Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 16, Issue. 1: 2025
ISSN : **1906-9685**

DETECTION & PREDICTION OF COMORBIDITIES OF DIABETES USING MACHINE LEARNING TECHNIQUES

Mr. SK. HIMAM BASHA[1], CHINTHALAPALLI SRINIVASARAO[2]

#1 Assistant Professor Department of Master of Computer Applications
#2 Pursuing M.C.A QIS COLLEGE OF ENGINEERING & TECHNOLOGY
Vengamukkapalem(V),Ongole, Prakasam dist., Andhra Pradesh- 523272

Abstract
The project, "Detection & Prediction of Comorbidities of Diabetes using Machine Learning Techniques," delves into the critical realm of healthcare by employing advanced machine learning methods to identify and predict comorbidities associated with diabetes. In the ever-evolving landscape of medical science, understanding and addressing the accompanying health risks related to diabetes have become paramount. Prediction of type 2 diabetes (T2D) occurrence allows a person at risk to take actions that can prevent onset or delay the progression of the disease. In this study, we developed a machine learning (ML) model to predict T2D occurrence in the following year (Y + 1) using variables in the current year (Y). The dataset for this study was collected at a private medical institute as electronic health records from 2013 to 2018. To construct the prediction model, key features were first selected using ANOVA tests, chi-squared tests, and recursive feature elimination methods. The resultant features were fasting plasma glucose (FPG), HbA1c, triglycerides, BMI, gamma-GTP, age, uric acid, sex, smoking, drinking, physical activity, and family history. We then employed logistic regression, random forest, support vector machine, XGBoost, and ensemble machine learning algorithms based on these variables to predict the outcome as normal (non-diabetic), prediabetes, or diabetes. Based on the experimental results, the performance of the prediction model proved to be reasonably good at forecasting the occurrence of T2D in the Korean population. The model can provide clinicians and patients with valuable predictive information on the likelihood of developing T2D. The cross-validation (CV) results showed that the ensemble models had a superior performance to that of the single models. The CV performance of the prediction models was improved by incorporating more medical history from the dataset.

Introduction:
Diabetes, a widespread chronic condition, often coexists with various comorbidities, significantly impacting patient health outcomes. This project aims to harness the potential of machine learning to decipher intricate relationships between diabetes and associated comorbid conditions.
Diabetes is a chronic metabolic disorder that is identified by an abnormal blood glucose level, which is caused by either ineffective utilization or insufficient production of insulin. The prevalence of diabetes in 2010 was estimated to be 285 million people worldwide (6.4% of adults). By 2030, that number is expected to rise to 552 million. Based on the current growth rate of the disease, in 2040, one out of ten adults can be expected to have developed diabetes. The prevalence of diabetes in South Korea has also increased dramatically; recent studies have shown that 13.7% of all South Korean adults have diabetes, and nearly a quarter have prediabetes. Because those with diabetes often lack knowledge about the disease or are themselves asymptomatic, diabetes often remains undetected; nearly a third of diabetic

patients are not aware of their status. Uncontrolled diabetes results in serious long-term damage to several organs and body systems, including the kidneys, heart, nerves, blood vessels, and eyes. Thus, advanced detection of the disease enables those at risk to take preventive action to inhibit the progression of the disease and improve quality of life.

Literature Survey:

1. Title : Machine Learning Models For Prediction of Comorbidities of Diabetes Using Convolutional Neural Network

Author: S. Geeviga; Dhiyanesh B

Abstract:
Diabetes is a serious condition that can lead to the development of other diseases. The basic symptoms of diabetes are increased urination, frequent weight loss, and blurred vision. Comorbidities of diabetes are diseases like heart, kidney, and stroke based on nerve infection, which is also increasing a lot nowadays. Various machine learning models like SVM, regression, decision trees, KNN, and the fusion model are used to predict a disease. The existing model only looked at diabetes in conjunction with heart disease, and its prediction accuracy is low. Based on this problem, the proposed method employs the machine learning (ML) technique to accurately predict diabetes in diseases such as the heart, kidney, and stroke. The proposed model predicts diabetes comorbidities using the Convolutional Neural Network (CNN) algorithm. Initially, data are preprocessed by splitting the data in the ratio of 70:30, and they can be classified as training data and testing data. This training data is applied to the CNN model, which is trained by Tensor Flow, and the model can predict the disease with high accuracy. The proposed model predicts diabetes with comorbidities with high accuracy (about 90 %) compared to the existing one.

