

Transforming Type 2 Diabetes Management Through Telemedicine, Data Mining and Environmental Insights

Sapna S. Basavaraddi¹ and A. S. Raju²

¹SSIT, SSAHE, Agalakote, B.H. Road, Tumakuru-572105, Karnataka, India

²Department of Medical Electronics, SSIT, SSAHE, Agalakote, B.H. Road, Tumakuru-572105, Karnataka, India

†Corresponding author: Sapna S. Basavaraddi; sapnasbasavaraddi@gmail.com , asraju74@gmail.com

Abbreviation: Nat. Env. & Poll. Technol.
Website: www.neptjournal.com

Received: 11-08-2024

Revised: 23-09-2024

Accepted: 21-10-2024

Key Words:

Diabetes mellitus
 Type-2 diabetes
 Telemedicine
 Environment
 Arsenic
 Machine learning
 Data mining
 Predictive analytics
 Healthcare management

Citation for the Paper:

Sapna S. Basavaraddi and A. S. Raju., 2025. Transforming type 2 diabetes management through telemedicine, data mining and environmental insights. *Nature Environment and Pollution Technology*, 24(2), p. B4256. <https://doi.org/10.46488/NEPT.2025.v24i02.B4256>

Note: From year 2025, the journal uses Article ID instead of page numbers in citation of the published articles.



Copyright: © 2025 by the authors
Licensee: Technoscience Publications
 This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

ABSTRACT

Diabetes mellitus is a prevalent chronic disease with significant implications for public health, including an expanded chance of coronary heart malady, stroke, persistent kidney illness, misery, and useful inability. In India, the predominance of diabetes among grown-ups matured 20 a long time and more seasoned rose from 5.5% in 1990 to 7.7% in 2016. Traditionally, diabetes management involves costly consultations and diagnostic tests, presenting challenges for timely diagnosis and treatment. Additionally, a comprehensive study was conducted to investigate the relationship between the incidence of type 2 diabetes mellitus (T2DM) and environmental exposure to arsenic in the form of air, water, and food pathways. The majority of the analyzed studies examined the levels of arsenic in water samples, with analyses of urine, blood, serum, and plasma samples coming next. Groundwater supplies may get contaminated by arsenic, especially in regions where arsenic deposits are naturally occurring or as a result of industrial activity. Additionally, various meals contain it, particularly rice, seafood, and poultry. Besides, it might be released into the environment by industrial processes such as coal combustion, smelting, and mining, which could lead to occupational exposure. There may be a genetic component to the association between arsenic exposure and the onset of diabetes. Ultimately, diabetes mellitus is enhanced by arsenic pollution through air, food, and drinking water. Advances in machine learning and telemedicine offer innovative solutions to address these challenges. Data mining, a crucial aspect of machine learning, facilitates the extraction of valuable insights from extensive datasets, enabling more efficient and effective diabetes management. This study explores a telemedicine-based system utilizing five classification techniques Tree, Naive Bayes, Support Vector Machine, and others to predict Type-2 diabetes. By leveraging real-time data analysis, the system aims to enhance early diagnosis and management of Type-2 diabetes, potentially preventing progression to critical conditions. The results evaluate the effectiveness of these models in a telemedicine context, identifying the best-performing model to assist healthcare professionals in making informed decisions for early intervention and improved patient outcomes.

INTRODUCTION

Diabetes has developed as a worldwide well-being emergency, with the scourge especially articulated in Southeast Asia, eminently in India. Around 72 million individuals in India are evaluated to have diabetes, whereas an extra 80 million are accepted to have pre-diabetes. Type 2 diabetes accounts for over 90-95% of these cases (Gupta et al. 2024, Kumar 2024). According to the World Wellbeing Organization (WHO), roughly 347 million individuals around the world are influenced by diabetes, and it is anticipated to become the seventh driving cause of passing all-inclusive by 2030 (“WHO/diabetes”, Rahman et al. 2023). The management of Type 2 diabetes has evolved rapidly over the past decade, with several new therapeutic agents introduced. In 2005, the Indian Council of Medical Research (ICMR) released guidelines for the management of Type 2 diabetes, which have been extensively embraced throughout the nation. A

subsequent ICMR workshop in 2018 focused on refining these guidelines (Kumar 2024). The short-term goals of this work are to discuss certain methodological modifications related to the environment, possible fixes, and further research opportunities. At this point, the most prevalent attributes are walking distance, air pollution, food, exercise, and environmental pathways (Schulz et al. 2016).

Mechanism Linking the Environment and T2DM

Environmental, biochemical, and behavioral risk variables are thought to combine to cause type 2 diabetes (Kahn et al. 2014, Chatterjee et al. 2017). If no environment encourages healthy lifestyles, it is taught that these lifestyles are undesirable. A comprehensive search was carried out in Pubmed, Scopus, web of Science, and Embase databases. The inclusion of some criteria was studied. Diabetes is a chronic metabolic disease characterized by the body's inability to appropriately control blood sugar levels. Diabetes mellitus type 2 is a chronic metabolic disease that usually shows up as adult food and a complex interaction between environmental and genetic factors causes it. This condition is characterized by insulin resistance and decreased insulin production, both of which cause long-term disruptions in glucose metabolism. As a result, high blood sugar levels can lead to many complications, such as cardiovascular disease, neurological conditions, kidney problems, and retinopathy (Barbagallo et al. 2022, Kakavandi et al. 2023, Armghan et al. 2023).

Empirical data has demonstrated a probable association between diabetes and exposure to environmental arsenic. Numerous studies have shown that extended exposure to arsenic may increase the likelihood of developing diabetes. There is evidence linking the high levels of arsenic in the population's drinking water to both occupational settings and demographics (Halim & Halim 2019). Studies revealed a 27% increase in the risk of developing diabetes is linked to a drinking water arsenic concentration rise of 15 micrograms per liter (Ke et al. 2023, Rahmani et al. 2023). Insulin signaling and secretion are negatively impacted by arsenic, which results in decreased glucose tolerance and elevated insulin resistance. This event raises the possibility that arsenic plays a role in the onset or exacerbation of diabetes mellitus 2. Furthermore, it has been demonstrated that exposure to arsenic increases oxidative stress and inflammation, two processes that are known to be important in the development and maintenance of diabetes (Nikraves et al. 2023, Colwell et al. 2023, Rangel-Moreno et al. 2022).

Arsenic's effects on the advancement of diabetes have been linked in numerous research. The data points to a possible connection between inorganic arsenic and the onset

of diabetes, especially in populations exposed to different concentrations of the metal. Chronic low-level arsenic exposure has been linked in certain publications to a higher prevalence of diabetes mellitus 2 (Li et al. 2023, Shokat et al. 2024, Nurchi et al. 2020, Tinkelman et al. 2020). Polluted air can raise type 2 diabetes risk. Studies from the USA and Europe reveal that there is a definite link between incident diabetes mellitus 2 and ambient PM 2.5(fine particulate matter 2.5 micrometer diameter). In a systematic review and meta-analysis, some countries showed that the risk of diabetes mellitus 2 rose by 8-10% per 10 microgram/meter increase in exposure, and the association was stronger in females. However, conclusions drawn from these studies cannot be directly collated to developing countries.

