

Analyzing the Role of Stress and Sleep Patterns in Diabetes Risk Prediction Using Machine Learning

Vidhi¹, Dr. Vishal Verma², Ashish³

¹MCA Student, ²Professor, ³Asst. Prof.

Deptt. of Comp. Sc. and Application, Chaudhary Ranbir Singh University, Jind, Haryana, India

Abstract: Diabetes is an emerging health issue today. Existing work on diabetes prediction is essentially limited to medical factors like blood glucose level, blood pressure and other clinical factors. Also lifestyle risk factors such as stress, sleep and activity are important as these can impact overall health and have a role in developing diabetes. These factors have been important in most cases, but are under utilized in traditional prediction models.

This study investigated, using Machine Learning methods, stress and sleep pattern effects on its relationships with diabetes risk prediction. A data-set on diabetes that consists of demographic, lifestyle and clinical information was utilized for the above purpose. Firstly, Exploratory Data Analysis (EDA) was carried out to get an insight into the data, look for patterns and identify relationships between various variables. Following the data pre-processing phase, 4 Machine Learning algorithms were used and tested: Logistic Regression, Decision Tree, Support Vector Machine (SVM) and Random Forest.

The accuracy of the experimental results proved that the model Random Forest is the best model because it got 85.5% accuracy, better than the other models. In addition, the feature importance analysis revealed that child's BMI, age, family history of diabetes, duration of sleep, and exercise habits were important variables for the prediction of diabetes. Results of this study indicate that by integrating the lifestyle related information with conventional physician parameters, a better performance for the diabetes prediction models is achieved. This could be a way to detect high-risk individuals at early stages and aid in providing better preventive healthcare.

Keywords: Diabetes Mellitus, Machine Learning, Random Forest, Stress, Sleep Patterns, Risk Prediction.

Introduction

One of the most common chronic diseases globally, diabetes also reportedly costs many health-care systems a great deal, with millions of lives impacted [1]. Risk assessment in early stages can be important because if the person can take preventive measures, they will have a reduced risk for future severe health issues. The majority of diabetes prediction models are based on clinical factors, such as blood glucose and blood pressure. But it's generally a lifestyle choice that is addressed by very few people, even though this can have a significant effect on diabetes risk: stress and sleep patterns.

In this study, we employed machine learning techniques to explore whether there was a relationship between levels of stress and sleep, and between these and prediction of diabetes. The data subjected to analysis were a diabetes risk data with demographic, lifestyle and clinical data. First, Exploratory Data Analysis (EDA) was conducted to gain insight into the data, discover patterns and look for correlations between various variables. Pre-processing the data and then using and testing four machine learning algorithms; logistic regression, decision tree, support vector machine (SVM), and random forest were followed.

The performance of the model was analyze, and it was found that the model RM giving RANDOM FOREST performed better with accuracy around 85.5%. Based on these additional feature importance analyses, authors concluded that the most important factors contributing to diabetes risk prediction were the following: Body Mass Index (BMI); age; family history of diabetes; sleep duration and physical activity. Past research has also found that repeated sleep restriction has the

potential to have a strong impact on metabolic and endocrine functions [3]. The results and analyses of this research indicate that incorporating the following features into the written curriculum might be helpful for teachers:

Lifestyle-related factors along with the traditional clinical factors can help boost the accuracy of the early diabetes prediction models, which can aid in creating more efficient preventive healthcare strategies.

A lot of studies says that long-term stress and poor sleep are connected to insulin resistance and high blood sugar. But still, most diabetes models still ignore these factors. We think machine learning can help here. ML can look at how age, medical data, and lifestyle factors work together. This can make diabetes risk prediction more accurate.

Main goals of the research:

1. Check the link between stress, sleep, and diabetes risk.
2. Build ML models for predicting diabetes risk.
3. Compare how well different algorithms perform.
4. See how important lifestyle factors are for diabetes prediction.

• LITERATURE REVIEW

Many researchers used ML models such as Logistic Regression, Decision Trees, Support Vector Machines, and Random Forest for diabetes prediction. Most of the research has reported prediction accuracies ranging from 85% to 95% using clinical features of data only. However, most studies have focused primarily on medical indicators such as glucose levels and body mass index. Both quantity and quality of sleep significantly predict the risk of developing type 2 diabetes[4].

