ORIGINAL RESEARCH

Predicting Diabetes in United Arab Emirates Healthcare: Artificial Intelligence and Data Mining Case Study

Saada Khadragy[1], Mohamed Elshaeer[2], Talal Mouzaek[3], Demme Shammass[4], Fanar Shwedeh[5], Ahmad Aburayya[5], Ammar Jasri[6], Shaima Aljasmi[6]

[1]Assistant Professor, MIS Department, Business College, City University Ajman, Ajman, United Arab Emirates
[2]Pharma Program, College of Pharmacy, Gulf Medical University, Ajman, UAE
[3]Senior Specialist Physician, Sheikh Khalifa General Hospital Umm Al Quwain, Umm Al Quwain, UAE
[4]Specialist Internal Medicine, Intensive Care Unit, Midclinic City Hospital, Dubai, UAE
[5]Assistant Professor, MBA Department, Business College, City University Ajman, Ajman, UAE
[6]Senior Specialist Registrar, Dubai Academic Health Corporation, Dubai, UAE

Corresponding author: Dr. Fanar Shwedeh
MBA Department, Business College, City University Ajman, Ajman, UAE.

Abstract

Aim: The primary aim of this article is to address the scarcity of tools available to examine the relationships between different attributes in medical datasets within the healthcare industry. Specifically, the focus is on developing a predictive model for diabetes using Artificial Intelligence and Data Mining techniques in the United Arab Emirates healthcare sector.

Methods: The paper follows a comprehensive approach, employing the four data mining steps: data preprocessing, data exploration, model building, and model evaluation. To build the predictive model, the decision tree algorithm is utilized. Data from 2856 patients, collected from prime hospitals in Dubai, United Arab Emirates, are analyzed and used as the basis for model development.

Results: The research findings indicate that several factors significantly influence the likelihood of developing diabetes. Specifically, age, gender, and genetics emerge as critical determinants in predicting the onset of diabetes. The developed predictive model demonstrates the potential to provide accurate and easy-to-understand results regarding the likelihood of diabetes in the future.

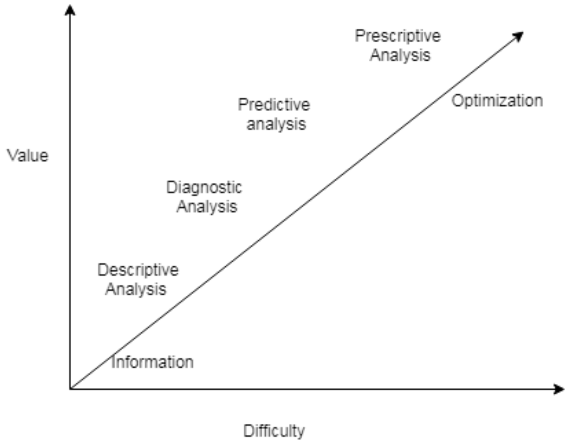
Conclusion: This study highlights the importance of Artificial Intelligence and Data Mining techniques in predicting diabetes within the United Arab Emirates healthcare sector. The findings emphasize the significance of age, gender, and genetics in diabetes prediction. This research addresses the current data scarcity and offers valuable insights for healthcare professionals. Furthermore, the study recommends further research to enhance diabetes prediction models and their application in clinical settings.

Keywords: Artificial Intelligence, Data mining, Decision Tree algorithm, Diabetes, Healthcare industry, Medical Center Data, Patient attributes, Predictive Modeling, Risk factors.

Introduction

The healthcare industry has accumulated vast data through record keeping and patient care (1,2). Efforts have been made to bridge the gap between generated and stored data, and automated medical records have facilitated the archiving of doctor-patient interactions. This emergence of "big healthcare data" presents opportunities for data analysis, healthcare management, and treatment outcome prediction. Big Data analytics has the potential to transform the healthcare sector by improving effectiveness, disease outbreak response, medical experimentation, and healthcare expenditure optimization (3,4). With the increasing aging population, the prevalence of chronic diseases, and expensive medical technology, there is a need to enhance the sustainability of healthcare models. By improving healthcare system efficiency, substantial savings in community spending can be achieved, reaching up to 20% of the gross domestic product on average within the OECD, potentially amounting to €330 billion in the United States of America (USA) and Europe based on 2021 figure 1.

Figure 1: Efficiency of the health care system



The healthcare industry has generated a significant amount of data that has impacted various fields. However, adopting big data approaches has been slow in the industry (3,4,8). This paper aims to explore the reasons behind this reluctance and highlight the potential of big data in extracting valuable insights from existing datasets. The report will examine the four perspectives of care data analytics (descriptive, diagnostic, prognostic, and prescriptive) to demonstrate the opportunities for innovation and the development of business models (9,10,11). Additionally, it will discuss how implementing big data technologies can enhance healthcare productivity and accessibility and outline the steps required to achieve this goal.