Ambient pollution and the levels are low in high-income countries. There have been a few attempts to assess the risk of diabetes associated with air pollution in developing nations where the air quality tends to be much poorer. A study from China showed 155 increased hazards of incidents of diabetes from every 10 microgram/meter increase in fine particulate matter of diameter of less than or equal to 2.5 micrometers. Another study from China reported an increase in the fasting plasma glucose (FPG) and glycated hemoglobin (HBA1c) by 0.025 and 0.001 m mole/liter, respectively, for every 10 microgram/meter increase in particulate matter of diameter less than or equal to 10 micrometers and an increase by 0.061 to 0.016 m mole /liter respectively for every 10 microgram/meter increase in fine particulate matter of diameter 2.5 micrometers.

An analysis of data from nearly 4 lakh individuals showed that exposure to higher levels of fine particulate matter in air of diameter 2.5 micrometers was associated not only with the increased risk of diabetes 2 but also mortality risk from baseline of diabetes mellitus 2 and its complications. In such an environment, educational and behavioral measures may be severely reduced or completely useless. Numerous studies have demonstrated the importance of behavioral, socioeconomic, demographic, and individual-level factors in predicting type 2 diabetes (Gray-Webb et al. 2013, Agardh et al. 2011). Reviews from the past also point to a connection between the environment and conditions like obesity, cardiovascular disease, hypertension, metabolic syndrome, and physical activity that are directly linked to type 2 diabetes (Black & Macinko 2008, Sallis et al. 2008, Poortinga 2006).

The hypothetical system shown in Fig. 1 describes the various ways that unique environmental factors could influence type 2 diabetes. Socio-ecological theories form the basis of this paradigm; they highlight that human behavior is determined by capacity and is supported by the socio-demographic, psycho-social, economic, organizational, and

physical environments (Sallis et al. 2012). In Fig. 1.1, the methods by which environmental factors affect the possibility of Type-2 diabetic complications are diagrammatically represented (Northridge et al. 2023). Within this framework, newer advancements have led to the development of categorization models utilizing a range of machine-learning approaches. By leveraging extensive or real-time data sets to aid in diagnosis and prediction, these models assist medical professionals in the identification and management of Type 2 diabetes (Kumar 2024).

Investigation

This study's main objective is to examine diabetes datasets and apply machine learning algorithms-namely Decision Tree, Random Forest, Neural Network, Support Vector Machine, and Naive Bayes-for the prediction of Type 2

diabetes (Teimoory & Keyvanpour 2024, Chakraborty et al. 2023, Uddin & Ali 2023). The research seeks to develop a robust prediction engine and a corresponding web application that enables users to predict diabetes using these algorithms. This study also explores the application of statistical models in machine learning and aims to enhance understanding of how these algorithms function in the context of diabetes prediction (Teimoory & Keyvanpour 2024).

MATERIALS AND METHODS

The study was done with preferred reporting items for systematic reviews and meta-analysis (PRISMA) statements. Based on study criteria such as

1. Article published in English
2. Full-text records

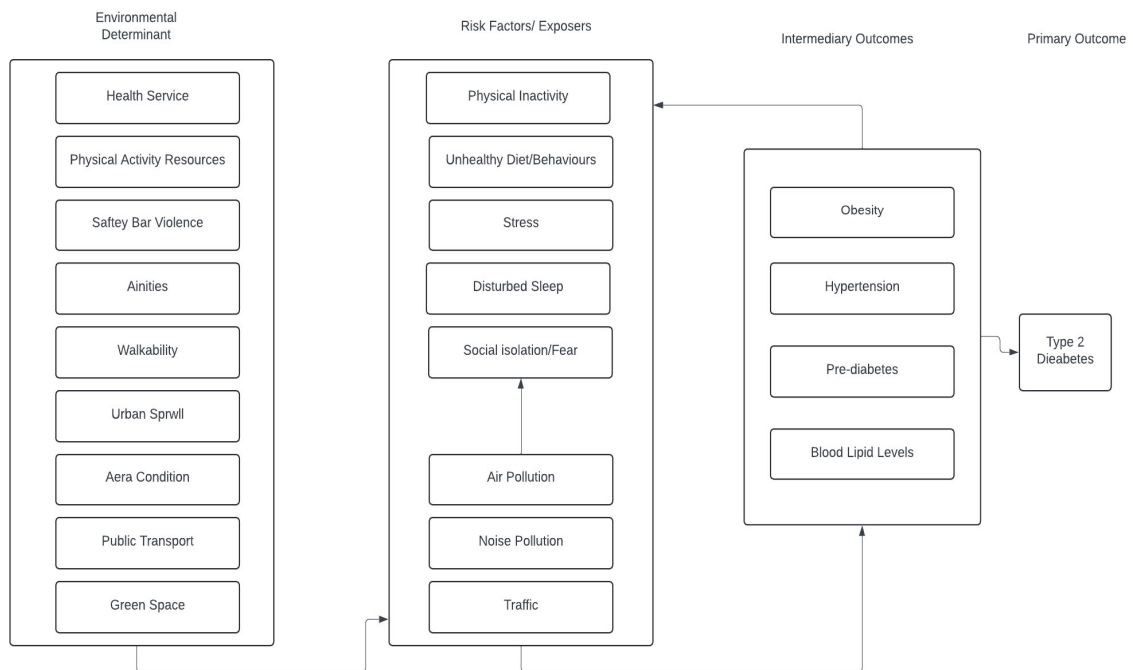


Fig. 1: Diagrammatic representation of the mechanisms by which environmental factors affect the risk of Type-2 diabetes mellitus.

Table 1: Clinical Differentiation between Type 1 and Type 2 Diabetes (LIANA 2024).

Clinical Point	Type 1 Diabetes	Type 2 Diabetes
Age	Typically diagnosed in youth	Commonly diagnosed in adults
Family History of Diabetes	Uncommon	Common
Ketosis at Diagnosis	Can occur	Rare
Insulin Markers	Absence	Presence
C-Peptide Assay	Lack of reserve beta cells	Presence of reserve beta cells
Pancreatic Autoantibodies	Presence	Absence

- 3. Articles relating arsenic exposure through the drinking water, air, and food pathways and its effects on diabetes mellitus 2.
- 4. Exclusion criteria like books, clinical trials and conference articles.

Along with these criteria, followed by a real-time data set and online dataset to study the prevalence accessibility of diabetes mellitus 2 and its management to reduce mortality rate, it also emphasizes the linkage between air pollution in the form of fine particulate matter less than or equal to 2.5-micrometer diameter and also arsenic pollution through drinking water and food and other pollutant toxins.