Recent studies indicate that stress increases cortisol production, which contributes to insulin resistance, while lack of sleep negatively affects metabolic regulation. Stress and the individual : Mechanisms leading to disease[2].Still, the combined influence of stress and sleep patterns in machine learning-based diabetes prediction remains insufficient.

• METHODOLOGY

• Dataset Description

We uses a diabetes risk dataset containing demographic, lifestyle, and clinical attributes. It includes age, body mass index, sleep duration, stress level, physical activity, family history, and diabetes risk category.

• Data Preprocessing Stages

First, we removed all the duplicate records from our dataset. After that, we filled the missing values using imputation techniques. We also converted the categorical variables into numbers. Then we standardized all the features so they were on the same scale. In the end, we split the data into 80% for training and 20% for test.ing.

- **Machine Learning Models:**

1. Logistic Regression
2. Decision Tree
3. Support Vector Machine
4. Random Forest

- **Evaluation Metrics**

- Accuracy
- Precision
- Recall
- F1-Score
- ROC-AUC Score

- **RESULTS AND DISCUSSION**

- Exploratory Data Analysis :

The Exploratory Data Analysis (EDA) revealed a clear relationship between lifestyle factors and diabetes risk. The results indicated that individuals with shorter sleep duration and higher stress levels were more likely to have an increased risk of diabetes. In contrast, individuals who engaged in regular physical activity showed a significantly lower risk. These findings suggest that lifestyle factors, particularly sleep patterns, stress levels, and physical activity, play an important role in diabetes risk prediction.

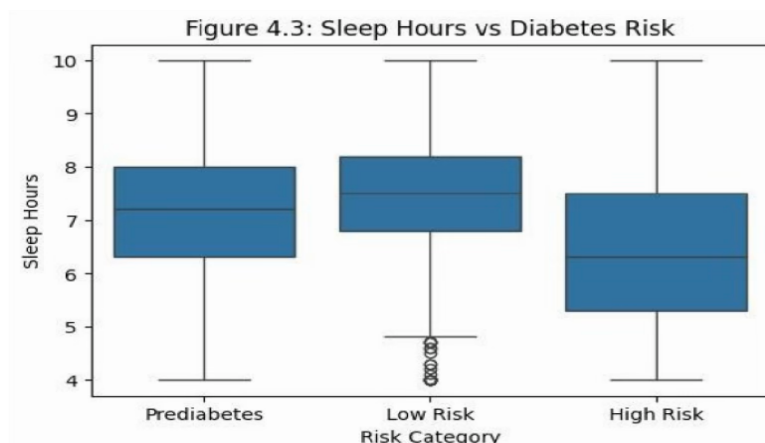


Table4.1: Classification Report of Random Forest Model

Class	Precision	Recall	F1-Score	Support
High Risk	0.89	0.89	0.89	449
Low Risk	0.91	0.92	0.93	520
Prediabetes	0.64	0.61	0.63	231

• Model Performance

After evaluating all the selected machine learning algorithms, the Random Forest model demonstrated the best performance, achieving an accuracy of approximately 85.5%. This accuracy was higher than that of the other models tested, making Random Forest the most effective algorithm for diabetes risk prediction in this study.