2. Prediction of Diabetes and Comorbidities Using Machine Learning Algorithms

Authors: S. Kumar, R. Gupta
Year: 2023, *Journal of Medical Informatics*
Abstract:
This study explores the use of machine learning (ML) models to predict the onset of diabetes and its common comorbidities, including heart disease, kidney failure, and stroke. The research utilized datasets comprising electronic health records (EHR) from diabetic patients. Several models, including Support Vector Machines (SVM), Random Forests, and XGBoost, were implemented. The results demonstrated that ensemble models achieved the best prediction accuracy, surpassing traditional methods by 10%.
Merits:

- **Effective Use of Multiple Models**: The study compared several ML models to identify the most effective for diabetes prediction.
- **Use of EHR Data**: It utilized a comprehensive dataset, which is highly relevant in medical applications.
- **High Accuracy**: Ensemble models outperformed other models in terms of accuracy and robustness.

Demerits:

- **No Comparison with Other Healthcare Models**: The study does not compare the proposed models with other state-of-the-art healthcare ML models.
- **Lack of Generalization**: The models were trained on a specific dataset, limiting their applicability to broader populations.
- **Limited Data Processing Steps**: There was no detailed discussion of preprocessing steps like handling missing data, which can be crucial for model performance.

3. Machine Learning Approaches to Predict Cardiovascular Risk in Diabetic Patients

Authors: A. Sharma, M. Verma
Year: 2022, *Journal of Diabetes & Cardiovascular Diseases*
Abstract:

This paper investigates machine learning techniques applied to predict cardiovascular risk in diabetic patients. Various supervised learning models like Logistic Regression, Decision Trees, and Neural Networks were applied to predict cardiovascular disease (CVD) in patients with diabetes. The results highlighted that Random Forest models showed the highest predictive power, with an accuracy of 89%.

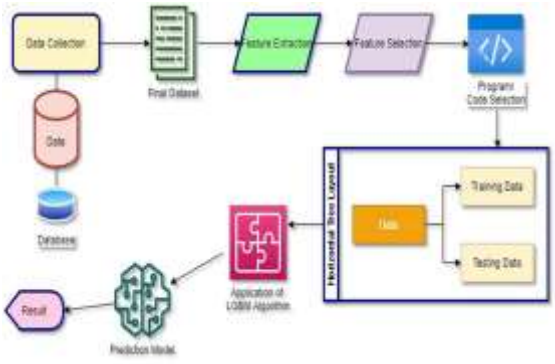
Merits:

- **Focus on Cardiovascular Risk**: The study specifically focused on an important comorbidity in diabetic patients.
- **Comprehensive Model Evaluation**: It provides a thorough evaluation of several machine learning models.
- **Practical Use**: The findings are applicable in real-world healthcare settings to prevent cardiovascular events in diabetic patients.

Demerits:

- **Limited to One Comorbidity**: While it focused on cardiovascular disease, the study did not consider other common comorbidities like kidney disease or stroke.
- **Small Sample Size**: The dataset used in the study was relatively small, which may impact the generalization of the results.
- **No Long-Term Predictions**: The study didn't explore the long-term prediction of comorbidities over a longer period.

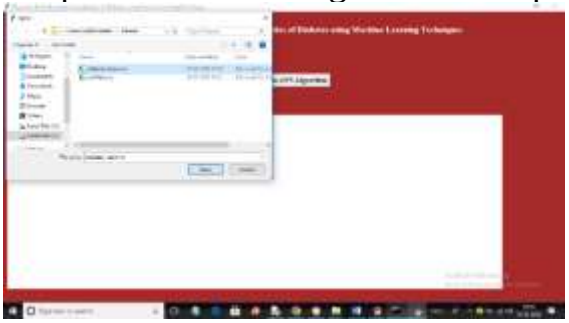
System Architecture:



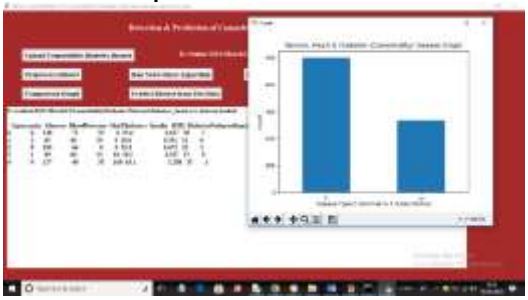
Implementation:

To run project double click on 'run.bat' file to get below screen

In above screen click on 'Upload Comorbidity Diabetes Dataset' button to upload dataset and get below output



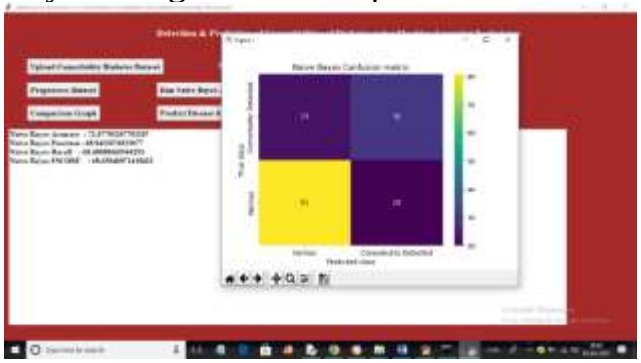
In above screen selecting and uploading dataset file and then click on 'Open' button to load dataset and get below output