Diabetes mellitus is classified into the following types:

- 1. Type 1 Diabetes: Previously known as “Insulin-Dependent Diabetes Mellitus,” this kind of disease is caused by the body’s inability to manufacture insulin (IDDM).
- 2. Type 2 Diabetes: This kind is marked by an absolute insulin deficit that can occasionally coexist with insulin resistance, a condition in which cells are unable to use insulin as intended. It was once referred to as “Adult-

Onset Diabetes” or “Non-Insulin Dependent Diabetes Mellitus” (NIDDM).

- 3. Type 3 diabetes, often known as gestational diabetes, is characterized by increased blood glucose levels and affects pregnant women who have never been diagnosed with the disease (Chakraborty et al. 2023, Uddin & Ali 2023).
- 4. The following Table 1 describes the Clinical Differentiation between Type 1 and Type 2 Diabetes.

Overview of Types-2 Diabetes

Diabetes type 2 may be a complicated metabolic and vascular condition primarily marked by insulin resistance and, to differing degrees, insulin secretory absconds. It may be a dynamic condition, as often as possible, related to central corpulence, dyslipidemia, and hypertension. Although type 2 diabetes is most common among overweight and hefty people of center to late age, it is progressively watched in more youthful populaces and those with lower body mass index (BMI). Outstandingly, South Asians display the “Asian Indian phenotype,” where, at a given BMI level, Compared to Caucasians, they typically

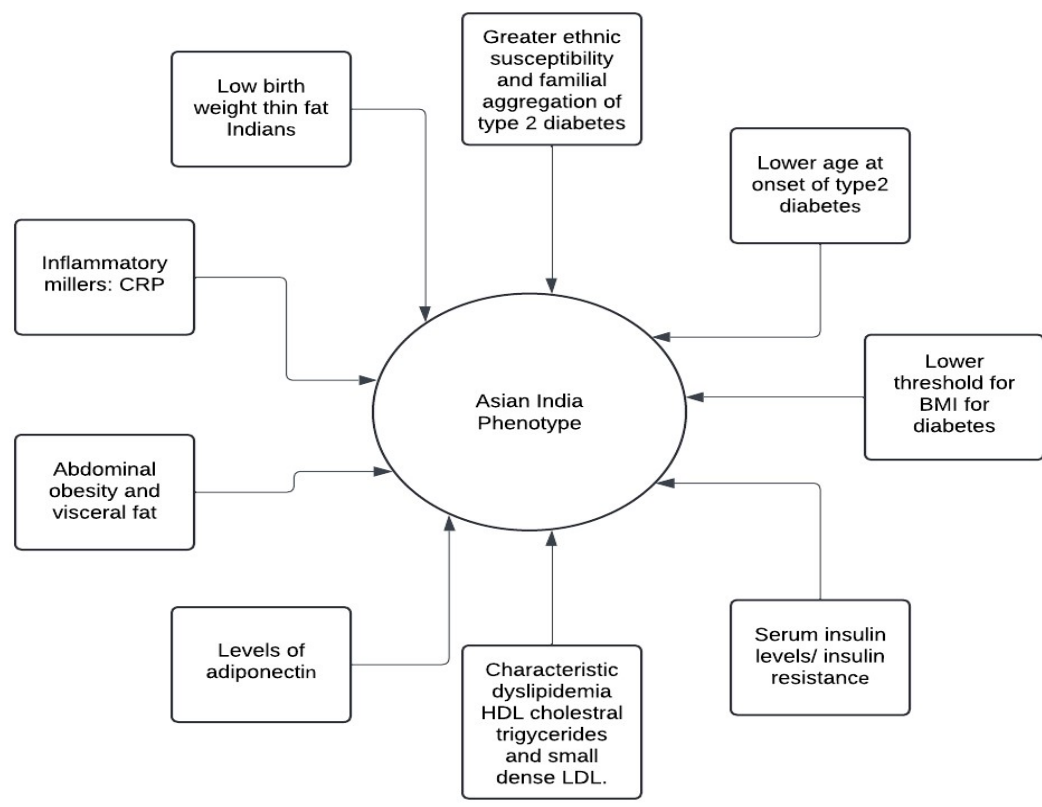


Fig. 2: Block diagram of Asian Indian phenotype.

have total body fat, extra visceral fat, increased insulin resistance, and a higher prevalence of diabetes (see Figure 2).

Data Collection

Actual data were gathered from diagnostic facilities in accordance with accepted clinical standards for the diagnosis of diabetes. The following Table 2 describes the diagnostic criteria for Diabetes and Pre-diabetes disease. Among the diagnostic standards used are:

- Random Plasma Glucose: $\geq 200 \text{ mg.dL}^{-1}$
- Fasting Plasma Glucose (FPG): $\geq 126 \text{ mg.dL}^{-1}$
- 2-Hour Post 75g Glucose (hPG): $\geq 200 \text{ mg.dL}^{-1}$
- Glycated Hemoglobin (HbA1c): $\geq 6.5\%$

Patients presenting with symptoms such as osmotic symptoms, weight loss, tiredness, weakness, recurrent urogenital infections, and delayed wound healing were assessed. Notably, some patients with diabetes may exhibit no symptoms. Pre-diabetes, identified as impaired glucose tolerance (IGT), was also evaluated to gauge early risk.

Based on family history, symptoms of insulin resistance, hypertension, dyslipidemia, and polycystic ovarian syndrome (PCOS), specialists frequently predict Type 2 diabetes. This research leverages real-time data collected at Clinical

diagnostic centers to enhance prediction accuracy. Machine learning algorithms were applied to analyze patterns in the collected data. This approach aimed to improve the prediction and classification of diabetes and pre-diabetes, providing a valuable tool for doctors in the diagnostic process. The data mining techniques used were selected to identify underlying patterns and trends, thus supporting more accurate diabetes classification and management.

Methodology

Data Description

The evaluation of the performance of various classification algorithms on two distinct datasets and reviews relating to arsenic in drinking water, food, fine particulate matter & pollution in the air to predict diabetes.

Dataset 1: the dataset is collected from the online Kaggle website. It consists of 5,000 samples with nine features, including gender, age, hypertension, heart disease, smoking history, BMI, HbA1c level, blood glucose level, and diabetes as the target variable. The following Table 3 is the sample table of the online dataset.

Dataset 2: the real-time data set is collected from the Clinical Diagnostic Center. It Comprises 48 samples with 19 features, including lab_id, gender, age, fasting blood sugar levels,

Table 2: Diagnostic criteria for diabetes and pre-diabetes.

Parameter	WHO/ADA Criteria	Diabetes
Fasting Plasma Glucose (FPG) [mg.dL ⁻¹]	< 110 mg.dL ⁻¹	126 mg.dL ⁻¹ or higher
2-Hour Postprandial Glucose(2-hPG) (mg/dl)	< 140 mg.dL ⁻¹	200 mg.dL ⁻¹ or higher
Glycated Hemoglobin (HbA1c)[%]	< 5.7%	6.5% or higher
Random Plasma Glucose	Not specified	200 mg.dL ⁻¹ or higher with symptoms

Table 3: The Sample Table of Online Dataset.