Literature Review

Big Data in the Healthcare Field

The healthcare industry generates significant data, encompassing patient records, diagnostic tests, treatment plans, and more. However, harnessing the potential of big data in healthcare requires advanced management and storage solutions. While data analytics is commonly used in the healthcare sector, implementing big data approaches is still in its early stages (6,7,11).

Healthcare data is often fragmented and stored in separate systems, making it challenging to access and analyze comprehensively. Integrating various data sources, including electronic medical records, vital signs, laboratory results, medication records, and patient-generated data from Internet of Things (IoT) devices, holds great potential for valuable insights (7,12). Linking these datasets can provide a holistic understanding of patient conditions, treatment outcomes, and population health.

This integrated approach to healthcare data analytics has implications for personalized treatment plans, disease progression prediction, resource allocation optimization, and lifestyle improvement programs. By identifying patterns and correlations, healthcare professionals can make informed decisions and improve patient outcomes (7).

Moreover, comprehensive analysis of healthcare data can drive innovation and inform business models in the healthcare industry. It enables the identification of trends, inefficiencies, and opportunities for improvement (13).

Although challenges exist in accessing and analyzing fragmented healthcare data, integrating and linking disparate datasets can unlock valuable insights for improved healthcare outcomes, resource allocation, and the development of innovative approaches in the field. The full potential of big data in healthcare is yet to be realized, but its impact holds promise for transformative advancements in patient care and healthcare management (14).

Medical Database

The elderly population is growing rapidly, leading to an increase in demand for healthcare services due to the prevalence of chronic diseases among the elderly. According to projections, the number of individuals aged 85 years and older will increase from 14 million to 19 million by 2020 and to 40 million by 2050. The impact of this growing demand is evident from a study by Accenture conducted in 2021, which found that one-third of European hospitals reported operating losses (1,3).

To analyze this challenge, the concept of the Triangle of Healthcare is often used, which comprises three elements: quality, access, and cost (4,5). The effectiveness, value, and resulting outcomes of care reflect the quality of a healthcare system. Access refers to who can receive care and when they need it. Cost represents the price tag of care and its affordability for patients and payers. The problem is that these elements often compete with each other in the healthcare sector. Healthcare optimization approaches can improve the Triangle of Health. Still, a groundbreaking breakthrough is needed to fully disrupt the Iron Triangle where Quality, Access, and Cost are optimized simultaneously. The healthcare industry holds vast amounts of valuable data, which has the potential to revolutionize the Triangle. Traditionally stored as text, there is a trend towards converting this data into more accessible formats (5).

Classification Models Based on Rule Sets

Identifying complex decision trees is challenging, especially when data is consistently presented within the tree. C4.5 is often referred to as a statistical classifier (12,13). C4.5 introduced a structure that clusters the directions of statements together, simplifying the identification process. Each statement's directions are grouped, and a case is classified based on the first direction that meets its criteria. If no direction is satisfied, it is directed to a specified class (12).

Decision Tree algorithm

Decision tree algorithms include CART, ID3, and C4.5, (12,13) which differ in their splitting methods, stopping criteria, and class assignment. CART utilizes the Gini index to measure split impurity, making it suitable for high-dimensional numerical data. Decision trees can handle continuous data but require modifications for categorical data (15-17).

Methods

This study collected data from UAE medical centers to create a predictive model in SPSS (11,14) for assisting in the diagnosis of diabetic patients. The model aims to support healthcare professionals in making informed decisions using reliable electronic sources. The study utilized the "IF conditions THEN conclusion" rule and the K-means algorithm for clustering. Data from 2856 patients in leading Dubai hospitals were collected.

IF and THEN

Regulations encompass two fundamental elements: the regulation initiator (IF section) that outlines certain conditions associated with predictor attributes' significance, and the regulation result (THEN section) that articulates an anticipated value for a target attribute (12). Possessing a precise expectation regarding the target attribute's value can enhance the decision-making process. IF-THEN expectation regulations find widespread application in data analytics, as they embody explicit knowledge at a profound level of comprehension (12). In healthcare, this concept can be employed as follows: by considering specific indicators and past medical history, one can deduce the underlying cause of a particular disease.

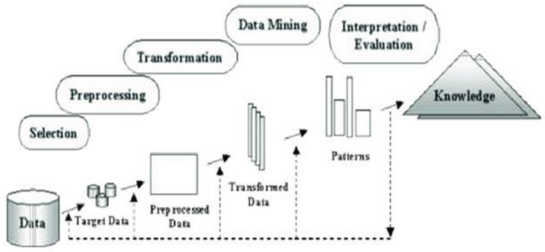
K-Mean Algorithm for Data Clustering

Algorithms aid in knowledge discovery by identifying attribute relationships and describing the nature of connections between them. Categorization and clustering are commonly utilized methods for gaining insights into this process. The categorized analysis involves supervised learning algorithms that examine pre-categorized datasets to generate classification rules. Conversely, clustering employs unsupervised learning algorithms to partition a dataset into cohesive groups or clusters. Clustering is a fundamental data mining technique across diverse domains like education and healthcare (18-20).

