest accuracy of 97.50% compared to other algorithms. © 2024 IEEE.

Author keywords

Decision tree (J48); Diabetes; Machine learning algorithms; Naïve Bayes; Random Forest

Indexed keywords

Engineering controlled terms

Barium compounds; Decision trees; Diseases; Error statistics; Geriatrics; Hospitals; Information use; Large datasets; Learning systems; Machine learning; Mean square error; Pediatrics; Random forests

Engineering uncontrolled terms

Age groups; Chronic disease; Decision tree (j48); Different ages; High-accuracy; Machine learning algorithms; Middle ages; Naive bayes; Performance metrics; Random forests

Engineering main heading

Medical problems

Funding details

Details about financial support for research, including funding sources and grant numbers as provided in academic publications.

Funding sponsor	Funding number	Acronym
Kulliyyah of Information and Communication Technology		
International Islamic University Malaysia See opportunities by IIUM ↗	IRAGS18-030-0031	IIUM
International Islamic University Malaysia See opportunities by IIUM ↗		IIUM

Funding text

The authors would like to thank Kulliyyah of Information and Communication Technology (KICT), IIUM for their support, especially for the publication fee. The second author would like to acknowledge support by IRAGS18-030-0031 for research activities.

© Copyright 2025 Elsevier B.V., All rights reserved.

Abstract

Author keywords