SI.No	Gender	Age	hypertension	heart_disease	smoking_history	BMI	HbA1c_Level	blood_glucose_level	diabetes
1	Female	80	0	1	never	25.19	6.6	140	0
2	Female	54	0	0	No Info	27.32	6.6	80	0
3	Male	28	0	0	never	27.32	5.7	158	0
4	Female	36	0	0	current	23.45	5	155	0
5	Male	76	1	1	current	20.14	4.8	155	0

Table 4: The Sample Table of Real-Time Dataset.

SI.No	lab_id	Gender	Age	fbs1	fns1	ppbs1	ppvs1	HbA1c	Mbg	result	Remarks
1	4	M	49	209	1	354	1.5	9.3	220.21	1	dm
2	11	M	28	81	0	105	0	5.4	108.28	0	nd
3	12	F	25	76	0	99	0	5.1	99.67	0	nd
4	15	M	65	69	0	104	0	5.6	114.02	0	nd
5	19	M	51	164	0.5	238	1	7.3	162.81	1	dm

postprandial blood sugar levels, HbA1c, and a target variable indicating diabetes presence. The following Table 4 is the sample table of the real-time dataset.

Data Preprocessing

Dataset 1:

- **Missing Value Imputation:** Examined for any missing data. The most frequent value was used to impute categorical data, and the mean was used for numerical features.
- **Encoding Categorical Variables:** One-hot encoding was used for categorical variables like “smoking history” and “gender,” whereas label encoding was used for other variables.
- **Feature Standardization:** Standardized numerical features were given a mean of 0 and a standard deviation of 1.

Dataset 2:

- **Missing Value Handling:** Missing values were handled appropriately, with numerical features filled using the mean and categorical features filled using the mode.
- **Feature Transformation:** Features were encoded and scaled as necessary for compatibility with the classification algorithms.

Managing Class Disparities

Dataset 1:

- Created synthetic samples for the minority class using the Synthetic Minority Over-sampling Technique (SMOTE), which balanced the distribution of classes.

Dataset 2:

- The dataset was relatively balanced, but preprocessing was applied to ensure class distributions were optimal for modeling.

Model Training And Evaluation

Dataset 1:

- Trained and evaluated six classifiers: Support Vector Machine (SVM), Decision Tree, Random Forest, Naive Bayes and Neural Network
- Split the dataset into training (70%) and testing (30%) sets.
- Employed Stratified K-Fold cross-validation (5 folds) to assess model performance.

Dataset 2:

1. Trained and evaluated the same classifiers with cross-validation techniques suited for smaller datasets.

2. Used the entire dataset for training and evaluation due to its smaller size.

Classifiers and configurations

- **SVM:** RBF kernel approximation.
- **Decision Tree:** Limited maximum depth to avoid overfitting.
- **Random Forest:** 100 trees with a maximum depth of 10.
- **Naive Bayes:** classifier with probability based on the Bayes theorem.
- **Neural Network:** 100 neurons in a single hidden layer with ReLU activation.

RESULTS AND DISCUSSION

Arsenic is a heavy metalloid that can be toxic to the body and has no place in human physiology. Type 2 diabetes is more common in people who have been exposed to high doses of arsenic over an extended period (Fan et al. 2022, Tsai et al. 1999, Guisela et al. 2022, Kile & Christiani 2008). Due to the fact that prolonged exposure to arsenic is associated with impaired glucose tolerance and insulin resistance, two conditions that are known to exist. The characteristic insulin resistance found in people with type 2 diabetes. Uncertainty exists regarding the exact mechanisms by which arsenic influences the development of type 2 diabetes (Castriota et al. 2018, Ghaedrahmat et al. 2021, Dalal et al. 2010, Spratlen et al. 2018). The likelihood of developing diabetes increases with continuous exposure to arsenic and with increasing concentrations of the element, according to weighted data (Spratlen et al. 2019, Wang et al. 2018, Chang et al. 2020). More research that looked into the quantity of arsenic in water samples was included in the review. Arsenic analysis in plasma, blood serum, and urine samples came in second. The inferior classes included food, diet, nails, and tears, in addition to air samples.

Environmental Exposure to Arsenic and Type 2 Diabetic Mellitus

Long-term exposure to arsenic through drinking water has been linked to an increased incidence of micro and macro vascular complications of diabetes type 2-such as myocardial infarctions, heart failure, stroke, retinopathy, diabetic foot, and polyneuropathy (Jovanović et al. 2019, Hsu et al. 2016, Bräuner et al. 2014, Rahman et al. 1998, Andrew et al. 1998, Chakraborti et al. 2011). This systematic review looked at several studies that looked into the relationship between the prevalence of diabetes and long-term exposure to arsenic via food, drink, and airborne environments (D’Ippoliti et

al. 2010, Huang et al. 2014, Rao et al. 2022). Consuming arsenic-tainted water can lower insulin production and increase insulin resistance, which can lead to beta-cell dysfunction and insulin resistance. Furthermore, oxidative stress and inflammation—both of which are connected to diabetes mellitus—can result from exposure to arsenic (Diaz-Villasenor et al. 2013).

These mechanisms may explain why consuming arsenic-contaminated water results in diabetes and other associated health problems (James et al. 2013, Fevrier-Paul et al. 2021, Drobná et al. 2013, Rahman et al. 1998, Sriparaya et al. 2017). Genetics and drinking water exposure to arsenic were also associated with an increased incidence of diabetes 2. People with particular mutations in the notch Recept (Spangler 2012) or 2 Gene (NOTCH2) were more likely to develop diabetes 2 when exposed to inorganic arsenic. There is a correlation between the death rate from diabetes mellitus and the county-level air concentration of arsenic in each of the 100 counties of North Carolina, in addition to drinking water sources. Furthermore, it is well acknowledged that exposure to particulate matter containing arsenic, beryllium, cadmium, and nickel has a negative effect on health, potentially increasing death rates and elevating the risk of Type 2 diabetes (Spangler 2012, Riseberg et al. 2021).

Demographic Characteristics and Socio-Economic Status

Based on demographic factors like gender, age, genetics, obesity, and socioeconomic level, some research demonstrates

Table 5: Test Set Accuracy.

Classifier	Dataset 1 Accuracy	Dataset 2 Accuracy
SVM	0.91	1.00
Decision Tree	0.92	1.00
Random Forest	0.98	1.00
Naive Bayes	0.86	1.00
Neural Network	0.93	0.50

a relationship between arsenic exposure and the development of Type 2 diabetes. Regarding this, some findings were noted across many populations, including those in the US, Taiwan, Korea, and China. Diabetes increases the incidence of internal malignancies, such as those of the stomach, colon, liver, pancreas, and lungs. When diabetes patients had elevated arsenic levels, the correlation was especially strong (Hendryx et al. 2021). For instance, the amount of 20 metals in the urine of middle-aged women participating in the study was measured. This finding indicated a link between T2DM in women and the excretion of metals like arsenic (Wang et al. 2020). Thus, to eliminate the accessibility of heavy metals that cause T2DM and ultimately result in mortality, the primary focus needs to be on drinking water, food, and crops. To reduce and assist physicians in identifying the primary causes of diabetes and to identify previous treatments based on reviews and datasets that are currently available, the best machine learning techniques or classifiers were developed. This allowed medical professionals to provide direct or telemedicine care while also lowering global death rates.