Gathering patient data and information from medical centers in person posed a significant challenge during the initial stages of this study. To overcome this obstacle, the following steps were taken:

• An official letter was issued to collect medical data. Regional hospitals didn't provide data, except for prime hospitals in Dubai. The researcher contacted a prime hospital, which requested data collection. The researcher then started the four-step data mining process, beginning with data cleaning (Figure 2).

Figure 2: The steps of data mining



- The initial form of the provided data was unsuitable for this research study as it contained negative values for age and weight and missing information on gender and genetics.

- The data was subsequently cleaned and organized to fit the requirements of this type of research. Table 1 represents the final form of the data after the cleaning process.

Results

After implementing the previous steps on the collected data, the results indicate a significant influence of all the independent variables (genetics, gender, and age) on the dependent variable (diabetes). Figure 3 depicts the resultant predictive model generated by a decision tree, which can aid in diagnosing whether an individual is likely to develop diabetes based on their genetic history, age, and gender. The conclusive outcomes establish the following regulations:

- If both parents have diabetes, then there are no significant statistical differences among the other variables.

- If a male is under 40 years old and has one parent with diabetes, then he has a 100% chance of having the same disease.

- If a female is under 40 years old and has one parent with diabetes, then she has a 51.8% chance of having the same disease.

- If a male is between 52 and 63 years old and has one parent with diabetes, then he has an 88.12% chance of having the same disease.

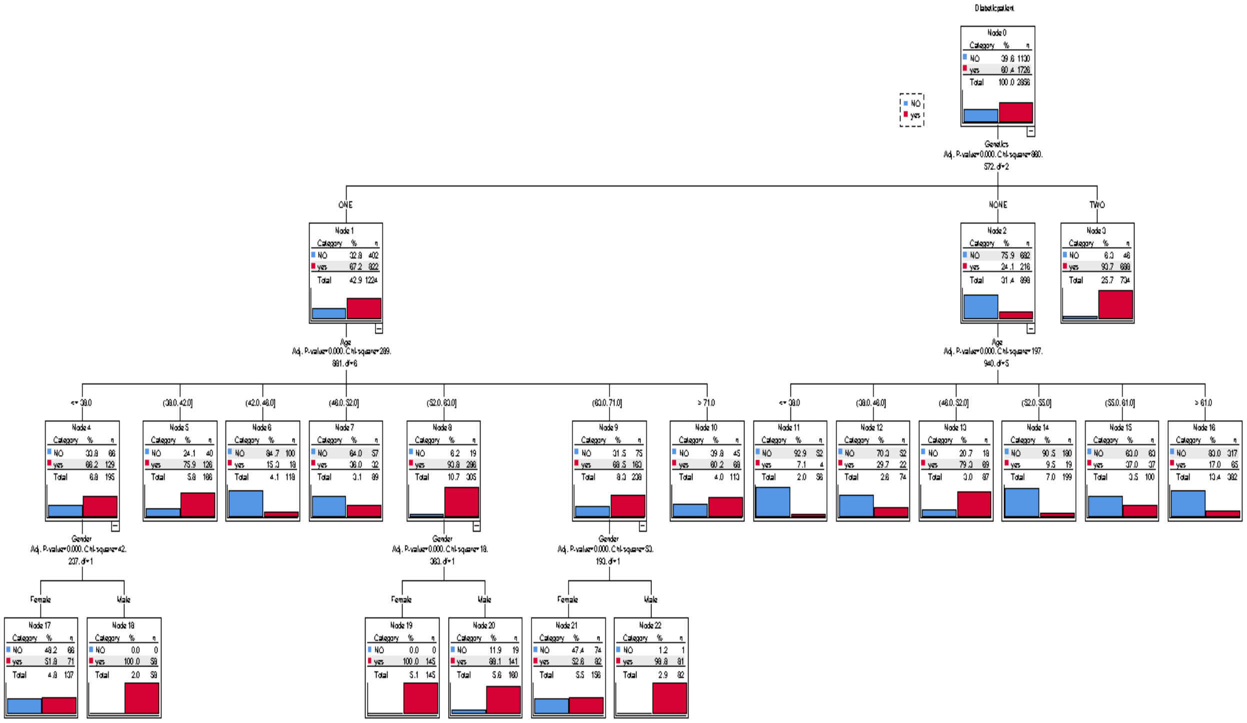
- If a female is between 52 and 63 years old and has one parent with diabetes, then she has a 100% chance of having the same disease.

- If a female is between 63 and 71 years old and has one parent with diabetes, then she has a 52.5% chance of having the same disease.

- If a male is between 63 and 71 years old and has one parent with diabetes, then he has a 98.7% chance of having the same disease.

