International Advance Journal of Engineering Research (IAJER)
Volume 4, Issue 06 (June- 2021), PP 10-21
ISSN: 2360-819X
www.iajer.com

Research Paper **Open Access**

Diabetes Mellitus of Type-1 & Type-2 Prediction and Detectionusing Data Miningand WarehousingTechniques

Deepa. S[1], Dr. B. Booba[2]

[1]Research Scholar (**UP19P9611053**), Department of Computer Science, Vels Institute of Science, Technology & Advanced Studies (VISTAS), Pallavaram,Chennai-600117, India

[2]Professor, Department of Information Technology, Vels Institute of Science, Technology & Advanced Studies (VISTAS), Pallavaram, Chennai-600117, India

Abstract: Diabetes is one of the most significant health issues which are faced by maximum number of human being now a day. Diabetes is a chronic, metabolic disease characterized by elevated levels of blood glucose (or blood sugar), which leads over time to serious damage to the heart, blood vessels, eyes, kidneys and nerves.There are three main types of diabetes: **Type 1 diabetes** is caused by an autoimmune reaction in which the body's immune system attacks the insulin-producing beta cells of the pancreas. As a result, the body produces very little or no insulin. **Type 2 diabetes** is the most common type of diabetes. Initially, hyperglycaemia (high blood glucose levels) is the result of the inability of the body's cells to respond fully to insulin, a situation termed 'insulin resistance'. **Gestational diabetes** (GDM) is characterised by high blood glucose levels during pregnancy. It may occur at any time during pregnancy (although most likely after week 24) and usually disappears after the pregnancy. But the present study is conducted for diabetes mellitus of Type I and Type II and its prediction and detection using data mining and data warehousing techniques. The study uses various data mining and warehousing techniques to predict and detect the diabetes mellitus. So that a multidimensional diabetes data warehouse has been built to store and access the general as well as medical records of diabetes patients. Along with that a data mining model has also been proposed and implemented with this diabetes data warehouse to predict the diabetes mellitus among patients and the risk factor for a particular type i.e., Type I and Type II diabetes and also the exact method to diagnose it.In this study,OLAP operations are performed to segregate the data based on several attributes like gender, type of diabetes, localities, etc.The k-means clustering algorithm is used for partitioning the data into diabetes and non-diabetes clusters.Weighted Average Method (WAM) is used to improve the accuracy of analytic predictive performance models for diabetes prevention systems with more number of new patients.

Keywords: Architectural Data Warehouse**,** Diabetes Mellitus, Data Mining and Warehousing Techniques, Multidimensional Diabetes Data Warehouse, MultidimensionalStar Schema, k- means Algorithm, OLAP Operations, Types of Diabetes- Type-I & Type-II,WAM.

I. Introduction

In 2000, the global estimate of adults living with diabetes was 151 million. By 2009 it had grown by 88% to 285 million. Today, we calculate that 9.3% of adults aged 20–79 years – a staggering 463 million people – are living with diabetes**.** A further 1.1 million children and adolescents under the age of 20, live with type 1 diabetes.

A decade ago, in 2010, the global projection for diabetes in 2025 was 438 million andWorld Health Organization (WHO) estimates that diabetes was the seventh leading cause of death in 2016. With over five years still to go, that prediction has already been surpassed by 25 million.

International Diabetes Federation (IDF) estimates that there will be 578 million adults with diabetes by 2030, and 700 million by 2045.The global diabetes prevalence in 2019 is estimated to be 9.3% (463 million people), rising to 10.2% (578 million) by 2030 and 10.9% (700 million) by 2045. The prevalence is higher in urban (10.8%) than rural (7.2%) areas, and in high-income (10.4%) than low-income countries (4.0%).

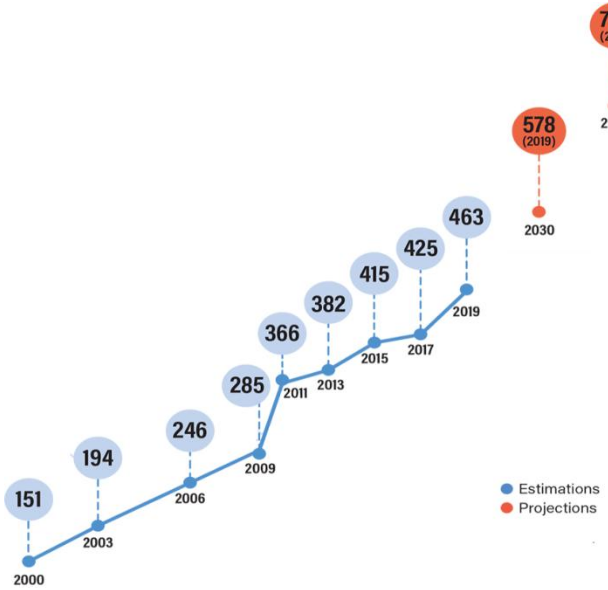


Figure 1: Estimated no. of adult diabetes (in millions)

India is home to 69.1 million people with DM and is estimated to have the second highest number of cases of DM in the world after China in 2015. The prevalence of DM in India ranges from 5–17%, with higher levels found in the southern part of the country and in urban areas. DM continues to increase as a result of rapid cultural and social changes, which include: ageing populations, increasing urbanization, dietary changes, reduced physical activity and unhealthy behavior.