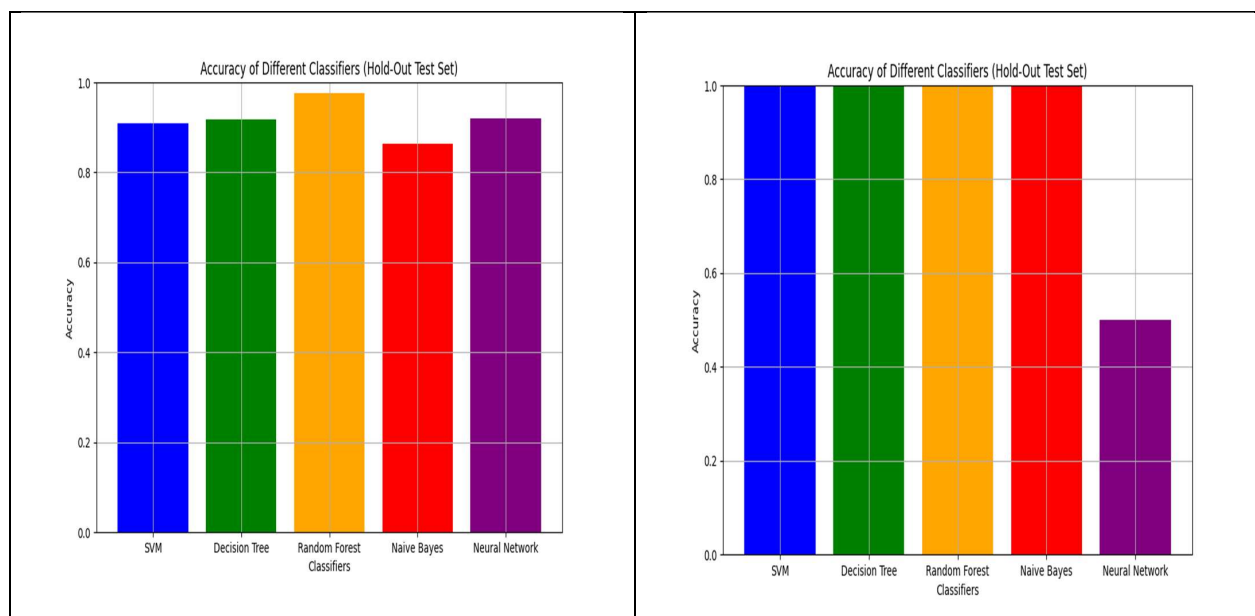


Fig. 3: Accuracy of different classifiers on Dataset 1 and Dataset 2.

Table 6: Cross-Validation Accuracy.

Classifier	Dataset-1 Accuracy	Dataset-1 Std Dev	Dataset-2 Accuracy	Dataset-2 Std Dev
SVM	0.91	0.01	1.00	0.00
Decision Tree	0.92	0.02	1.00	0.00
Random Forest	0.98	0.00	1.00	0.00
Naive Bayes	0.86	0.01	1.00	0.00
Neural Network	0.92	0.02	0.50	0.00

Dataset Results

The test accuracy of five classifiers, SVM, Decision Tree, Random Forest, Naive Bayes, and Neural Network, was evaluated across two datasets and is given in Table 5. On Dataset 1, Random Forest achieved the highest accuracy at 98%, followed by Neural Network (93%), Decision Tree (92%), SVM (91%), and Naive Bayes (86%). In contrast, for Dataset 2, all classifiers except the Neural Network reached 100% accuracy. The Neural Network, however, showed a significant drop to 50% accuracy on Dataset 2, suggesting potential challenges such as overfitting or data imbalance.

Table 7: Confusion Matrices.

Dataset 1

SVM:

	Predicted Non-Diabetic	Predicted Diabetic
Actual Non-Diabetic	1357	45
Actual Diabetic	85	513

Decision Tree:

	Predicted Non-Diabetic	Predicted Diabetic
Actual Non-Diabetic	1360	42
Actual Diabetic	72	526

Random Forest:

	Predicted Non-Diabetic	Predicted Diabetic
Actual Non-Diabetic	1392	10
Actual Diabetic	32	566

Naive Bayes:

	Predicted Non-Diabetic	Predicted Diabetic
Actual Non-Diabetic	1300	102
Actual Diabetic	95	503

Neural Network:

	Predicted Non-Diabetic	Predicted Diabetic
Actual Non-Diabetic	1362	40
Actual Diabetic	72	526

Dataset 2

SVM:

	Predicted Non-Diabetic	Predicted Diabetic
Actual Non-Diabetic	17	0
Actual Diabetic	0	31

Decision Tree:

	Predicted Non-Diabetic	Predicted Diabetic
Actual Non-Diabetic	17	0
Actual Diabetic	0	31

Random Forest:

	Predicted Non-Diabetic	Predicted Diabetic
Actual Non-Diabetic	17	0
Actual Diabetic	0	31

Naive Bayes:

	Predicted Non-Diabetic	Predicted Diabetic
Actual Non-Diabetic	17	0
Actual Diabetic	0	31

Neural Network:

	Predicted Non-Diabetic	Predicted Diabetic
Actual Non-Diabetic	17	0
Actual Diabetic	0	31

The bar graph in Fig. 3 illustrates the accuracy of five classifiers, SVM, Decision Tree, Random Forest, Naive Bayes, and Neural Network, across two datasets. Random Forest consistently achieved the highest accuracy on Dataset 1, while all classifiers except the Neural Network reached 100% accuracy on Dataset 2. The Neural Network’s significant drop in performance on Dataset 2, with only 50% accuracy, stands out, highlighting potential model-specific issues.

In Table 6, the cross-validation accuracy and standard deviation of five classifiers are compared across two datasets. On Dataset 1, Random Forest achieved the highest accuracy (0.98), while Naive Bayes had the lowest (0.86). For Dataset 2, all classifiers except the Neural Network achieved perfect accuracy (1.00) with no variation. The Neural Network showed a notable drop in performance on Dataset 2, with an accuracy of 0.50 and no deviation.

The confusion matrices compare the performance of five classifiers-SVM, Decision Tree, Random Forest, Naive Bayes, and Neural Network—across two datasets, as shown in Table 7. On Dataset 1, Random Forest exhibited the

highest accuracy, with only 10 false positives and 32 false negatives, while Naïve Bayes showed the most errors. For Dataset 2, all classifiers performed flawlessly, correctly classifying both non-diabetic and diabetic cases with no misclassifications.

Key Findings

1. Dataset Characteristics and Performance

- **Dataset 1** (5,000 samples) and **Dataset 2** (48 samples) presented different challenges due to their size and feature composition. Dataset 1, with its larger size and varied features, demonstrated generally lower but still competitive accuracy scores compared to Dataset 2, which, despite its small size, showed perfect classification results for most models.