Prior to developing the final model, which is illustrated in Figure 3, we utilized the entropy manual method to determine which independent variable to initiate the classification model. Consequently, the information gained from the independent variable "genetics" was found to be greater than that of the other independent variables. Therefore it was prioritized as the starting point for the predictive model, as shown in the decision tree (fig. 3).

Figure 3: Predictive model with decision tree algorithm



We identified relevant attributes for the healthcare predictive model and used SPSS No ??? for the interpretation.

After data cleaning, we obtained the final database with 2857individuals from the prime hospital in Dubai (Table 1).

Table 1. Dependent and independent variables.

		Gender	Age	Genetics	Diabetic patient
N	Valid	2857	2857	2857	2857
	Missing	0	0	0	0

Table 1 displays the independent variables, including age, gender, and genetics, alongside the dependent variable, which is the diabetic patient.

Table 2. Data classification

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NO	1131	39.7	39.7	39.7
	Yes	1727	60.5	60.5	100.1
	Total	2857	100.1	100.1	

Table 2 displays the dataset comprising2856 individuals, of which 1130 (39.6%) are non-diabetic, and 1726 (60.4%) are diabetic. The following tables illustrate the statistical associations between the dependent variable, diabetes, and the independent variables, genetics, age, and gender.

Table 3. The effect of genetics on diabetes.

Genetics

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NONE	899	32.5	32.5	32.3
	1	1225	41.8	41.8	75.4
	2	735	24.8	24.8	101.1
	Total	2857	101.1	101.1	

Table three displays the association between the initial independent variable (Genetics) and the dependent variable (Diabetes). With the IF-THEN rule applied, the following outcomes can be obtained:

-If both parents have diabetes, then 90.7% of the total number of individuals will have diabetes.

As can be observed from the table, the number of diabetic patients who have both parents diagnosed with diabetes is 734, representing 25.7% of the total sample size. Out of these, 699 individuals could be diabetic patients due to genetic reasons since both of their parents have diabetes.

-If both parents have diabetes, THEN 9.3% of the total number will not be diabetic patients.

Furthermore, it can be observed that there are 45 individuals in the sample who do not have diabetes despite having both of their parents with the disease.

Upon analyzing the data, it was discovered that 1224 individuals (42.9% of the total sample size) have one parent with diabetes. Of these individuals, 402 do not have diabetes while 822 have the disease. This implies that:

-IF one parent has diabetes, THEN the probability of having the same disease would be decreased for the next generations.

The final rule pertains to individuals whose parents are not diabetic patients. Out of the total sample size, 899 (32.4%) individuals were found to have non-diabetic parents. Among them, 682 (76.9%) do not have diabetes, whereas 217 (24.2%) are diabetic patients, indicating that:

-If both parents do not have diabetes, THEN the disease for the next generations would be rare.

We have deduced the following findings from our analysis using the IF conditions THEN conclusion rule:

-If both parents have diabetes, THEN there will not be any significant statistical differences between age and diabetes.

-To obtain accurate results for the first case where both parents are not diabetic patients, we applied the frequency method to clean and verify the age data provided. We then divided the data into six age intervals as follows:

For the first age interval (≥ 39), a frequency analysis was conducted to clean and organize the data. The interval had a total of 56 cases, which accounted for 2% of the entire dataset. Of those 56 cases, 52 (92.9%) did not have diabetes, while only 4 (7.1%) had the disease. This suggests that age may not be a significant factor in the development of diabetes for individuals in this age range.

•If the person's age is below 40 and both of their parents are not diabetic patients, then the likelihood of developing diabetes in the future is low with a percentage of 7.1%.

•If the person's age is below 40 and both of their parents are not diabetic patients, then the probability of not having diabetes in the future is high with a percentage of 92.9%.

The second age interval is [39, 46], and it includes 74 individuals, which is 2.6% of the total sample. Among them, 52 individuals do not have diabetes, which accounts for 70.3%, while 22 individuals have diabetes, which accounts for 29.7%. This implies:

•If the person is between 39 and 46 years old and has non-diabetic parents, the likelihood of having diabetes in the future is lower, with a percentage of 29.7%.

•If the person is between 39 and 46 years old and has non-diabetic parents, the likelihood of not having diabetes in the future is higher, with a percentage of 70.3%.

The third interval, which is [46, 52], includes 87 individuals, representing 3% of the total sample. Of these individuals, 18 do not have diabetes, accounting for 20.7% of the group, while 69 are diabetic, representing 79.3% of the group. This indicates that:

- IF the person is between 46 and 52 years old and has non-diabetic parents, THEN the likelihood of developing diabetes in the future would increase with a percentage of 79.3%.

-On the other hand, IF the person is in this age group and has non-diabetic parents, THEN the probability of not having diabetes in the future would decrease with a percentage of 20.7%.

The fourth interval, which is [52, 55], consists of 199 individuals, accounting for 7% of the total sample. Out of these, 19 individuals do not have diabetes while 190 individuals have the disease.