Each year 7 million people develop Diabetes and the most dramatic increases in type 2 Diabetes have occurred in populations where there have been rapid and major changes in lifestyle, demonstrating the important role played by lifestyle factors and the potential for reversing the global epidemic. A person with type 2 diabetes is 2 – 4 times more likely to get cardiovascular disease, and 80% of people with Diabetes will die from it.

As an alternative to the tedious physical storage of resources it is more important to build a data warehouse specific to diabetes disease of type 1 and type 2 and a data mining model to predict these types of diabetes earlier.A data warehouse is a powerful repository of large, multidimensional data cubes which answers complex queries and helps in decision making. If a machine learning technique is used to do these tasks then the physicians can directly start treatment without wasting much time in different kinds of diagnosis. There are several data mining and warehousing techniques in health care sector with respect to type 1 and type 2 of diabetes mellitus.

A data warehouse is usually built on OLAP technology which supports data integration, analyses and decision making, while the OLTP database is used to update and process individual records of a patient on a day-to-day basis. A data warehouse that combines these two techniques to make decision individually by answering complex queries as well as with a help of data mining model can be of immense help in health care to predict life-threatening disease such as diabetes in advance.

The present study is carried out to predict and detect diabetes mellitus of type 1 and type 2 category using several data mining and data warehousing techniques. For that the data will be collected from Andaman & Nicobar Islands Institute of Medical Science (ANIIMS) which is the only government tertiary care hospital in Andaman and Nicobar Islands. All the patients diagnosed or suspected with diabetes would be registered in G.B. Pant Hospital.

II. Aims and Objectives of the Research

The aim of the study is to build a multidimensional diabetes architectural data warehouse to store and process the diabetes related data. It is done by building OLAP and OLTP technologies for fetching useful and necessary information using query engines. And also data mining model has been built with these data warehouse to predict a person's predisposition towards diabetes and risk level for a particular type of diabetes such as Type 1 and Type 2 and also the exact method to diagnosis it.

The objective is to develop a method for the multidimensional diabetes data warehouse to extract the data from the different database and a user interface transform them into multidimensional data cubes and upload itself automatically for each new entry. And to enhance the diabetes data warehouse such that it supports OLTP and progress multidimensional diabetes data warehouse OLAP consecutively and produce its results individually.

Then the next objective is to develop a data mining model to predict a person's risk level of diabetes disease, the type of diabetes and to suggest an exact clinical diagnostic method using classification clustering and prediction techniques. And also to generate the data mining model results in the knowledge base and compare each new case with the already existing similar cases in the data warehouse before predicting results, thereby making it intelligent.

III. Review of RelatedLiterature

The combination of data warehousing and data mining technologies in the healthcare field proves to be a valuable tool to researchers and patients alike. This helps to prevent, diagnose, and medicate diseases using the medical database. There are several studies have been conducted on diabetes mellitus in Abroad as well as in India.

From this literature review, it has been found that the use of IT systems in medical sciences to mitigate, diagnose and prevent such life threatening diseases like diabetes mellitus has garnered the attention of IT researchers worldwide.

The review of related literature on diabetes related databases helped us to realize the fact that suitable data warehouse architecture should be implemented for the efficient functioning of the objective data analysis.

IV. Methodology of Research

Due to Advanced Technology, increasing number of hospitals are using electronic medical records to accumulate substantial amount of data of their patients with the associated clinical conditions and treatment details.They capture clinical data, store in personal database as well as mirror it in local and regional database. The present study uses the following methodologies to predict and detect the two types of diabetes mellitus.

A. Building Architectural Data Warehouse for Diabetes Disease

Data Warehouse is a phenomenon that grew from the huge amount of electronic data stored in recent years and for emergency need the data is used to accomplish goals that go beyond the routine tasks linked to daily processing**.**

Detection and prediction system is very essential in a Data warehouse process in the scenario of a big chain of multispecialty or diabetes institute which have many branches,also patient admin managers need to quantify and evaluate how eachbranch can update theirpatientdetails to the central hospital. The group database stores detailed patient data on the task performed by branches.

A data warehouse is the storage of convenient, consistent, complete and consolidated data, which is collected for the purpose of making quick analysis for the end users who take part in Decision Support Systems (DSS).These data is obtained from different operational sources and kept in separate physical store.

A data warehouse is a collection of data that supports decision-making processes. It provides the features like Subject-oriented, Integrated and consistent and Shows its evolution over time and it is not volatile.In addition, the data warehouse involves data accessible and process oriented features as well.