2. Classifier Performance

- **Random Forest** emerged as the most effective classifier for Dataset 1, achieving the highest cross-validation accuracy (0.98) and test set accuracy (0.98). This robustness indicates Random Forest's ability to handle complex feature interactions and its generalization capability.
- For Dataset 2, all classifiers achieved perfect test set accuracy except the neural network. However, permutation testing showed a p-value of 1.0000 for the SVM classifier, suggesting that while the model performed well, the results could be influenced by the small sample size.

3. Model Evaluation Metrics

- **Cross-Validation Accuracy:** Random Forest consistently outperformed other models in Dataset 1. For Dataset 2, classifiers achieved perfect accuracy, but this is likely influenced by the small sample size.
- **Confusion Matrices:** The confusion matrices highlighted the models' ability to correctly classify diabetic and non-diabetic instances. For Dataset 1, Random Forest had the lowest number of misclassifications. For Dataset 2, all classifiers achieved perfect results, indicating a lack of variance in the data.

4. Feature Importance

- In Dataset 1, Random Forest's feature importance analysis identified Blood Glucose Level, HbA1c Level, and BMI as the most influential features, aligning with clinical expectations. Dataset 2's limited feature set and size did not permit a detailed feature importance analysis.

Implications

- The findings suggest that while Random Forest is a highly effective algorithm for larger, complex datasets, smaller datasets may yield inflated performance results due to limited data variability. Future research should focus on expanding dataset sizes, exploring additional features, and applying more advanced preprocessing techniques to enhance model performance and generalizability. This research work emphasizes the purpose of the application of data mining techniques in telemedicine.
- In the telemedicine system, regular visits to the hospitals for patients suffering from diseases are minimized since they can be expensive for rural background patients.
- During the COVID-19 pandemic, the physical presence of the patients and doctors became risky. People preferred the Telemedicine system.
- Telemedicine services are done through video conferences & smartphones, which reduce the time and cost to patients.
- Furthermore, the telemedicine system has fast and advantageous characteristics. It can streamline the workflow of hospitals and clinics.
- This method is more useful in natural calamities, floods, war zones, and rural areas all over the world.
- The telemedicine system would make it easier to monitor discharged patients, manage their recovery from disease, and reduce mortality rates.

CONCLUSIONS

This systematic review and dataset study discusses the relationship between type 2 diabetes and arsenic exposure. Research has indicated that extended exposure to arsenic by ingestion or work-related exposure heightens the likelihood of acquiring Type 2 diabetes. Insulin resistance is thought to be caused by exposure to arsenic, which is thought to alter how the body uses insulin and how it is released. It is linked to inflammation and stress as well. The information now available suggests a possible link between T2DM and environmental arsenic exposure from food, water, and the air. The review reveals a close connection between T2DM and environmental degradation. This study presents a straightforward approach for assessing the prevalence of type 2 diabetes through dataset analysis utilizing different classifiers. It finds that the Random Forest classifier consistently delivers superior performance, achieving perfect accuracy of 100% on both real-time datasets. The research compares various classifiers, including SVM, Random Forest,

Naive Bayes, Neural Network, and Decision Tree, ultimately identifying Random Forest as the most effective method for disease prediction in telemedicine. With accuracy rates of 98% and 100% for online datasets, Random Forest is highlighted as the best option for detecting, controlling, and monitoring prevalent diseases while helping to reduce mortality rates. Thus, data mining techniques emerge as powerful tools for predicting and monitoring diseases within telemedicine systems. Doctors can use these accurate classifiers to help patients lead healthy lives, control the mortality rates in remote areas using telemedicine systems, or use a personal approach. To stop the disease's escalating burden, a fuller understanding of the connection between the environment and type 2 diabetes can assist in developing health-promoting policies and offer opportunities for people to translate their intentions into long-lasting behavioral adjustments.

Future Work

Further investigations are needed to:

- **Expand Dataset Size:** Get more samples to make sure the models are reliable and applicable to a wider range of situations.
- **Explore Additional Features:** Integrate more relevant features that could enhance predictive performance.
- **Hyper-parameter Tuning:** Perform extensive hyper-parameter optimization to further improve model accuracy.
- **Advanced Techniques:** Implement ensemble methods and deep learning approaches to address limitations observed in smaller datasets.

For future research, several key areas warrant attention: First, expanding the dataset size is crucial to obtain a more comprehensive range of samples, which will enhance the reliability and generalizability of the models across diverse situations. Second, exploring additional relevant features may significantly improve predictive performance, as incorporating more data points can provide a richer context for classification. Third, hyperparameter tuning should be prioritized to optimize model accuracy, allowing for fine-tuning of the algorithms to achieve better results. Finally, implementing advanced techniques, such as ensemble methods or deep learning approaches, can help overcome the limitations associated with smaller datasets and improve predictive capabilities in the context of telemedicine. This study underscores the importance of selecting appropriate models and evaluation techniques based on dataset characteristics and highlights the need for careful consideration of model performance metrics, especially in cases with small sample sizes.