Model| Accuracy

Logistic Regression| 81.2%

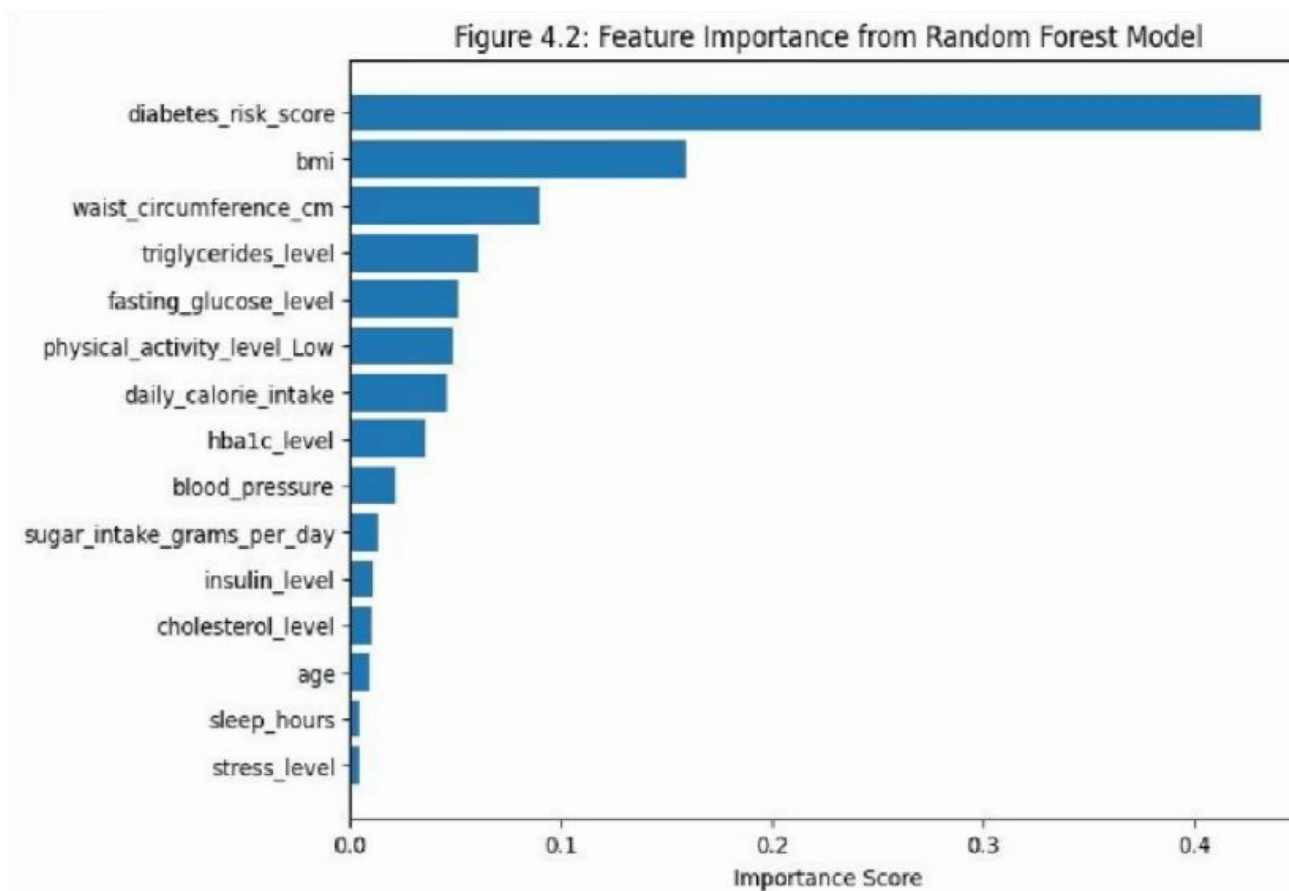
Decision Tree| 83.4%

Support Vector Machine| 82.7%

Random Forest| 85.5%

Feature Importance Analysis

Figure 4.2 presents the feature importance analysis, highlighting the variables that contributed most significantly to diabetes risk prediction. The results showed that age, Body Mass Index (BMI), family history, sleep duration, and physical activity were among the most important predictive factors. These findings emphasize the importance of incorporating lifestyle-related factors, along with traditional clinical indicators, into early diabetes screening and risk assessment models.



• CONCLUSION AND FUTURE WORK

Overall, the results of this study show that machine learning models can effectively predict diabetes risk when both medical and lifestyle factors are considered together. In particular, stress levels and sleep patterns were found to be important indicators in identifying individuals at risk. Among all the models tested, Random Forest performed the best and gave the highest accuracy, making it the most suitable model for this dataset.

These findings suggest that such models could be useful for early screening and risk assessment in real-world healthcare settings. By combining clinical data with lifestyle information, predictions become more accurate and meaningful.

For future work, there are several possible improvements. A larger and more balanced dataset could be used to improve model performance and generalization. More advanced techniques such as deep learning models could also be explored. In addition, developing a real-time web or mobile application would make the system more practical for everyday use. Finally, integrating data from wearable devices like smart watches could help capture sleep and stress patterns more accurately, further improving prediction reliability.

REFERENCES :

1. World Health Organization, Global Report on Diabetes, Geneva, Switzerland, 2016.
2. B. S. McEwen, "Stress and the individual: Mechanisms leading to disease," Annual Review of Public Health, vol. 36, pp. 441–462, 2015.
3. C. Spiegel et al, "Impact of sleep deprivation on metabolic and endocrine function," The Lancet, vol. 354, no. 9188, pp. 1435–1439, 1999.
4. F. P Cappuccio et al, "Quantity and quality of sleep and incidence of type 2 diabetes," Diabetes Care, vol. 33, no.2, pp. 414–420, 2010