•If the individual is between 52 and 55 years old and neither of their parents have diabetes, then there is a high probability of having diabetes in the future with a percentage of 90.5%.

•If the individual is between 52 and 55 years old and neither of their parents have diabetes, then the probability of not having diabetes in the future is low with a percentage of 9.5%.

The fifth age interval is [55, 61]. There are 100 individuals in this group, which represents 3.5% of the total population. Out of these individuals, 63 do not have diabetes, while 37 have been diagnosed with the disease.

•If the individual is between 55 and 61 years old and neither of their parents are diabetic patients, then the likelihood of developing diabetes in the future would increase by 63%.

•If the individual is between 55 and 61 years old and neither of their parents are diabetic patients, then the likelihood of not developing diabetes in the future would decrease by 37%.

The sixth interval is for individuals older than 61 years. Out of the total population, 382 individuals with a percentage of 13.4% fall into this category. Among them, 317 individuals do not have diabetes, while 65 individuals have the disease.

•For individuals up to 61 years old and with non-diabetic parents, the likelihood of developing diabetes in the future would increase to 17%.

•For individuals up to 61 years old and with non-diabetic parents, the likelihood of not developing diabetes in the future would decrease to 83%.

•Age ≥ 39: Among 56 cases (2% of the dataset), 92.9% did not have diabetes and 7.1% had the disease. Age may not be a significant factor for diabetes in this age range.

•Age [39, 46] Out of 74 individuals (2.6% of the sample), 70.3% did not have diabetes and 29.7% had the disease. The likelihood of diabetes is lower in this age group.

•Age [46, 52]: Among 87 individuals (3% of the sample), 79.3% had diabetes and 20.7% did not. The likelihood of diabetes increases in this age range.

•Age [52, 55]: Out of 199 individuals (7% of the sample), 90.5% had diabetes and 9.5% did not. There is a high probability of having diabetes in this age group.

•Age [55, 61]: Among 100 individuals (3.5% of the population), 63% had diabetes and 37% did not. The likelihood of developing diabetes increases in this age range.

•Age > 61: Out of 382 individuals (13.4% of the population), 83% did not have diabetes and 17% had the disease. The likelihood of developing diabetes increases for individuals up to 61 years old.

In the second scenario, where only one parent has diabetes, we also classified age into five intervals as follows:

Interval 1: Age ≥ 39

Out of 195 cases, 6.8% were aged 39 or younger.

Among them, 66 cases (33.8%) didn't have diabetes, while 129 cases (66.2%) had diabetes.

Interval 2: Ages 39-42

166 individuals (5.8% of the total) fall within this range.

Among them, 40 individuals (24.1%) don't have diabetes, while 126 (75.9%) have it.

Interval 3: Ages 42-46

There are 118 cases (4.1% of the total).

Among them, 18 individuals (15.3%) don't have diabetes, while 100 (84.7%) have it.

Interval 4: Ages 46-52

There are 89 cases (3.1% of the total).

Among them, 57 individuals don't have diabetes (64%), while 32 (36%) have it.

Interval 5: Ages 52-63

There are 305 cases in this range.

Among them, 19 individuals (6.2%) don't have diabetes, while 296 (96.7%) have it.

•If an individual is between 52 and 63 years old and has one parent with diabetes, then there is a 97.04% probability of developing diabetes in the future.

•If an individual is between 52 and 63 years old and has one parent with diabetes, then the probability of not developing diabetes in the future is only 6.2%.

Table 4: The effect of the gender variable on diabetes.

Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Female	68	2.6	2.5	2.3
	Female	1361	47.7	47.7	50.1
	Male	1427	50.1	50.1	100.1
	Total	2857	100.1	100.1	

Table 5 illustrates the correlation between the predictor variable (gender) and the response variable (diabetes). Based on our data set, we obtained the following findings:

•If both parents are diabetic patients, there is no significant association between the gender of the patient and the occurrence of the diabetic disease.

•If one parent is a diabetic patient, there are significant associations between the gender and age group of the patient and the occurrence of the diabetic disease.

The first interval, which includes individuals aged ≤ 39, yielded 58 male patients with diabetic disease in our sample. Among the 137 female individuals in this age group, 66 were not diabetic patients, while 71 had the disease. This result confirms the following pattern:

•If a male is under 40 years old and has one parent who is a diabetic patient, then he is certain to have the disease.

•If a female is under 40 years old and has one parent who is a diabetic patient, then there is a 51.8% chance that she will have the disease.

Table 5 demonstrates the correlation between gender and diabetes. Our findings are as follows:

Both parents' diabetic: No significant association between gender and diabetes.

One parent diabetic: Significant associations between gender, age group, and diabetes.

Interval 1: Age ≤ 39

58 male patients with diabetes.

Among 137 females, 66 were not diabetic, while 71 had the disease.

Patterns:

Male under 40 with one diabetic parent always has the disease.