REFERENCES

- Agardh, E., Allebeck, P., Hallqvist, J., Moradi, T. and Sidorchuk, A., 2011. Type 2 diabetes incidence and socio-economic position: A systematic review and meta-analysis. *International Journal of Epidemiology*, 40(3), pp.804–818. DOI
- Andrew, A.S., Jewell, D.A., Mason, R.A., Whitfield, M.L., Moore, J.H. and Karagas, M.R., 2008. Drinking water arsenic exposure modulates gene expression in human lymphocytes from a US population. *Environmental Health Perspectives*, 116(4), pp.524–531.
- Armghan, J., Logeshwaran, S.M., Sutharshan, K., Aliqab, M., Alsharari, M. and Patel, S.K., 2023. Design of biosensor for synchronized identification of diabetes using deep learning. *Results in Engineering*, 20, p.1382.
- Barbagallo, M., Veronese, N. and Dominguez, L.J., 2022. Magnesium in type 2 diabetes mellitus, obesity, and metabolic syndrome. *Biomedicine and Pharmacotherapy*, 87(3), p.714.
- Black, J.L. and Macinko, J., 2008. Neighborhoods and obesity. *Nutrition Reviews*, 66(1), pp.2–20. DOI
- Bräuner, E.V., Nordsborg, R.B., Andersen, Z.J., Tjønneland, A., Loft, S. and Raaschou-Nielsen, O., 2014. Long-term exposure to low-level arsenic in drinking water and diabetes incidence: A prospective study of the diet, cancer, and health cohort. *Environmental Health Perspectives*, 122(10), pp.1059–1065.
- Castriota, F., Acevedo, J., Ferreccio, C., Smith, A.H., Liaw, J., Smith, M.T. and Steinmaus, C., 2018. Obesity and increased susceptibility to arsenic-related type 2 diabetes in Northern Chile. *Environmental Research*, 167, pp.248–254.
- Chakraborti, D., Singh, S.K., Rashid, M.H. and Rahman, M.M., 2011. Arsenic: Occurrence in groundwater. *Encyclopedia of Environmental Health*, 2, p.117.
- Chakraborty, S., Panda, A.R., Vadiya, P., Samal, S., Gupta, I. and Ray, N.K., 2023. Predicting diabetes: A comparative study of machine learning models. *2023 OITS International Conference on Information Technology (OCIT)*, Raipur, India, pp.743–748.
- Chang, A., Ridpath, A., Carpenter, J., Kieszak, S., Sircar, K., Espinosa-Bode, A., Nelson, D. and Martin, C., 2020. Urine bisphenol A and arsenic levels in residents of the Cheyenne River Sioux Tribe, South Dakota, with and without diabetes. *Journal of Medical Toxicology*, 16, pp.276–283.
- Chatterjee, S., Khunti, K. and Davies, M.J., 2017. Type 2 diabetes. *The Lancet*, 389(10085), pp.2239–2251. DOI
- Colwell, M.L., Flack, N., Rezabek, A. and Faulk, C., 2023. Intergenerational arsenic exposure on the mouse epigenome and metabolic physiology. *Environmental and Molecular Mutagenesis*, 64(2), pp.72–87.
- Dalal, B., Nayak, K., Gangopadhyay, P. and Mukherjee, A., 2010. Identification of indicators of arsenic-induced hepatic damage in humans. *Internet Journal of Toxicology*, 7(1), p.71.
- Diaz-Villasenor, A., Cruz, L., Cebrian, A., Hernandez-Ramirez, R.U., Hiriart, M., Garcia-Vargas, G., Bassol, S., Sordo, M., Gandolfi, A.J., Klimecki, W.T. and Lopez-Carillo, L., 2013. Arsenic exposure and calpain-10 polymorphisms impair the function of pancreatic beta-cells in humans: A pilot study of risk factors for T2DM. *PLoS One*, 8(1), p.642.
- D'Ippoliti, D., Santelli, E., De Sario, M., Scorticini, M., Davoli, M. and Michelozzi, P., 2015. Arsenic in drinking water and mortality for cancer and chronic diseases in Central Italy, 1990–2010. *PLoS One*, 10(9), p.e013.
- Drobná, Z., Del Razo, L.M., García-Vargas, G.G., Sánchez-Peña, L.C., Barrera-Hernández, A. and Stýblo, M., 2013. Environmental exposure to arsenic, AS3MT polymorphism and prevalence of diabetes in Mexico. *Journal of Exposure Science and Environmental Epidemiology*, 23(2), pp.151–155.
- Fan, C., Zhan, Z., Zhang, X., Lou, Q., Guo, N., Su, M., Gao, Y., Qin, M.,