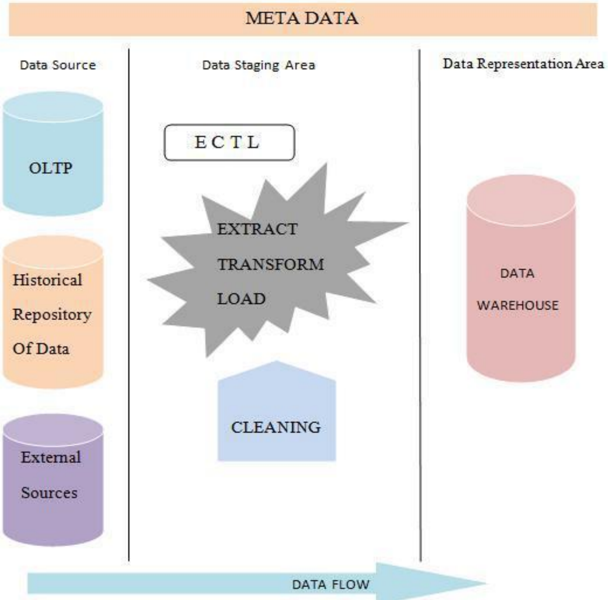


Figure 2: Data Flow diagram of Data Warehouse

It is well known that the Data Warehouse is very important in the large chain of organization level to maintain the massive data records/information. So it is very necessary to implement a data warehouse system in health care sectors.

In Data Warehouse the data comes from operational systems and external sources. To create the data warehouse, diabetes data are extracted from source systems like questionnaire, diabetes institute database, etc… and then it is cleaned (e.g., to detect and correct errors) and then transformed (e.g., put into subject groups or summarized), and finally loaded into a data store (i.e., placed into a data warehouse).

It includes tools for extracting data from multiple operational database and external sources. In addition to the main warehouse, there may be several departmental data marts.

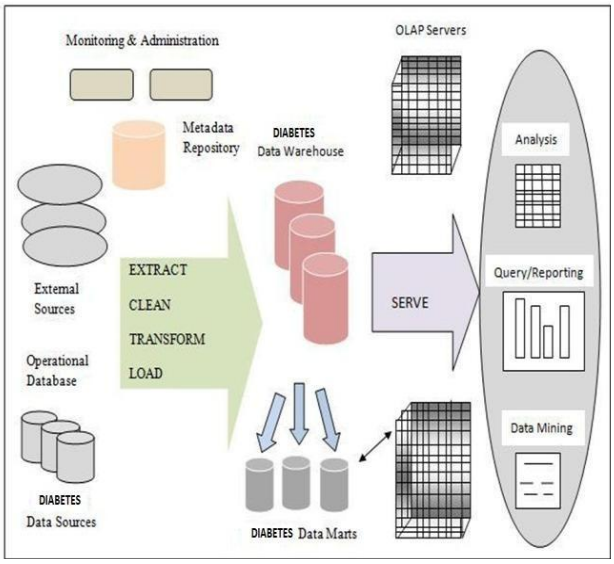


Figure 3: Diabetes Data Warehousing Architecture

Data in the warehouse and data marts are stored and managed by one or more warehouse servers, which present multidimensional views of data to a variety of front end tools: query tools, report writers, analysis tools, and data mining tools. Finally, there is a repository for storing and managing metadata, tools for monitoring and administering the warehousing system. Data warehousing technologies have been successfully deployed in healthcare.

B. Multidimensional Diabetes Data Warehouse

The multidimensional data model is based on the key concepts such as cube, dimension and hierarchy. This model allows its users to view and extract results from the cancer data warehouse in several different ways.

The OLAP structure comprises cubes, dimensions, measures, hierarchies and levels. A data cube is a multidimensional data storage unit. Habits, risk factors and symptoms are few dimensions of diabetes data warehouse. Measures are the facts that are to be analyzed. This diabetes data warehouse contains a Fact table named Medical that has fields for Patient_id as primary key, Habit_id, Risk_factor_id, Symptom_id, Blood group_id Diabetes_Id as foreign keys representing other dimensions and Treatment, Diagnostic status, Districts and Date of Entry as measures. If a medical analyst generally analyzes the quantity of drugs required for a month, the Diagnostic status of patients in a month will be the measure and the diabetes cube will be the dimension.

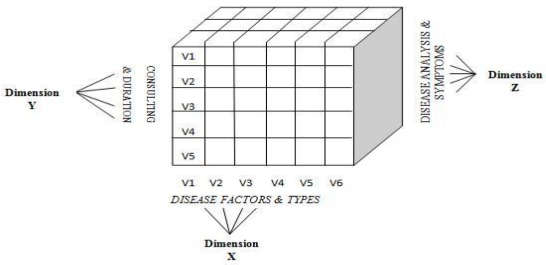


Figure 4: Multidimensional Diabetes Data Warehouse with dimension of X, Y, Z

Multidimensional analysis has become a popular way to extend the capabilities of query and reporting. That is, rather than submitting multiple queries, data is structured to enable fast and easy access. Dimensions can have individual entities or a hierarchy of entities, such as region, storage and sector. Multidimensional analysis enables users to look at a large number of interdependent factors involved in a medical problem and to view the data in complex relationships.