Female under 40 with one diabetic parent has a 51.8% chance of having the disease.

The second interval of the age, [age 52-63], includes 160 males in our sample. Among them, 19 do not have diabetes while 141 have the disease. Additionally, 145 females are in the same age range and all of them have diabetes. This implies that:

•If a male is between 52 and 63 years old and has one diabetic parent, the probability of having diabetes in the future is 88.1%.

•If a female is between 52 and 63 years old and has one diabetic parent, the probability of having diabetes in the future is 100%.

The third age interval is [63, 71]. Among the participants in our sample, there were 82 males between the ages of 63 and 71, with only one not having the diabetic disease and the remaining 81 being diabetic patients. Among the 156 females in the same age interval, 74 do not have the disease while 82 are diabetic patients. This suggests that:

•For males aged between 63 and 71 years with one diabetic parent, 81 out of 82 have the disease, which means the probability of having diabetes is 98.7%.

•For females aged between 63 and 71 years with one diabetic parent, 82 out of 156 have the disease, which means the probability of having diabetes is 52.5%.

Application of K-means algorithm for clustering

The study utilizes the K-means clustering analysis algorithm to perform vector quantization and cluster the large groups of variables, as stated by Jain (2020). The age-independent variable is categorized into four clusters with mean ages of 31, 68, 49, and 88. The results indicate that individuals in the first cluster are less vulnerable to diabetes, while those in the fourth cluster are more susceptible to the disease. Please refer to table 5 for details.

Table 5: The initial cluster centers.

Initial Cluster Centers

	Cluster			
	1	2	3	4
Age	32	67	48	89

a.Convergence achieved due to no or small change in cluster centers. The maximum absolute coordinate change for any center is .000. The current iteration is 6. The minimum distance between initial centers is 18.000 (see table6)

Table 6: Iteration History

Iteration History

	Change in Cluster Centers			
Iteration	1	2	3	4
1	4.411	2.094	.445	8.608
2	.828	1.592	.628	3.363
3	.752	.525	.663	1.231
4	.597	.593	.581	1.071
5	.596	.203	.614	.347
6	.000	.000	.000	.000

-The subsequent procedure involved taking a subset of our data and calculating the distance between each age and the cluster center, as shown in (See Appendix 11)Table 8, using the K-means algorithm. Table 8 shows that the first case has a distance of 7.183 points from the cluster center. Similarly, case number 10 has a distance of 7.138 points from the cluster center. Tables 10, 11, and 12 provide two additional sets of samples where the distances between each case and its respective cluster center are directly recorded.

The last step involves creating new clusters, which are presented in table13. The table indicates changes in the mean values, where the initial cluster had a mean of 31, while it increased to 38 in the final clustering.

Table 12: The final cluster centers.

Final Cluster Centers

	Cluster			
	1	2	3	4
Age	39	64	52	74

The distances between the final cluster centers are presented in Table 13. It can be observed that the first and second clusters are 24.811 units apart, the distance between the first and third clusters is 12.857 units, the third and fourth clusters are separated by

22.342 units, and the first and fourth clusters are 35.199 units apart, indicating the variation between the initial and final clusters.

Table 13: Final Cluster Centers Distances.

Final Cluster Centers Distances

Cluster	1	2	3	4
1		24.812	12.858	35.198
2	24.811		11.955	10.388
3	12.857	11.955		22.342
4	35.199	10.388	22.342	

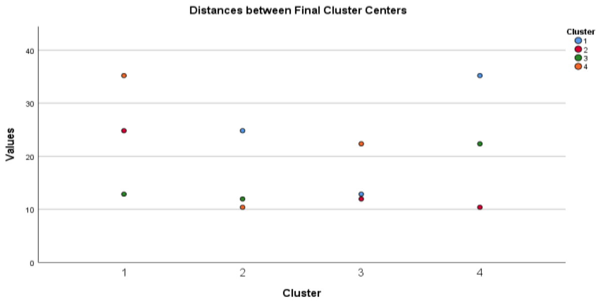
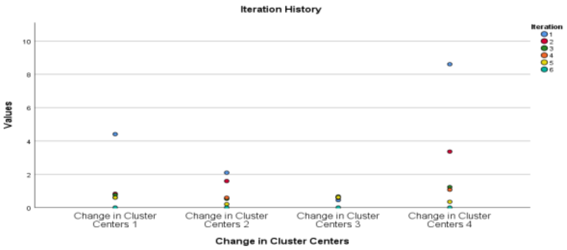


Table 14: Cases in each cluster

Number of Cases in each Cluster

Cluster	1	742.001
	2	942.001
	3	722.001
	4	450.001
Valid		2856.001
Missing		.001



The analysis of the provided data reveals several important insights regarding the relationship between the study variables: genetics, age, and diabetes.