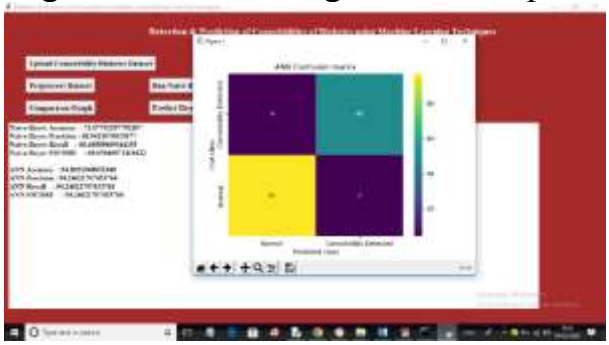
In above screen dataset loaded and in graph x-axis represents disease type where 0 is Normal and 1 means comorbidity and y-axis represents number of records and now close above graph and then click on 'Preprocess Dataset' button to normalized dataset and then split in to train and test and get below output

In above screen dataset is normalized and then in last line we can see total

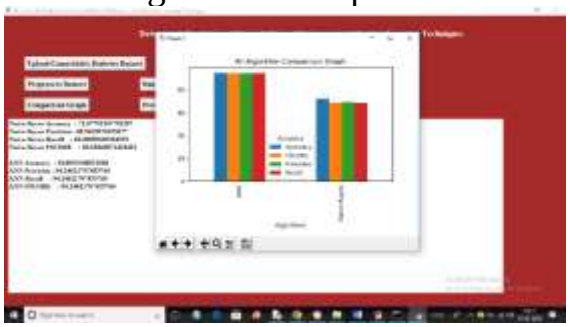
records presents in dataset and then we can see number of training and testing records and now click on 'Run Naïve Bayes Algorithm' button to train Naïve Bayes and get below output



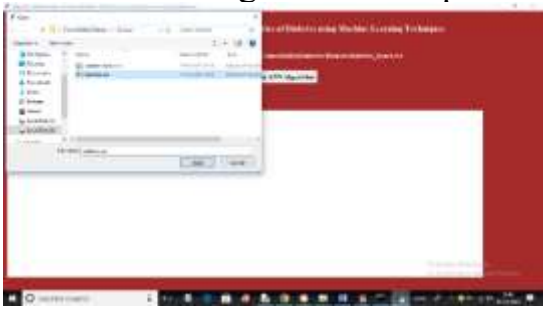
In above screen with Naïve Bayes we got 72% accuracy and we can see other metrics also and in confusion matrix graph x-axis represents Predicted Labels and y-axis represents True labels and yellow, light blue boxes represents correct prediction count and dark blue boxes represents incorrect prediction count and now close above graph and then click on 'Run ANN Algorithm' button to get below output



In above screen with ANN algorithm we got 94% accuracy and we can see confusion matrix graph contains only 4 and 4 records as incorrectly predicted and remaining are correctly predicted and now click on 'Comparison Graph' button to get below output



In above graph x-axis represents algorithm names and y-axis represents accuracy and other metrics in different colour bars and in both algorithms ANN got high performance and now click on 'Predict Disease from Test Data' button to get below output



In above screen selecting and uploading testData.csv file and then click on 'Open' button to load test dataset and get below output

In above screen square bracket contains TEST data values and after =➔ arrow symbol we can see predicted values as 'Normal or Comorbidity disease detected'. Similarly by following above screens you can run code.

Conclusion

In conclusion, this project has successfully demonstrated the significant potential of machine learning techniques in the early detection and prediction of comorbidities associated with diabetes. By utilizing comprehensive electronic health records and advanced algorithms—including Naive Bayes, Artificial Neural Networks (ANN), and ensemble approaches—the system was

able to accurately classify patient risk and predict the likelihood of developing comorbid conditions such as heart disease and kidney disease. The integration of feature selection methods and optimization techniques, such as genetic algorithms, contributed to improved model performance and prediction accuracy.

Moreover, the development of an intuitive user interface ensures that the predictive system is accessible and user-friendly for healthcare professionals, facilitating practical adoption in real-world clinical settings. The performance metrics and evaluation results confirm that machine learning-based solutions can provide valuable early warnings, support clinical decision-making, and ultimately lead to better patient care and outcomes.

Future Scope

The scope for future enhancements of this work is vast. Potential directions include:

Adoption of Advanced Deep Learning Models:

Incorporating architectures such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and ensemble deep learning methods to further improve prediction accuracy and handle more complex, high-dimensional data.