C. Multidimensional Star Schema

A Star Schema is a schema Architectural structure used for creation and implementation of the Data Warehouse systems, where there is only one fact table and multiple dimension tables connected to it. It is structured like a star in shape of appearance. This is one of the efficient data warehouse schema types, which can use simple querying for accessing the data from the system, in order to derive logical contents for analytical and report generation purposes.

It is a relational schema where a relational schema whose design represents a multidimensional data model. It is the explicit data warehouse schema. It is also known as star schema because the entity-relationship diagram of this schema simulates a star, with points, diverge from a central table. The center of the schema consists of a large fact table, and the points of the star are the dimension tables.

In this study the fact table contains measurements (e.g. Patient History, Risk Factor, Type of Diabetes, Symptoms, Treatment, and Diagnosis) which may be aggregated in various ways. The dimension tables provide the basis for aggregating the measurements in the fact table. The fact table is linked to all the dimension tables by one-to-many relationships. The primary key of the fact table is the concatenation of the primary keys of all the dimension tables.

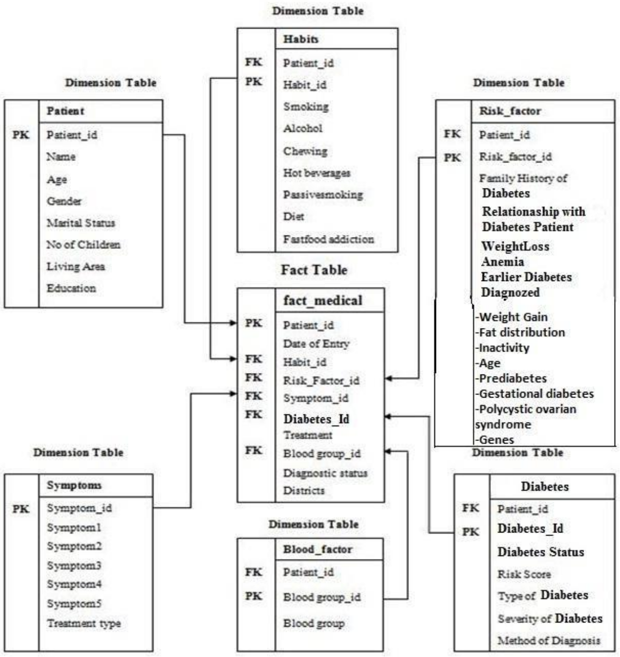


Figure 5: Star Schema representing Diabetes Data Warehouse

Fact Table

The Fact table that describes the subject matter is named fact_medical. The table consists of Foreign Keys (Patient_id, Date of Entry, Habit_id, Risk_factor_id, Symptoms_id, Diabetes_Id, Treatment, Blood group_id, Diagnose status and Districts) that related with dimensional table and Measures (Types of Diabetes, Severity of Diabetes, and Method of diagnosis).

Dimension Tables

There are six dimension tables that detailed each entity in the Patient table, Habits table, Risk_factor table, Symptoms table, Blood_factor table and Diabetes table.

D. OLAP operations of Diabetes Data Warehouse

OLAP operations are performed on diabetes data warehouse and data marts. It provides a user friendly environment for interactive data analysis. Following for OLAP operations have been performed to fetch and segregate the collected diabetes dataset.

1. Roll-Up (Drill-Up)
2. Drill Down
3. Slice and Dice and

4. Pivot

Roll-Up (Drill-Up)
The roll-up operation performs aggregation on a data cube either by climbing up the hierarchy or by dimension reduction.
It does not remove any data but changes the level of granularity of a particular dimension.
It is based on location which is equivalent to grouping the data by districts like South Andaman District, North & Middle Andaman District and Nicobar District.
And also it is equivalent to grouping the data based on Areas like Rural and Urban Area, Gender, etc.

For Example:
Rolling up (sometimes to as drilling up) is the reverse. For example, types of disease symptoms {types} = {{symptoms 1}, {symptoms 2}, {symptoms 3}, {symptoms 4}, {symptoms 5}}; is rolled up into symptoms '{types} = {{symptoms 1, symptoms 2, symptoms 3, symptoms 4, symptoms 5}} in dimension Z.
Example of Dimension Y Gender {types} = {{Male}, {Female}, {others}} is rolled up into gender. Gender {types} = {{Male, Female, others}}.