Genetics: The findings indicate a strong association between genetics and diabetes. Specifically, if both parents have diabetes, there is a high probability that the individual will also have diabetes. This suggests a hereditary component to the disease. However, it's worth noting that there are cases where individuals do not develop diabetes despite having both parents with the disease. This implies that other factors, such as environmental or lifestyle factors, may also play a role in the development of diabetes. Further research is needed to explore these additional factors and their interactions with genetic predisposition.

Age: The analysis highlights the influence of age on the likelihood of developing diabetes.

For individuals with non-diabetic parents, the probability of having diabetes generally increases with age. This aligns with the common understanding that age is a risk factor for diabetes. However, the data also reveals variations in the likelihood of diabetes depending on specific age intervals. For instance, individuals aged 39 or younger, regardless of parental diabetes status, have a relatively low likelihood of having diabetes. This suggests that younger age groups are less prone to the disease. On the other hand, individuals between the ages of 52 and 63 with one parent having diabetes have a significantly higher probability of developing diabetes. This finding underscores the importance of considering both age and parental diabetes status when assessing an individual's risk for the disease.

Gender: The provided data does not include a separate analysis of gender and its relationship with diabetes. As a result, no specific conclusions or insights can be drawn regarding the influence of gender on diabetes based on the given information. It is possible that gender was not a variable of interest in this particular study or that data regarding gender was not collected or included in the dataset. Further investigation or additional data would be required to assess the potential impact of gender on diabetes risk in this context.

Overall, the results emphasize the significant role of genetics and age in determining the likelihood of developing diabetes. The findings also suggest that there may be complex interactions between these factors and other potential risk factors for diabetes, highlighting the need for further research to gain a comprehensive understanding of the disease's etiology.

Study limitations

- Sample size: The paper's findings are based on data collected from a specific set of medical centers in the UAE. The limited sample size may not fully represent the diverse population and healthcare settings in the country.

- Generalizability: Due to the focus on UAE healthcare sector and specific patient attributes, the generalizability of the predictive model to other regions or populations may be limited.

- Data quality: The accuracy and completeness of the medical center data used in the analysis may have inherent limitations.

Inaccurate or missing data points can affect the reliability of the predictive model.

- Causality vs. correlation: While the paper identifies risk factors associated with diabetes, it is important to note that the analysis establishes correlations rather than causation. Other unmeasured factors may also contribute to the development of diabetes.

Conclusion

To summarize, there have been numerous attempts to create predictive models in the healthcare sector, specifically targeting the diagnostic phase of various diseases, including diabetes, which has been a focus due to the large volume of data it generates. Despite the considerable effort invested, a significant number of these models have not been implemented in practice. This has resulted in an increase in the amount of data produced, which can be challenging to manage without effective tools for analysis and decision-making.

However, despite the current lack of implementation of predictive models in healthcare, many experts (ref???) believe that such models represent a promising way to revolutionize the use of technology in the field. These models can be utilized as a reliable source of electronic information to assist hospitals in making informed decisions about their patients and guided planning of resources. By analyzing and processing large amounts of data, predictive models can provide valuable insights into patients' health and well-being, allowing healthcare providers to deliver more personalized and effective care.

In conclusion, although there is a long way to go in terms of fully realizing the potential of predictive models in healthcare, they represent an exciting avenue for exploration and innovation. With further research and development, these models have the potential to transform the way healthcare is delivered, enabling more efficient and effective decision-making that can ultimately lead to improved patient outcomes.

References

1. Thangarasu G, Subramanian K. Big data analytics for improved care delivery in the healthcare industry. International journal of online and biomedical engineering. 2019;
2. Dirienzo G. Informatics and the clinical lab: Present and future. Biochim Clin. 2020;

