|  |  |  |  |
| --- | --- | --- | --- |
| Title | A stacked ensemble machine learning approach for the prediction of diabetes | | |
| Author(s) Name | Khondokar Oliullah, Mahedi Hasan Rasel, Md. Manzurul Islam, Md. Reazul Islam, Md. Anwar Hussen Wadud & Md. Whaiduzzaman | | |
| Contact Email(s) | reazul@aiub.edu | | |
| Published Journal Name | Journal of Diabetes & Metabolic Disorders | | |
| Type of Publication | Journal | | |
| Volume | 22 | Issue | 2 |
| Publisher | Springer International Publishing | | |
| Publication Date | 22 November 2023 | | |
| ISSN | 2251-6581 | | |
| DOI | 10.1007/s40200-023-01321-2 | | |
| URL | https://link.springer.com/article/10.1007/s40200-023-01321-2 | | |
| Other Related Info. |  | | |
|  | | | |

|  |  |
| --- | --- |
| Abstract |  |
| Diabetes has become a leading cause of mortality in both developed and developing countries, impacting a growing number of individuals worldwide. As the prevalence of the disease continues to rise, researchers have diligently worked towards developing accurate diabetes prediction models. The primary aim of this study is to utilize a diverse set of machine learning algorithms to detect the presence of diabetes, particularly in females, at an early stage. By leveraging these methods, this research seeks to provide physicians with valuable tools to identify the disease early, enabling timely interventions and improving patient outcomes. In this study, some state-of-the-art machine learning techniques, such as random forest classifiers with gridsearchCV, XGBoost, NGBoost, Bagging, LightGBM, and AdaBoost classifiers, were employed. These models were chosen as the base layer of our proposed stacked ensemble model because of their high accuracy. Before feeding the data into the models, the dataset was preprocessed to ensure optimal performance and obtain improved results. The accuracy achieved in this study was 92.91%, which demonstrates its competitiveness with the existing approaches. Moreover, the utilization of the Shapley additive explanation (SHAP) facilitated the interpretation of machine learning models. We anticipate that these findings will be beneficial to healthcare providers, stakeholders, students, and researchers involved in diabetes prediction research and development. | |