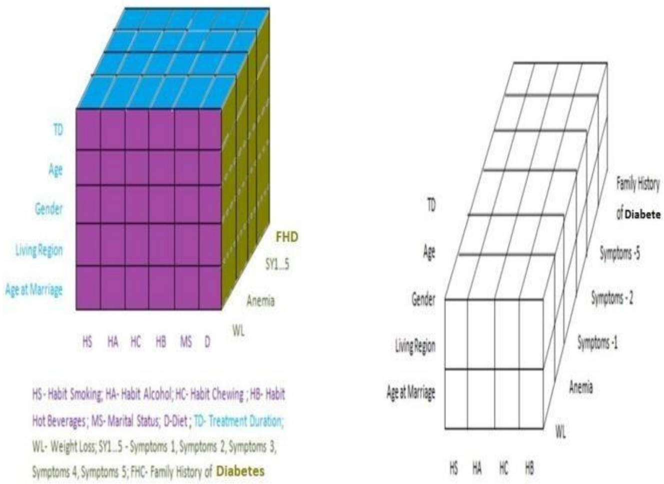


Figure 6(a): Originalview **Fig 6(b):** Roll-Up View

Drill-Down (Roll Down)
Drill-down is the reverse of roll-up operation. That means lower level summary to higher level summary. It can be performed either by stepping down a concept hierarchy for a dimension or by introducing a new dimension.This operation does not remove any events but change the level of granularity of a particular dimension.

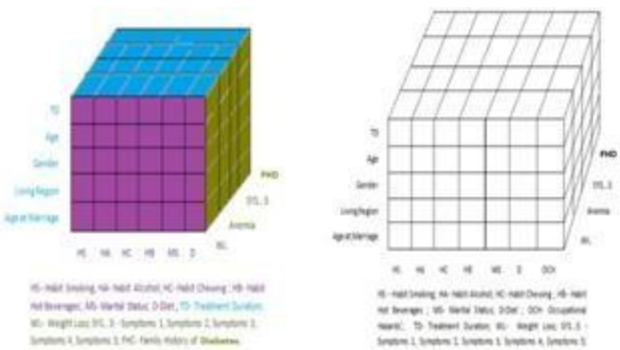


Figure 7 (a): Original view **Fig7(b):** Drill-Down View

Slice and Dice
The slice operation performs a selection on one dimension of the given cube, resulting in asub-cube. It reduces the dimensionality of the cubes. It produces a sliced OLAP cube by allowing the analyst to pick specific value for one of the dimensions.

For Example:

The Slicing is performed in the Diabetes Data warehouse The dimension Y Age, thus the Age dimension is removed the original cube and only Age of Dimension Y, remaining dimension of X Habits like (Smoking, Chewing, Alcohol, Hot beverage), dimension of Z (Weight loss, Anemia, Symptoms 1…5, Family History of Cancer are same. It implies the removing of the Age dimension and only considers the Age factor related to Habits, disease symptoms.

The dice operation defines a sub-cube by performing a selection on two or more dimensions. It produces a sub-cube by performing the system to pick specific values for multiple dimensions.

For example, one could dice the Treatment/Duration “2017” & “2018”, Diabetes types wise ‘Type 1’, ‘Type 2’, Gender wise ‘Male’, ‘Female’ and disease factor of “Anemia”. Here, no dimensions are removed, but only Treatment / Duration in 2017 and 2018 as Gender wise data warehouse dicing process is analyzed the Z-dimension disease factor “Anemia” and/or “Weight Loss” are considered.

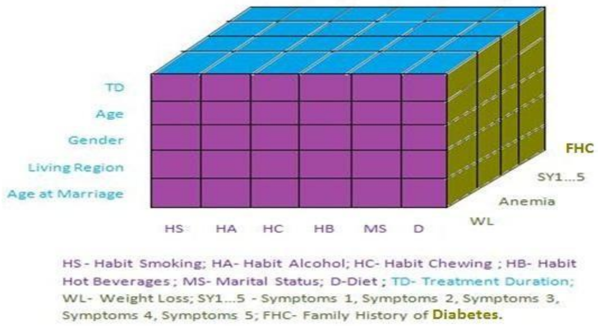
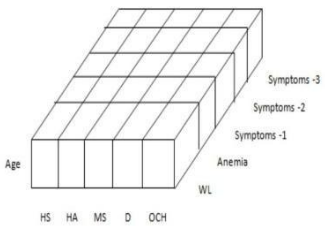


Figure 8(a): Original view of Diabetes Data Warehouse



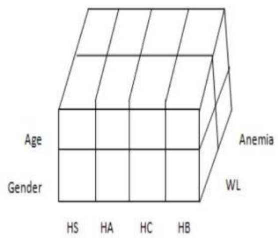


Figure 8(b): Slice View **Figure 8(c):**Dice View

Pivot

Pivot is also known as rotate.This operation rotates the data axes in view in order to provide an alternative presentation or different perspectives of the data. It removes a measure.

For Example:

Data for Marital Status, Diet, and Habit of Hot Beverage is obtained from the input cube and querying is allowed only for specific measures. We evaluate Projection by removing a measure from the sub cube query.