3. Ravikumar R, Kitana A, Taamneh A, Aburayya A, Shwedeh F, Salloum S, et al. The Impact of Big Data Quality Analytics on Knowledge Management in Healthcare Institutions: Lessons Learned from Big ' 'Data's Application within The Healthcare Sector. South Eastern European Journal of Public Health (SEEJPH) [Internet]. 2022 Dec 22 [cited 2023 Apr 27]; Available from: https://www.biejournals.de/index.php/seejph/article/view/6194
4. Rejitha Ravikumar AKATAAFSSSKS. Impact of knowledge sharing on knowledge Acquisition among Higher Education Employees. Computer Integrated Manufacturing Systems [Internet]. 2022 Dec 9 [cited 2023 Apr 27];28(12):827–45. Available from: http://cims-journal.com/index.php/CN/article/view/462
5. Wang YC, McPherson K, Marsh T, Gortmaker SL, Brown M. Health and economic burden of the projected obesity trends in the USA and the UK. The Lancet [Internet]. 2011 Aug 27 [cited 2023 Apr 27];378(9793):815–25. Available from: http://www.thelancet.com/article/S0140673611608143/fulltext
6. Shwedeh F, Hami N, Zakiah S, Baker A. Effect of Leadership Style on Policy Timeliness and Performance of Smart City in Dubai: A Review.
7. Shwedeh F, Hami N, Bakar SZA, Yamin FM, Anuar A. The Relationship between Technology Readiness and Smart City Performance in Dubai. Journal of Advanced Research in Applied Sciences and Engineering Technology [Internet]. 2022 Dec 23 [cited 2023 Apr 27];29(1):1–12. Available from: https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/article/view/996
8. Dragneva R, Wolczuk K. Russia, the Eurasian Customs Union and the EU: Cooperation, Stagnation or Rivalry? SSRN Electronic Journal [Internet]. 2012 Aug 6 [cited 2023 Apr 27]; Available from: https://papers.ssrn.com/abstract=2125913
9. Shwedeh F, Adelaja AA, Ogbolu G, Kitana A, Taamneh A, Aburayya A, et al. Entrepreneurial innovation among international students in the UAE: Differential role of entrepreneurial education using SEM analysis. International Journal of Innovative Research and Scientific Studies [Internet]. 2023 [cited 2023 Apr 27];6(2):266–80. Available from: https://ideas.repec.org/a/aac/ijirss/v6y2023i2p266-280id1328.html
10. Aburayya A, Salloum SA, Alderbashi KY, Shwedeh F, Shaalan Y, Alfaisal R, et al. SEM-machine learning-based model for perusing the adoption of metaverse in higher education in UAE. International Journal of Data and Network Science. 2023 Mar 1;7(2):667–76.
11. Shwedeh F, Aburayya A, Alfaisal R, Adelaja AA, Ogbolu G, Aldhuhoori A, et al. SMEs&rsquo; Innovativeness and Technology Adoption as Downsizing Strategies during COVID-19: The Moderating Role of Financial Sustainability in the Tourism Industry Using Structural Equation Modelling. Sustainability 2022, Vol 14, Page 16044 [Internet]. 2022 Dec 1 [cited 2023 Apr 27];14(23):16044. Available from: https://www.mdpi.com/2071-1050/14/23/16044/htm

12. Lavrač N. Selected techniques for data mining in medicine. Artif Intell Med [Internet]. 1999 May [cited 2023 Apr 27];16(1):3–23. Available from: https://pubmed.ncbi.nlm.nih.gov/10225344/
13. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health Inf Sci Syst [Internet]. 2014 Feb 7 [cited 2023 Apr 27];2(1). Available from: /pmc/articles/PMC4341817/
14. Dahu BM, Aburayya A, Shameem B, Shwedeh F, Alawadhi M, Aljasmi S, et al. The Impact of COVID-19 Lockdowns on Air Quality: A Systematic Review Study. South East Eur J Public Health [Internet]. 2023 Jan 24 [cited 2023 Apr 27]; Available from: https://seejph.com/index.php/seejph/article/view/312
15. Abdullah El Nokiti KSSSAAFS& BS. Is Blockchain the answer? A qualitative Study on how Blockchain Technology Could be used in the Education Sector to Improve the Quality of Education Services and the Overall Student Experience. Computer Integrated Manufacturing Systems [Internet]. 2022 Nov 14 [cited 2023 Apr 27];28(11):543–56. Available from: http://cims-journal.com/index.php/CN/article/view/237
16. Wong TY, Rosamond W, Chang PP, Couper DJ, Sharrett AR, Hubbard LD, et al. Retinopathy and risk of congestive heart failure. JAMA [Internet]. 2005 Jan 5 [cited 2023 Apr 27];293(1):63–9. Available from: https://pubmed.ncbi.nlm.nih.gov/15632337/
17. ``Sherief Abdallah BAAAENSSSAAA& FS. A COVID19 Quality Prediction Model based on IBM Watson Machine Learning and Artificial Intelligence Experiment. Computer Integrated Manufacturing Systems [Internet]. 2022 Nov 14 [cited 2023 Apr 26];28(11):499–518. Available from: http://cims-journal.com/index.php/CN/article/view/235
18. Ahmad A, Dey L. A k-mean clustering algorithm for mixed numeric and categorical data. Data Knowl Eng. 2007 Nov 1;63(2):503–27.
19. Mohammad Salameh ATAKAAFSSSKSDV. The Impact of Project Management ' 'Office's Role on Knowledge Management: A Systematic Review Study. Computer Integrated Manufacturing Systems [Internet]. 2022 Dec 9 [cited 2023 Apr 27];28(12):846–63. Available from: http://cims-journal.com/index.php/CN/article/view/463
20. Salloum S, Al Marzouqi A, Alderbashi KY, Shwedeh F, Aburayya A, Rasol M, et al. Sustainability Model for the Continuous Intention to Use Metaverse Technology in Higher Education: A Case Study from Oman. Sustainability 2023, Vol 15, Page 5257 [Internet]. 2023 Mar 16 [cited 2023 Apr 27];15(6):5257. Available from: https://www.mdpi.com/2071-1050/15/6/5257/htm

© 2023 S Khadragy et al.; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.