- Wu, L., Huang, W. and Zhang, M., 2022. Research for type 2 diabetes mellitus in endemic arsenic areas in central China: Role of low level of arsenic exposure and KEAP1 rs11545829 polymorphism. *Archives of Toxicology*, 96(6), pp.1673–1683.
- Fevrier-Paul, A.K., Soyibo, K., De Silva, N., Mitchell, S., Nwokocha, C. and Voutchkov, M., 2021. Addressing the challenge of potentially hazardous elements in the reduction of hypertension, diabetes, and chronic kidney disease in the Caribbean. *Journal of Health Pollution*, 11(30), Article 210613.
- Ghaedrahmat, Z., Cheraghian, B., Jaafarzadeh, N., Takdastan, A., Shabbazian, H.B. and Ahmadi, M., 2021. Relationship between urinary heavy metals with metabolic syndrome and its components in population from Hoveyeh cohort study: A case-control study in Iran. *Journal of Trace Elements in Medicine and Biology*, 66, Article 126757.
- Gray-Webb, T.L., Suglia, S.F. and Tehranifar, P., 2013. Social epidemiology of diabetes and associated conditions. *Current Diabetes Reports*, 13(6), pp.850–859.
- Guisela, B.Z., Ohana, N.D.A., Dalvani, S.D., Fermin, G.V., Francisco Hm, L. and Luis, N.-G., 2022. Adsorption of arsenic anions in water using modified lignocellulosic adsorbents. *Results in Engineering*, 13, p.340.
- Gupta, S., Sharma, N., Arora, S. and Verma, S., 2024. Diabetes: A review of its pathophysiology and advanced methods of mitigation. *Current Medical Research and Opinion*, 40(5), pp.773–780.
- Halim, M. and Halim, A., 2019. The effects of inflammation, aging, and oxidative stress on the pathogenesis of diabetes mellitus (type 2 diabetes). *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 13(2), pp.1165–1172.
- Hendryx, M., Luo, J., Chojenta, C. and Byles, J.E., 2021. Exposure to heavy metals from point pollution sources and risk of incident type 2 diabetes among women: A prospective cohort analysis. *International Journal of Environmental Health Research*, 31(4), pp.453–464.
- Hsu, L.I., Cheng, Y.W., Chen, C.J., Wu, M.M., Hsu, K.H. and Chiou, H.Y., 2016. Cumulative arsenic exposure is associated with fungal infections: Two cohort studies based on southwestern and northeastern basins in Taiwan. *Environmental International*, 96, pp.173–179.
- Hsu, L.I., Wang, Y.H., Chiou, H.Y., Wu, M.M., Yang, T.Y. and Chen, Y.H., 2013. The association of diabetes mellitus with subsequent internal cancers in the arsenic-exposed area of Taiwan. *Journal of Asian Earth Sciences*, 73, pp.452–459.
- Huang, J.W., Cheng, Y.Y., Sung, T.C., Guo, H.R. and Shihannopkao, S., 2014. Association between arsenic exposure and diabetes mellitus in Cambodia. *BioMed Research International*, 2014, p.54.
- James, K.A., Marshall, J.A., Hokanson, J.E., Meliker, J.R., Zerbe, G.O. and Byers, T.E., 2013. A case-cohort study examining lifetime exposure to inorganic arsenic in drinking water and diabetes mellitus. *Environmental Research*, 123, pp.33–38.
- Jovanović, D., Paunović, K., Manojlović, D. and Rasic-Milutinović, Z., 2019. Arsenic exposure from drinking water and the occurrence of micro- and macrovascular complications of type 2 diabetes. *Environmental Arsenic in a Changing World*, CRC Press, pp.331–332.
- Kahn, S.E., Cooper, M.E. and Del Prato, S., 2021. Pathophysiology and treatment of type 2 diabetes: Perspectives on the past, present, and future. *Lancet*, 383, pp.1068–1083.
- Kakavandi, N.R., Mousavi, T., Asadi, T., Moradi, A., Esmaeili, M., Sezavar, A.H., Nikfar, S. and Abdollahi, M., 2023. An updated systematic review and dose-response meta-analysis on the relation between exposure to arsenic and risk of type 2 diabetes. *Toxicology Letters*, 384, pp.115–127. DOI.
- Ke, C., Narayan, K.V., Chan, J.C., Jha, P. and Shah, B.R., 2022. Pathophysiology, phenotypes and management of type 2 diabetes mellitus in Indian and Chinese populations. *Nature Reviews Endocrinology*, 18(7), pp.413–432.
- Kile, M.L. and Christiani, D.C., 2008. Environmental arsenic exposure and diabetes. *JAMA*, 300(7), pp.845–846.
- Kumar, A., Gangwar, R., Ahmad Zargar, A., Kumar, R. and Sharma, A., 2024. Prevalence of diabetes in India: A review of IDF Diabetes Atlas 10th Edition. *Current Diabetes Reviews*, 20(1), e130423215752.
- Li, W., Li, Z., Yan, Y., Zhang, J., Zhou, Q., Jia, C., Xu, Y., Cui, H., Xie, S., Liu, Q. and Guan, Y., 2023. Urinary arsenic metabolism, genetic susceptibility, and their interaction on type 2 diabetes. *Chemosphere*, 345, Article 140536.
- Nikraves, M., Mahdavinia, M., Neisi, N., Khorsandi, L. and Khodayar, M.J., 2023. Citicoline ameliorates arsenic-induced hepatotoxicity and diabetes in mice by overexpression of VAMP2, PPAR- α , AS3MT, and SIRT3. *Pesticide Biochemistry and Physiology*, 192, p.391.
- Northridge, M.E., Sclar, E.D. and Biswas, P., 2003. Sorting out the connections between the built environment and health: A conceptual framework for navigating pathways and planning healthy cities. *Journal of Urban Health*, 80, pp.556–568.
- Nurchi, V.M., Buha Djordjevic, A., Crisponi, G., Alexander, J., Bjørklund, G. and Aaseth, J., 2020. Arsenic toxicity: Molecular targets and therapeutic agents. *Biomolecules*, 10(2), p.235.
- Poortinga, W., 2006. Perceptions of the environment, physical activity, and obesity. *Social Science & Medicine*, 63, pp.2835–2846.
- Rahman, M., Tondel, M., Ahmad, S.A. and Axelsson, O., 1998. Diabetes mellitus associated with arsenic exposure in Bangladesh. *American Journal of Epidemiology*, 148(2), pp.198–203.
- Rahman, M., Tondel, M., Chowdhury, I.A. and Axelsson, O., 1999. Relations between exposure to arsenic, skin lesions, and glucosuria. *Occupational and Environmental Medicine*, 56(4), pp.277–281.
- Rahman, T., 2023. A proposed method to identify the occurrence of diabetes in the human body using data analysis. *International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET)*, 10(3), pp.399–429.
- Rahmani, A., Khamutian, S., Doosti-Irani, A., Shokohizadeh, M.J., Shirmohammadi-Khorram, N., Sahraei, F., Khodabakhshi, M. and Ahangaran, N., 2023. The association of arsenic exposure with mortality due to cancer, diabetes, Alzheimer's, and congenital anomalies using Poisson regression. *Scientific Reports*, 13(1), p.456.
- Rangel-Moreno, K., Gamboa-Loira, B., López-Carrillo, L. and Cebrián, M.E., 2022. Prevalence of type 2 diabetes mellitus in relation to arsenic exposure and metabolism in Mexican women. *Environmental Research*, 210, p.48.
- Rao, N.S., Raju, G.N., Tiwari, M., Naidu, B. and Sarita, P., 2022. Serum elemental analysis of type 2 diabetes patients using SRXRF. *Biological Trace Element Research*, 200(4), pp.1485–1494.
- Riseberg, E., James, K.A., Woodin, M., Melamed, R., Alderete, T. and Corlin, L., 2021. Multipollutant, longitudinal analysis of the association between urinary tungsten and incident diabetes in a rural population. *Environmental Epidemiology*, 5(6), p.615.
- Sallis, J.F., Floyd, M.F., Rodriguez, D.A. and Saelens, B.E., 2012. Role of built environments in physical activity, obesity, and cardiovascular disease. *Circulation*, 125, pp.726–737.
- Sallis, J.F., Owen, N. and Fisher, E.B., 2008. Ecological models of health behavior. *Health Behavior and Health Education: Theory, Research, and Practice*, 4, pp.465–486.
- Schulz, M., Romppel, M. and Grande, G., 2016. Built environment and health: A systematic review of studies in Germany. *Journal of Public Health*, 18, pp.1–8.
- Shokat, S., Iqbal, R., Riaz, S. and Yaqub, A., 2024. Association between arsenic toxicity, AS3MT gene polymorphism, and the onset of type 2 diabetes. *Biological Trace Element Research*, 202(4), pp.1550–1558.
- Spangler, J.G., 2012. Diabetes mortality and environmental heavy metals in North Carolina counties: An ecological study. *Dubai Diabetes and Endocrinology Journal*, 20, pp.71–74.
- Spratlen, M.J., Grau-Perez, M., Umans, J.G., Yracheta, J., Best, L.G. and Francesconi, K., 2018. Arsenic, one-carbon metabolism and diabetes-

- related outcomes in the Strong Heart Family Study. *Environmental International*, 121, pp.728–740.
- Spratlen, M.J., Grau-Perez, M., Umans, J.G., Yracheta, J., Best, L.G. and Francesconi, K., 2019. Targeted metabolomics to understand the association between arsenic metabolism and diabetes-related outcomes: Preliminary evidence from the Strong Heart Family Study. *Environmental Research*, 168, pp.146–157.
- Sripaoraya, K., Siri Wong, W., Pavitranon, S. and Chapman, R.S., 2017. Environmental arsenic exposure and risk of diabetes type 2 in Ron Phibun subdistrict, Nakhon Si Thammarat Province, Thailand: Unmatched and matched case-control studies. *Risk Management and Healthcare Policy*, 2017, pp.41–48.
- Teimoory, G.K. and Keyvanpour, M.R., 2024. An effective feature selection for Type II diabetes prediction. *Proceedings of the 10th International Conference on Web Research (ICWR)*, Tehran, Iran, pp.64–69.
- Tinkelman, N.E., Spratlen, M.J., Domingo-Relloso, A., Tellez-Plaza, M., Grau-Perez, M. and Francesconi, K.A., 2020. Associations of maternal arsenic exposure with adult fasting glucose and insulin resistance in the Strong Heart Study and Strong Heart Family Study. *Environmental International*, 137, p.553.
- Tsai, S.M., Wang, T.N. and Ko, Y.C., 1999. Mortality for certain diseases in areas with high levels of arsenic in drinking water. *Archives of Environmental Health*, 54(3), pp.186–193.
- Uddin, M. and Ali, A., 2023. Diabetes prediction with machine learning techniques. *2023 Global Conference on Information Technologies and Communications (GCITC)*, Bangalore, India, p.16.
- Wang, X., Karvonen-Gutierrez, C.A., Herman, W.H., Mukherjee, B., Harlow, S.D. and Park, S.K., 2020. Urinary metals and incident diabetes in midlife women: Study of Women's Health across the Nation (SWAN). *BMJ Open Diabetes Research and Care*, 8(1), p.233.
- Wang, X., Mukherjee, B. and Park, S.K., 2018. Associations of cumulative exposure to heavy metal mixtures with obesity and its comorbidities among US adults in NHANES 2003–2014. *Environmental International*, 121, pp.683–694.