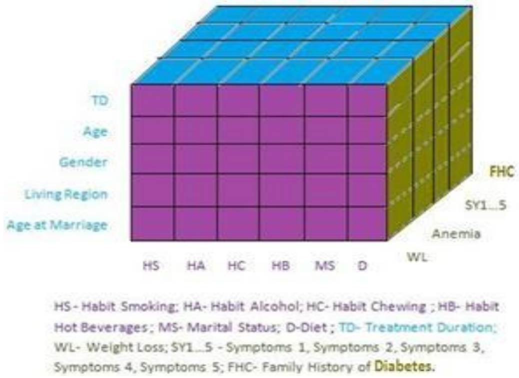


Figure 9(a): Original view of Diabetes Data Warehouse

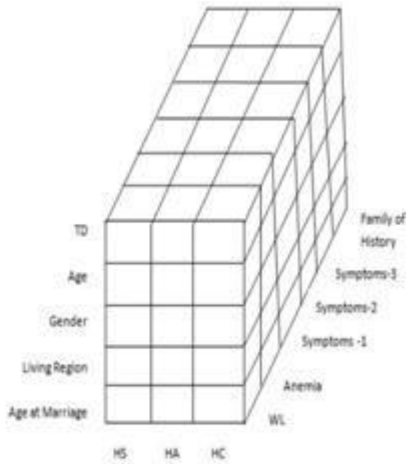


Figure 9**(b)**: **Pivot view**

The present study uses1500 raw samples of diabetes data which are shown in following table.

Table 1: Site wise distribution of Diabetes

S. No.	Type of Diabetes	Male	Female	Total
1.	Type - I	375	375	750
2.	Type - II	375	375	750

Graph 1: Site wise diabetes distribution

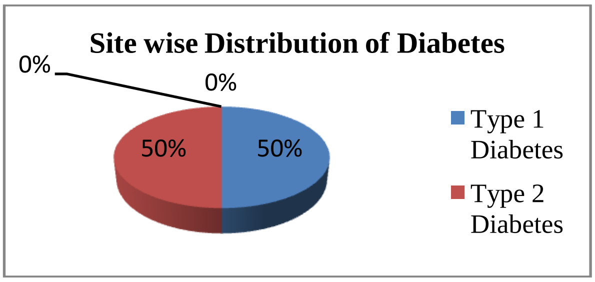


Table 2 Correlation of Age and Sex of Diabetes patient

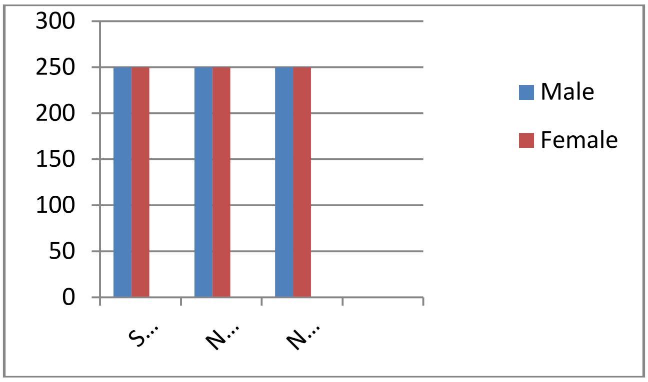
Age Group	Male	Female	Total
<10	750	750	1500
10-19			
20-29			
30-39			
40-49			
50-59			
60-69			
70-79			
80-89			
>90			

Through this study we can easily correlate the various age groups of diabetic patients with respect to their sex. We can also segregate the data according to three districts of Andaman & Nicobar Islands namely South Andaman District, North & Middle Andaman District and Nicobar District which are shown in below table.

Table 3: District-wise data of Diabetes Patients

District	Male	Female	Total
North and Middle Andaman District	250	250	500
South Andaman District	250	250	500
Nicobar District	250	250	500

Graph 2 :District - Wise Diabetic Patients distribution



Like-wise we can easily distribute the data based on localities such as Rural Area and Urban Area. We can also correlate the gender with these localities like Rural Area Male Diabetic Patients, Rural Area Female Diabetic Patients, Urban Area Male Diabetic Patients and Urban Area Female Diabetic Patients. We can also fetch the data like Rural Area Male Diabetic Patients in South Andaman District, Rural Area Female Diabetic Patients in South Andaman District,etc.

V. Developing a Data Mining Model to diagnose diabetes mellitus

Diabetes Mellitus is the most important cause of death globally. The disease diagnosis is a major process to treat the patients who are affected by diabetes. The diagnosis process is more difficult comparatively known about the diabetes disease detection. Developing a proposed data mining model is useful to diagnose the diabetes mellitus once the detection is accomplished.

In this study, a proposed data mining model has been separated into two different techniques which performs consecutively. These techniques are classification and clustering method of conceptual modeling.

Thus the diabetes data has to be converted into a knowledge base which is called as training data. The diabetic patient data and training data have beenbuilt with classification using decision tree.

The clustering techniques analyze the frequent pattern mining of diabetic patient whose data has been clustered based on the type of diabetes mellitus.

A. Clustering for Diabetics Disease Prediction and Detection

The clustering is one of the major decision processes which help to decide whether the group of disease data attributes has the disease defect or not.