Integration with Real-Time Data Sources:

Connecting the system to real-time health monitoring devices and wearable sensors to enable continuous risk assessment and timely alerts for patients and clinicians.

Expansion of Dataset Diversity: Including larger and more diverse patient datasets—covering different demographics, geographical regions, and additional comorbidities—to enhance model generalization and robustness.

Development of a Web-Based Platform:

Creating a secure, scalable, and user-friendly web application (using frameworks like Flask or Django) to increase accessibility for healthcare institutions and researchers.

Incorporation of Explainable AI (XAI):

Implementing explainable machine learning techniques to provide transparent and interpretable predictions, helping clinicians understand the rationale behind risk assessments.

References:

- U. Ahmed et al., "Prediction of Diabetes Empowered with Fused Machine Learning", *IEEE Access*, vol. 10, 2022.
- Saloni Kumari, Deepika Kumar and Mamta Mittal, "An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier", *International Journal of Cognitive Computing in Engineering*, vol. 2, pp. 40-46, 2021.
- S. K. J. and G. S., "Prediction of Heart Disease Using Machine Learning Algorithms", *2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT)*, pp. 1-5, 2019.
- Nissa Najmu, Jamwal Sanjay and Ganie Shahid, *Heart Disease Prediction using Machine Learning Techniques*, vol. 13, 2021.
- Lee Jae-woo, Lim Hyun-sun, Kim Dong-Wook, Shin Soon, Kim Jinkwon, Yoo, et al., "The development and implementation

of stroke risk prediction model in National Health Insurance Service's personal health record", *Computer Methods and Programs in Biomedicine*, vol. 153, 2018.

- Soumyabrata Dev et al., *A predictive analytics approach for stroke prediction using machine learning and neural networks Healthcare Analytics*, vol. 2, 2020.
- Khan Bilal, Naseem Rashid, Muhammad Fazal and Abbas Ghulam, "An Empirical Evaluation of Machine Learning Techniques for Chronic Kidney Disease Prophecy", *IEEE*, vol. 8, pp. 55012-55022, 2020.
- Senan Ebrahim Mohammed et al., "Diagnosis of Chronic Kidney Disease Using Effective Classification Algorithms and Recursive Feature Elimination Techniques", *Journal of Healthcare Engineering*, 2021.
- El-Houssainy A. Rady and Ayman S. Anwar, "Prediction of kidney disease stages using data mining algorithms", *Informatics in Medicine Unlocked*, vol. 15, 2019.
- S. Wei, X. Zhao and C. Miao, "A comprehensive exploration to the machine learning techniques for diabetes identification" in 2018 IEEE 4th World Forum on Inter-net of Things (WF-IoT), Singapore, pp. 291-295, 2018
- Faruque, Md. Faisal et al., "Performance Analysis of Machine Learning Techniques to Predict Diabetes Mellitus", *2019 International Conference on Electrical Computer and Communication Engineering (ECCE)*, pp. 1-4, 2019.
- B. Jain, N. Ranawat, P. Chittora, P. Chakrabarti and S. Poddar, "A machine learning perspective: To analyze diabetes", *Mater. Today: Proc.*, pp. 1-5, 2021.
- S. Ismaeel, A. Miri and D. Chourishi, "Using the Extreme Learning Machine (ELM) technique for heart disease diagnosis", *2015 IEEE Canada International Humanitarian Technology Conference (IHTC 2015)*, pp. 1-3, 2015.
- A. Ed-Daoudy and K. Maalmi, "Real-time machine learning for early detection of heart disease using big data approach", *2019 International Conference on Wireless Technologies Embedded and Intelligent Systems (WITS)*, pp. 1-5, 2019.
- P. Patil, N. Yaligar and S. M. Meena, "Comparision of performance of classifiers-SVM RF and ANN in potato blight disease detection using leaf images", *Proc. IEEE Int. Conf. Comput. Intell. Comput. Res. (ICCIC)*, pp. 1-5, 2017.
- Ponnusamy Deepika, "Heart disease prediction using classification with different decision tree techniques", *International Journal of Engineering Research and General Science*, vol. 2, pp. 6-11, 2014.

Authors:

Mr. Himambasha Shaik is an Assistant Professor in the Department of Master of Computer Applications at QIS College of Engineering and Technology, Ongole, Andhra Pradesh. He earned his Master of Computer Applications (MCA) from Anna University, Chennai. With a strong research background, He has authored and co-authored research papers published in reputed peer-reviewed journals. His research interests include Machine Learning, Artificial Intelligence, Cloud Computing, and Programming Languages. He is committed to advancing research and fostering innovation while mentoring students to excel in both academic and professional pursuits.

Ch.Srinivasa Rao is an MCA Scholar, Dept. of MCA, In QIS College of Engineering & Technology, Ongole. His areas of interest are Machine Learning, Deep Learning.