Cluster builds the attributes set into a correlation based grouping method. It givesthe different cluster groups in relation with the cancer disease types.

In this study, the clustering is performed in two stages of process: *k-means* clustering of diabetes disease attributes cluster and Weighted Average Method (WAM) performed on extended attributes cluster which get the result from *k*-means.

k- means clustering Algorithm

K-Means Clustering is an Unsupervised Learning algorithm, which groups the unlabeled dataset into different clusters. Here K defines the number of pre-defined clusters that need to be created in the process, as if K=2, there will be two clusters, and for K=3, there will be three clusters, and so on.

It is an iterative algorithm that divides the unlabeled dataset into k different clusters in such a way that each dataset belongs to only one group that has similar properties.

k-means **Clustering for Classified Significant Pattern**

The instances are now clustered into a number of classes where each class is identified by a unique feature based on the significant patterns mined by the decision tree algorithm.

At the beginning, the data is assigned to a non-diabetic cluster and then based on the intensity of diabetes given by its weightage, it is either moved to the diabetic cluster or gets retained in the non-diabetic cluster and further the data object is moved between the subgroups of the hierarchical diabetic cluster based on the symptoms of the data object contains.

To calculate the mean of the cluster centroid, the symptoms are assigned certain values where the average represents each distinguished cluster. The data objects are distributed to the cluster based on the cluster centroid to which it is nearest. It also searches the entire database to find a match to a single input data. The data is moved to that particular cluster if and only if an exact match is found. This technique minimizes the error rate of clustering.

The data in the cluster is again fed into *k-means* clustering algorithm to further subdivide it. The resulting clusters are separated based on particular symptoms associated with any one type of diabetes mellitus i.e. Type 1 & Type 2.

Algorithm 1: *k*-means

$W(S, C) = {}_{k=1}^{K}\, {}_{i \in S_k}\, d(i, Ck)....$

Input: k: the number of clusters. D: training dataset containing n objects.
Output: A set of group of clusters
Step-1: Chooses two mean values from weightage of significant patterns as initial cluster centers;
Step-2: Assign each object to the cluster to which it is most similar based on the mean value of the weightage.
Step-3: Update the cluster means by calculating the mean value of all the objects in the cluster.
Given - number of clusters of a dataset containing a set of N entities; I, and M measurements; Here S = Two Diabetes types; C – Class;

Now two clusters have been generated based on the weightage scores of the significant pattern. These two clusters are named as non-diabetic and diabetic clusters.

B. Weighted Average Method (WAM) k-means clustering for detecting diabetes mellitus

Weighted Average Method (WAM) is used to improve the accuracy of analytic predictive performance models for diabetes prevention systems with more number of new patients.

Weighted Average method is a modification of the *k*-means criterion to include unknown weights, to the features v, v =1... Db1, Db2, Db3 (1,2,3) in the datasets.

Weighted Average cluster k-means could cluster the patient data of medical records to different groups, and divided into two groups which mapped as diabetes types based on DB1, DB2, DB3, etc.

The subset of diabetes types of instances with clusters will be processed towards weighted of *k*-means in specific features, parameter range. The sum of all the predicted values can be averaged by set of instances.

Algorithm 2: Weighted Average Method (WAM)

$$W(S, C, w) = \Sigma_{k=1}^{k} \Sigma_{i \in I} \Sigma_{v=1}^{M} S_{ik} w_v^{\beta} (y_{iv} - C_{kv})^2 / i;$$

Input : I- user selected parameters; ' '- user-defined parameter; i – new Instance (patient); $Y_{i\,v}$ – Class value; W_v – unknown weighted feature;

Output : weighted average value of clustered group.

Step 1 : Initial setting Similar to *k*-means, the method uses an external parameter to define the number K of clusters. WAM *k*-means then randomly initializes the centroids and the feature weights, ensuring that they add up to unity.

Step 2 : Cluster updates the assigns each entity to their closest centroid, using the adjusted distance measure that takes into account the feature weights.

Step 3 : Stop condition...As per the *k*-means clustering algorithm.

Step 4 : Centroids update centroids to the centre of gravity of its cluster. Again, as the method is based on Euclidean distance. It is achieved by moving the centroids to the mean of their clusters.

Step 5 : Weights Average updates the weight of feature vector of according to WAM *k*- means, constrained by the condition that the sum of weights should be one.

VI. Experimental Result Analysis

Multiple Data mart has developed which trained the existing patient data and build the well structure Diabetic Data Warehouse (DDW).In this connection, establishment of the data design such as diabetic data modeling, diabetic disease normalization and their attributes which facilitate measurements of the effectiveness of treatment, relationships between casualty and treatment practice for systematic diseases and conditions are captured.

Here we have done the weighted average method based functional performance analysis for diabetic disease attribute impacts. Based on the Impact analysis, the system will obtain different resultant information for diabetic diagnosis and treatment guidance.However, knowing the risk factors of diabetes disease attributes and resultant information is the well knowledge obtained from our cancer detection & prevention system.

Through this research,several data has processed from multidimensional diabetes data warehouse and multivariate data has derived from the data mining model. The diabetes disease prediction based on *k*- means clustering has been better improved by Weighted Average-based prediction value which maps the diabetes types and disease factors as symptoms along with the blood group factor.

VII. Conclusions

The data warehouse is a data structure that is optimized for mass data storage and complex query processing which is used to analyze the data information. The patient data has uploaded in the data storage called database it performs basis of On-Line Transaction Processing. The data uploaded may contain irrelevant to cancer disease attributes, missing patient details, incorrect user entries are known as data noise. The data staging area of ECTL which extract the patient data, cleansing which remove the noisy data, transact and load into the data warehouse. The various Data mart has been formed as diabetes symptoms, diagnosis recommendation, and decision system. The data mart in the diabetes data warehouse has been planned to consult the diabetes patient attribute set. The diabetes data attributes has analyzed and developed the multidimensional star-schema which contains fact, dimensional tables can perform the OLAP operations.

There's no cure for type 1 diabetes. It requires lifelong disease management. But with consistent monitoring and treatment, you may be able to avoid more serious complications of the disease.If you work closely with your doctor and make good lifestyle choices, type 2 diabetes can often be successfully managed.

REFERENCES

[1]. Amira Hassan Abed and Mona Nasr (2019), "Diabetes Disease Detection through Data Mining Techniques", Int. J. Advanced Networking and Applications, Volume: 11 Issue: 01 Pages: 4142-4149(2019) ISSN: 0975-0290.

[2]. Seyed Ataaldin Mahmoudinejad Dezfuli et. al. (2019), "Early Diagnosis of Diabetes Mellitus Using Data Mining and Classification Techniques", Jundishapur Journal of Chronic Disease Care, Volume 8, Issue 3.

[3]. Sneha N. and Tarun Gangil (2019), "Analysis of diabetes mellitus for early prediction using optimal features selection", Article No. 13.

[4]. Deepti Sisodia and Dilip Singh Sisodia (2018), "Prediction of Diabetes using Classification Algorithms", Procedia Computer Science, Volume 132, Pages 1578-1585.

[5]. Deeraj Shetty et. al. (2017), "Diabetes disease prediction using data mining", International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), **INSPEC AccessionNumber:** 17558894,**DOI:** 10.1109/ICIIECS.2017.8276012 (IEEE).

[6]. Nimna Jeewandar et. al. (2017), "Data Mining Techniques in Prevention And Diagnosis of Non Communicable Diseases", international journal of Research in Computer Applications and Robotics, Volume 5, Issue, 11, Pp. 11-17, ISSN 2320-7345.

[7]. Loannis Kavakiotis et. al. (2017), "Machine Learning and Data Mining Methods in Diabetes Research", Computational and Structural Biotechnology Journal(Elsevier), Volume 15, Pages 104-116.

[8]. Harleen and Bhambri Pankaj (2016), "A Prediction Technique in Data Mining for Diabetes Mellitus", Apeejay-Journal of Management Sciences and Technology, Volume 4, Issue 1, ISSN -2347-5005.

[9]. Shivakumar B and Alby S (2014), "A survey on data-mining technologies for prediction and diagnosis of diabetes", International Conference on Intelligent Computing Applications, ICICA 2014 (2014) 167-173

[10]. Ravneet Jyot Singh and Williamjeet Singh (2012), "Data Mining in Healthcare for Diabetes Mellitus", International Journal of Science and Research (IJSR), ISSN (Online): 2319-7064

[11]. Miroslav Marinov et. al. (2011), "Data-Mining Technologies for Diabetes: A Systematic Review", Journal of Diabetes Science and Technology, Volume 5, Issue 6.

[12]. https://www.indiatoday.in/education-today/ gk-current-affairs/story/98-million-indians-diabetes-2030-prevention-1394158-2018-11-22

[13]. https://www.mayoclinic.org/diseases-conditions/diabetes/symptoms-causes/syc-20371444

[14]. https://www.who.int/news-room/fact-sheets/detail/diabetes

[15]. https://www.who.int/health-topics/diabetes#tab=tab_1

[16]. https://www.diabetesatlas.org/en/sections/worldwide-toll-of-diabetes.html

[17]. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3068646/

[18]. https://www.diabetesresearchclinicalpractice.com/article/S0168-8227(19)31230-6/fulltext

[19]. https://dmsjournal.biomedcentral.com/articles/10.1186/s13098-017-0207-3

[20]. https://www.educba.com/what-is-star-schema/

[21]. http://www.lastnightstudy.com/Show?id=54/OLAP-Operations#:~:text=Some%20popular%20OLAP%20operations%20for%20multidimensional%20data%20are%3A-,3%20Slice%20and%20dice.%20...%204%20Pivot.%20

[22]. https://www.javatpoint.com/k-means-clustering-algorithm-in-machine-learning